



Ministério da
Ciência e Tecnologia



INPE-15325-TDI/1368

CLASSIFICAÇÃO IMPRECISA DE DADOS BASEADA EM LÓGICA DIFUSA E TEORIA DE POSSIBILIDADES

Isabela Neves Drumond

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada,
orientada pelos Drs. Solon Venâncio de Carvalho e Sandra Aparecida Sandri,
aprovada em 14 de de setembro de 2007.

Registro do documento original:

<<http://urlib.net/sid.inpe.br/mtc-m17@80/2007/12.07.10.59>>

INPE
São José dos Campos
2008

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3945-6911/6923

Fax: (012) 3945-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO:

Presidente:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Membros:

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Haroldo Fraga de Campos Velho - Centro de Tecnologias Especiais (CTE)

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Jefferson Andrade Ancelmo - Serviço de Informação e Documentação (SID)

Simone A. Del-Ducca Barbedo - Serviço de Informação e Documentação (SID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Marilúcia Santos Melo Cid - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Viveca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)



Ministério da
Ciência e Tecnologia



INPE-15325-TDI/1368

CLASSIFICAÇÃO IMPRECISA DE DADOS BASEADA EM LÓGICA DIFUSA E TEORIA DE POSSIBILIDADES

Isabela Neves Drumond

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada,
orientada pelos Drs. Solon Venâncio de Carvalho e Sandra Aparecida Sandri,
aprovada em 14 de de setembro de 2007.

Registro do documento original:

<<http://urlib.net/sid.inpe.br/mtc-m17@80/2007/12.07.10.59>>

INPE
São José dos Campos
2008

Dados Internacionais de Catalogação na Publicação (CIP)

D845c Drummond, Isabela Neves.

Classificação imprecisa de dados baseada em lógica difusa e teoria de possibilidades/ Isabela Neves Drummond. – São José dos Campos: INPE, 2008.
157p. ; (INPE-15325-TDI/1368)

Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2007.

1. Imprecisão. 2. Incerteza. 3. Classificação. 4. Teoria de possibilidades. 5. Conjuntos difusos. I. Título.

CDU 551.58

Copyright © 2008 do MCT/INPE. Nenhuma parte desta publicação pode ser reproduzida, armazenada em um sistema de recuperação, ou transmitida sob qualquer forma ou por qualquer meio, eletrônico, mecânico, fotográfico, microfílmico, reprográfico ou outros, sem a permissão escrita da Editora, com exceção de qualquer material fornecido especificamente no propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

Copyright © 2008 by MCT/INPE. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the reader of the work.

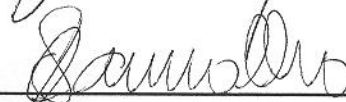
Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de Doutor(a) em
Computação Aplicada

Dr. José Demisio Simões da Silva



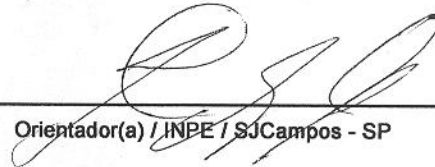
Presidente / INPE / SJC Campos - SP

Dr. Solon Venâncio de Carvalho



Orientador(a) / INPE / SJC Campos - SP

Dra. Sandra Aparecida Sandri



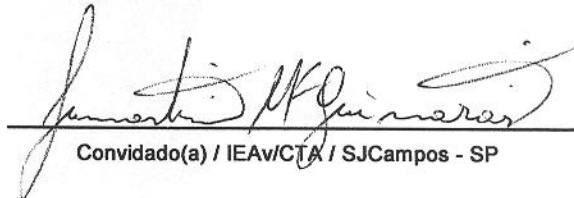
Orientador(a) / INPE / SJC Campos - SP

Dra. Corina da Costa Freitas



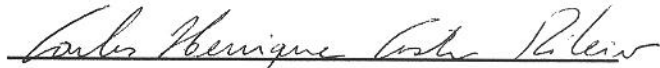
Membro da Banca / INPE / SJC Campos - SP

Dr. Lamartine Nogueira Frutuoso
Guimarães



Convidado(a) / IEAV/CTA / SJC Campos - SP

Dr. Carlos Henrique Costa Ribeiro



Convidado(a) / ITA / São José dos Campos - SP

Dr. Jacques Wainer



Convidado(a) / UNICAMP / Campinas - SP

Aluno (a): Isabela Neves Drummond

São José dos Campos, 14 de Setembro de 2007

*“Sede alegres na esperança, pacientes na tribulação
e perseverantes na oração”.*

(Rm, 12, 12)

A meus pais.

AGRADECIMENTOS

Agradeço a Deus.

Agradeço à minha orientadora, Dra. Sandra Aparecida Sandri, pelo apoio, incentivo e a amizade de sempre. E ao meu orientador Dr. Solon Venâncio de Carvalho.

Ao professor Dr. José Demísio Simões da Silva pela amizade e pela atenção dispensada a todas as minhas dúvidas.

Ao professor Dr. Juan Serrat pela orientação durante o trabalho de doutorado sanduíche na Universitat Autònoma de Barcelona (UAB).

Aos membros da banca examinadora pela disposição em analisar este trabalho.

À professora Dra. Corina da Costa Freitas e à aluna Flávia de Toledo Martins por disponibilizar os dados do Projeto GeoSchisto. E ao professor Dr. Joaquim Meléndez pelos dados das estações elétricas da Espanha.

Ao Dr. Ricardo Oliveira Falcão e à clínica Plani pela atenção dispensada e disponibilização das imagens de ressonância magnética empregadas neste trabalho.

Ao meu pai pelo incentivo e por acreditar em mim sempre, à minha mãe, pelo amor e pela dedicação, e ao meu irmão, que sempre torceu por mim.

Ao meu namorado Alexandre, por tudo.

Aos melhores amigos do mundo, Apac e Elcio, pela amizade sincera. A todos os amigos que sempre torceram por mim, em especial, Leo, Bet e Cris. Ao professor Ari, pelo incentivo. E a todos os amigos maravilhosos do Institut d'Investigació en Intel·ligència Artificial (IIIA). Moltes gràcies!

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro, ao Instituto Nacional de Pesquisas Espaciais (INPE) pela oportunidade e apoio, ao Serviço de pós graduação, em especial, à Carol; e a todos os professores e funcionários do Laboratório Associado de Computação e Matemática Aplicada (LAC), em especial, às secretárias Cristina, Vanessa, Neusa, Amanda e Fabiana por todos os favores prestados.

Agradeço a todos aqueles que direta e indiretamente colaboraram para a realização deste trabalho.

RESUMO

Este trabalho aborda estratégias de classificação de dados em espaços multidimensionais, que têm por base a lógica difusa e a teoria de possibilidades. São apresentados métodos de classificação que retornam distribuições de possibilidade para cada elemento do conjunto de dados, além de uma coleção de índices de classificação que auxiliam na busca pelo melhor classificador impreciso. Uma família parametrizada de classificadores imprecisos é definida a partir da distribuição de possibilidade proveniente de um método de classificação, e com base nos índices de classificação definem-se os conceitos de precisão, imprecisão parcial e imprecisão completa para o classificador. Estes índices favorecem o processo de combinação de classificadores, permitindo que seja ponderado o nível de imprecisão aceitável para uma aplicação específica, de acordo com as necessidades do usuário. É proposto o modelo de classificação AGRUP, um algoritmo supervisionado, baseado em agrupamento, que gera distribuições de possibilidade associadas a todos os elementos do conjunto de dados, a partir da aplicação de relações de similaridade. Por fim, na busca pela classificação precisa, uma heurística pode ser empregada, como o refinamento pela vizinhança, onde uma classificação imprecisa obtida em termos de classes puras e compostas é reduzida a classes puras somente. A classificação propriamente dita é feita no espaço de atributos, enquanto que a vizinhança é feita no espaço do problema. As abordagens propostas são testadas em diferentes aplicações: classificação de imagens de ressonância magnética (RM) do cérebro, imagens de satélite, além do conjunto de dados que representam registro de queda na distribuição de energia em subestações elétricas e de dados do Projeto GeoSchisto, relacionados à distribuição da esquistossomose no estado de Minas Gerais. Os resultados obtidos apontam para a viabilidade das estratégias propostas, no que diz respeito à flexibilidade, aplicabilidade e acurácia, tanto em dados sintéticos, quanto em dados reais.

IMPRECISE DATA CLASSIFICATION BASED ON FUZZY LOGIC AND POSSIBILITY THEORY

ABSTRACT

This work approaches strategies of data classification in multidimensional spaces, based on fuzzy logic and possibility theory. Classification methods that can yield possibilistic valuations as output for each element of the data set are presented, in addition to a collection of classification indexes that aid in finding the best imprecise classifier. A parameterized family of imprecise classifiers is defined from the possibilistic valuations provided by a given classification method and, based on classification indexes, accuracy measures for partial imprecision and complete imprecision are also defined for the classifier. These indexes favor the process of combination of classifiers, allowing a level of acceptable imprecision for a specific application, in agreement with the user's needs. The classification model, named AGRUP, is proposed in this work as a supervised algorithm based on a clustering approach, that generates possibilistic valuations associated to all elements of the data set, by the application of similarity relations on the cluster prototypes. Finally, in the search for the precise classification, a neighborhood heuristic can be employed as a refinement process, where an imprecise classification, obtained in terms of pure and composed classes, is reduced to pure classes only. The classification process is made in the space of attributes, while the neighborhood heuristic is made in the space of the problem. The proposed approaches are tested in different applications: classification of magnetic resonance images of the brain, satellite images, besides the data set representing the origin of sags registered in electricity distribution substations and another data set obtained from the GeoSchisto Project, related to the distribution of the Schistosomosis disease in the state of Minas Gerais, Brazil. The achieved results point to the viability of the proposed classification approaches, concerning the flexibility, applicability, accuracy, no matter whether synthetic or real, noisy or noiseless data are being classified.

SUMÁRIO

Pág.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

LISTA DE SÍMBOLOS

CAPÍTULO 1 - INTRODUÇÃO	27
CAPÍTULO 2 - ELEMENTOS DA TEORIA DOS CONJUNTOS DI- FUSOS	33
2.1 - Conjunto difuso e função de pertinência	33
2.2 - Termos básicos	34
2.3 - Operações	36
2.4 - Relações difusas	38
2.4.1 - Relações de similaridade	38
2.5 - Teoria de possibilidades	39
2.5.1 - Fundamentos	39
2.5.2 - Transformação da distribuição de possibilidade em probabilidade e vice- versa	41
CAPÍTULO 3 - A TAREFA DE CLASSIFICAÇÃO	43
3.1 - Definições	43
3.2 - Modelos de classificação	44
3.2.1 - Algoritmos de agrupamento	45
3.2.2 - Classificadores possibilistas	53
3.2.3 - Classificador Bayesiano	55
3.2.4 - Classificador por Máxima Verossimilhança (MAXVER)	56
3.2.5 - Classificador dos k vizinhos mais próximos (KNN)	56
3.2.6 - Árvores de decisão	60
3.2.7 - Redes Neurais Artificiais (RNA)	60
3.3 - Combinação de classificadores	62
3.4 - O desempenho de um classificador	64

3.4.1 - Taxa de acerto	65
3.4.2 - Conjuntos de treinamento e teste	66
CAPÍTULO 4 - ESTRATÉGIAS DE CLASSIFICAÇÃO POSSÍ- BILISTA	69
4.1 - Definições	70
4.2 - O classificador AGRUP	71
4.2.1 - Transformação do espaço de atributos	71
4.2.2 - Algoritmo de classificação	77
4.2.3 - Amostras definidas como centros	77
4.3 - Uma abordagem alternativa	78
4.4 - Normalização	79
CAPÍTULO 5 - CLASSIFICAÇÃO IMPRECISA	81
5.1 - Família parametrizada de classificadores imprecisos	82
5.2 - A matriz de confusão imprecisa	83
5.3 - Índices de acurácia e confiança do classificador impreciso	85
5.4 - Medida de utilidade de imprecisão	88
5.5 - Agregação de classificadores imprecisos	88
5.6 - Heurísticas de decisão	89
5.6.1 - Heurística do valor máximo: n-MAX	90
5.6.2 - Refinamento pela vizinhança	91
CAPÍTULO 6 - APLICAÇÕES	93
6.1 - Classificação de imagens	94
6.1.1 - Imagens de RM	96
6.1.2 - Imagens de sensoriamento remoto	98
6.2 - Experimentos	99
6.2.1 - Imagens de RM do cérebro com EM - Banco de dados sintético	100
6.2.2 - Imagens de RM do cérebro com EM - Banco de dados real	112
6.2.3 - Imagem LANDSAT	120
6.2.4 - Imagem CBERS	126
6.2.5 - Dados de registro de queda na distribuição de energia elétrica	131
6.2.6 - Dados do projeto GeoSchisto	137
6.3 - Considerações Finais	143
CAPÍTULO 7 - CONCLUSÃO	145
REFERÊNCIAS BIBLIOGRÁFICAS	149

LISTA DE FIGURAS

	<u>Pág.</u>
2.1 Funções de pertinência: a) Triangular; b) Trapezoidal; c) Gaussiana.	34
3.1 Exemplo de um dendrograma.	47
3.2 Conjunto de dados classificado em dois grupos por agrupamento difuso. . .	49
4.1 a) Conjunto sintético b) Conjunto sintético com centros associados às classes.	71
4.2 Parâmetros: d_{max} , d_{ext} , d_{prox} e d_{in}	73
4.3 Meta-agrupamento global com: a) 8 centros b) 9 centros e c) 10 centros. .	74
4.4 Fases do meta-agrupamento local.	75
5.1 Classificação precisa (Cl_1) e classificação imprecisa ($Cl_{0,4}$) para o conjunto de treinamento cone-torus (KUNCHEVA, 2000) (Número de centros associados a cada classe $\langle 5, 5, 10 \rangle$).	84
5.2 Índices de classificação de Cl_α para o conjunto de treinamento (número de centros por classe: $\langle 5, 5, 10 \rangle$): a) baseados em imprecisão, b) baseados em incerteza.	88
6.1 Metodologia.	93
6.2 Cortes de uma imagem de RM.	98
6.3 Modelos anatômicos: a) 11 classes e b) 5 classes	100
6.4 Cortes axiais: a) tipo T1 e b) tipo T2.	101
6.5 Espaço de atributos formado pelas amostras das imagens T1 e T2: a) sem ruído, b) com 3% de ruído e c) com 9% de ruído.	102
6.6 Gráficos de classificação para as imagens de RM sintética: a) 0% ruído, AGRUP Te2; b) 0% ruído, FKNN Te2; c) 3% ruído, AGRUP Te1; d) 3% ruído, FKNN Te1; e) 9% ruído, AGRUP Te5; f) 9% ruído, FKNN Te5 . . .	109
6.7 Combinação dos modelos AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ e FKNN com $k = 5$: a) experimento 4 e b) experimento 5	110
6.8 Gráficos de classificação para as imagens de RM sintética: a) 9% ruído, AGRUP Te4; b) 9% ruído, FKNN Te4; c) classificador $Cl_{0,10;\alpha_2}$	111
6.9 Cortes provenientes de volumes de RM do banco de dados real: a) corte 14 T2, b) corte 14 Flair, c) corte 15 T2 e d) corte 15 Flair.	112
6.10 Espaço de atributos formado pelas amostras das imagens T2 e Flair	113
6.11 Gráficos de classificação para a imagem de RM real: a) AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ Te2; b) AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ Te5; c) AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ Te2; d) AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ Te5; e) FKNN com $k = 5$ Te2 e f) FKNN com $k = 5$ Te5. . .	116

6.12	Combinação dos classificadores: a) AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ + AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ experimento 5 e b) AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ experimento 2 e FKNN experimento 5.	117
6.13	Classificador $Cl_{0,4;\alpha_2}$ proveniente da combinação apresentada em 6.12a). . .	117
6.14	Imagens de treinamento e teste classificadas: a) corte 14 e b) corte 15. . . .	119
6.15	Imagem LANDSAT.	120
6.16	Espaço de atributos normalizado formado pelas amostras da imagem LANDSAT nas bandas 2, 3 e 4.	120
6.17	Gráficos de classificação - imagem satélite LANDSAT - a) AGRUP Te4 e b) FKNN Te4.	123
6.18	Imagem LANDSAT classificada.	124
6.19	Imagem CBERS	126
6.20	Espaço de atributos normalizado formado pelas amostras da imagem CBERS nas bandas 2, 3 e 4.	126
6.21	Gráficos de classificação para imagem CBERS: a) classificador AGRUP Te1 e b) classificador FKNN Te1.	129
6.22	Espaço de atributos formado pelos atributos amplitude e duração do evento. 131	
6.23	Gráficos de classificação para os dados de registro de rede elétrica do classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ e FKNN com $k = 1$ e $k = 3$: a) AGRUP Te2; b) AGRUP Te4; c) FKNN com $k = 1$ Te2; d) FKNN com $k = 1$ Te4; e) FKNN com $k = 3$ Te2 ; f) FKNN com $k = 3$ Te4.	135
6.24	Gráficos de classificação, Cl^1 e Cl^2 , para o conjunto de dados de registros de rede elétrica. Ambos empregando o método AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ e número de centros fixos em 5 para cada classe.	136
6.25	Gráficos de classificação: a) $Cl_{\alpha_1;0,58}^{12}$ e b) $Cl_{0,60;\alpha_2}^{12}$	136
6.26	Espaço de atributos formado pelos dados do Projeto GeoSchisto, caracterizados pelas variáveis DEC, Veg_I e Tmin_V.	138
6.27	Imprecisão parcial referente ao experimento 1 (Te1) para os métodos AGRUP (met1) e FKNN (met2).	142
6.28	Gráficos de classificação para os dados do Projeto Schisto: a) AGRUP Te2; b) FKNN Te2; c) AGRUP Te3 e d) FKNN Te3.	142

LISTA DE TABELAS

	<u>Pág.</u>
2.1 Principais t-normas e t-conormas.	37
5.1 Matriz de confusão imprecisa para o classificador Cl_α num problema com m classes	83
5.2 Matriz de confusão imprecisa para o classificador Cl_α num problema com 3 classes	83
5.3 Matriz de confusão imprecisa para os classificadores a) preciso Cl_1 e b) impreciso $Cl_{0,4}$	84
6.1 Parâmetros usualmente empregados num sistema de RM	98
6.2 Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética sem ruído.	102
6.3 Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética sem ruído.	103
6.4 Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética sem ruído.	103
6.5 Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética sem ruído.	103
6.6 Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética sem ruído.	104
6.7 Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética sem ruído.	104
6.8 Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM sintética com 3% de ruído.	105
6.9 Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 3% de ruído.	105
6.10 Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM sintética com 3% de ruído.	105
6.11 Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 3% de ruído.	106

6.12	Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM sintética com 3% de ruído.	106
6.13	Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 3% de ruído.	106
6.14	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética com 9% de ruído.	107
6.15	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 9% de ruído.	107
6.16	Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética com 9% de ruído.	107
6.17	Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 9% de ruído.	108
6.18	Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética com 9% de ruído.	108
6.19	Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 9% de ruído.	108
6.20	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM real.	114
6.21	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ - imagem de RM real.	114
6.22	Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM real.	114
6.23	Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ - imagem de RM real.	115
6.24	Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM real.	115
6.25	Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ - imagem de RM real.	115
6.26	Índices <i>kappa</i> para a imagem de RM real.	119
6.27	Imprecisão completa em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - LANDSAT.	121
6.28	Imprecisão completa em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem LANDSAT.	121

6.29	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem LANDSAT.	122
6.30	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem LANDSAT.	122
6.31	Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem LANDSAT.	122
6.32	Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem LANDSAT.	123
6.33	Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - LANDSAT.	123
6.34	Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem LANDSAT.	123
6.35	Índices <i>kappa</i> e <i>acc</i> para a imagem LANDSAT após refinamento.	124
6.36	Índices <i>kappa</i> e <i>acc</i> para a imagem LANDSAT classificada por redes neurais em Silva (2004).	125
6.37	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem CBERS.	127
6.38	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem CBERS.	128
6.39	Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem CBERS.	128
6.40	Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem CBERS.	128
6.41	Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem CBERS.	129
6.42	Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem CBERS.	129
6.43	Valores de <i>kappa</i> antes e após o refinamento pela vizinhança.	130
6.44	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - Dados registro de rede elétrica.	132
6.45	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 1$ - Dados registro de rede elétrica.	132

6.46	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados registro de rede elétrica.	133
6.47	Imprecisão completa em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - Dados registro de rede elétrica.	133
6.48	Imprecisão completa em 5 experimentos usando o classificador FKNN com $k = 1$ - Dados registro de rede elétrica.	133
6.49	Imprecisão completa em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados registro de rede elétrica.	133
6.50	Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - Dados registro de rede elétrica.	134
6.51	Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 1$ - Dados registro de rede elétrica.	134
6.52	Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados registro de rede elétrica.	134
6.53	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - Dados Projeto GeoSchisto.	139
6.54	Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados Projeto GeoSchisto.	140
6.55	Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - Dados Projeto GeoSchisto.	140
6.56	Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados Projeto GeoSchisto.	140
6.57	Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - Dados Projeto GeoSchisto.	140
6.58	Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados Projeto GeoSchisto.	141
6.59	Imprecisão completa em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - Dados Projeto GeoSchisto.	141
6.60	Imprecisão completa em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados Projeto GeoSchisto.	141
6.61	Matrizes de confusão: a) amostras de treinamento e b) amostras de teste .	143

LISTA DE ABREVIATURAS E SIGLAS

AGRUP	–	Classificador possibilista baseado em AGRUPamento
ART	–	<i>Adaptive Resonance Theory</i>
CBERS	–	<i>China Brazil Earth Resources Satellite</i>
CPTEC	–	Centro de Previsão de Tempo Estudos Climáticos
EM	–	Esclerose Múltipla
FCM	–	<i>Fuzzy C-Means</i>
FKNN	–	<i>Fuzzy K-Nearest Neighbor algorithm</i>
IBGE	–	Instituto Brasileiro de Geografia e Estatística
KNN	–	<i>K-Nearest Neighbor algorithm</i>
LVQ	–	<i>Learning Vector Quantization</i>
MAXVER	–	Máxima Verossimilhança
MLP	–	<i>Multi-Layer Perceptron</i>
OWA	–	<i>Ordered weighted average</i>
RM	–	Ressonância Magnética
RNA	–	Redes Neurais Artificiais
RBF	–	<i>Base Radial Function</i>
SR	–	Sensoriamento Remoto
SOM	–	<i>Self Organizing Map</i>
SVM	–	<i>Support Vector Machine</i>

LISTA DE SÍMBOLOS

acc	–	Acurácia ou taxa de acerto do classificador
B	–	Conjunto de amostras $\in Z$
C	–	Conjunto de classes
Cl	–	Classificador
c^*	–	Classe correta de um elemento
$card$	–	Cardinalidade do conjunto difuso
d	–	Distância entre dois pontos
d_{ext}	–	Raio externo da classe
d_{in}	–	Raio interno
d_{max}	–	Raio máximo da classe
d_{min}	–	Raio mínimo da classe
d_{prox}	–	Raio do centro mais próximo
$d_{proxext}$	–	Ponto mais próximo de outra classe
d_{proxin}	–	Ponto mais próximo da mesma classe
err	–	Erro do classificador
imp	–	Imprecisão completa do classificador
k	–	Número de vizinhos de um elemento
mis	–	Índice de agrupamento
mm	–	Parâmetro de ajuste do suporte da relação de similaridade
nn	–	Parâmetro de ajuste do núcleo da relação de similaridade
N	–	Medida de necessidade
P	–	Medida de probabilidade
p	–	Distribuição de probabilidade
$qual^{*\kappa}$	–	Índice de qualidade do classificador
RF	–	Rádio Frequência
S^G	–	Relação de similaridade gaussiana
S^L	–	Relação de similaridade linear
T	–	Conjunto de protótipos de uma classe
T1	–	Imagem de RM com TR<800ms e TE<20ms
T2	–	Imagem de RM com TR>1500ms e TE>70ms
Te	–	Indica teste
TE	–	Tempo de decaimento
TI	–	Tempo de inversão
Tr	–	Indica treinamento
TR	–	Tempo de repetição dos pulsos de rádio frequência
Z	–	Conjunto de dados
Z_{te}	–	Conjunto de teste
Z_{tr}	–	Conjunto de treinamento
α	–	Parâmetro da família de classificadores imprecisos
γ	–	Núcleo da relação de similaridade
κ	–	Cardinalidade do elemento, ou seja, número de classes associadas

- λ – Suporte da relação de similaridade
- π – Distribuição de possibilidade
- Π – Medida de possibilidade
- Ψ – Espaço de atributos
- ρ – Dimensão de um vetor de atributos

CAPÍTULO 1

INTRODUÇÃO

A classificação é o processo pelo qual um elemento, pertencente a um conjunto qualquer de dados, é atribuído a uma classe definida em um conjunto finito de classes. Ou seja, um elemento é reconhecido como parte de um conjunto de entidades iguais ou similares a ele, possivelmente identificado por um nome. Num contexto mais geral, pode-se dizer que a classificação é parte de um processo mais complexo definido como reconhecimento de padrões, área de pesquisa que tem por objetivo a classificação de objetos (padrões).

Segundo Bezdek e Pal (1992), o reconhecimento de padrões pode ser definido como a busca de estruturas em um conjunto de dados. Para Duda *et al.* (2000), o processo de reconhecimento de padrões se divide em subproblemas, o que torna o processo mais compreensível. Dentro destes subproblemas destacam-se as etapas de pré-processamento, extração de feições e classificação. No pré-processamento, obtém-se o conjunto de dados no qual se deseja identificar os padrões. Em seguida, são estabelecidos critérios medindo-se características presentes nos dados como forma de redução do conjunto de dados, ou ainda, segundo Nadler e Smith (1993), torná-lo mais aceitável para o processo de decisão. E por fim, um classificador identifica os padrões com base nas feições definidas na fase anterior.

Um ser humano pratica com facilidade o ato de reconhecer padrões como, por exemplo, no reconhecimento de faces, na leitura de caracteres ou no entendimento da própria fala. Contudo, o processo é extremamente complexo, sendo natural buscar o projeto de máquinas que sejam capazes de desempenhar a atividade de reconhecer padrões (DUDA *et al.*, 2000). Dado um padrão, que é uma entidade, vagamente definida e possivelmente nomeada, sua classificação pode ser (i) supervisionada, na qual um padrão é identificado como membro de uma classe pré-definida, ou (ii) não-supervisionada, onde o padrão é atribuído a uma classe desconhecida segundo uma métrica de similaridade.

O problema de classificação de padrões pode ser identificado em diversas áreas de pesquisa como a biologia, psicologia, medicina, visão computacional, inteligência artificial e sensoriamento remoto, destacando o recente avanço de aplicações em reconhecimento de padrões em bioinformática, mineração de dados, automação industrial, análise de imagens médicas e de satélite, entre outras. O rápido crescimento e a demanda por métodos que possibilitem o processamento de conjuntos de dados cada vez maiores, facilita o uso e a elaboração de métodos para classificação de dados, com conseqüente necessidade do projeto de sistemas automáticos de reconhecimento

de padrões. Dado o desenvolvimento na área e os métodos existentes, é visível que em determinadas aplicações uma única abordagem de classificação não gera soluções “ótimas”, encorajando o emprego de múltiplas abordagens e a combinação de métodos.

Na literatura não existe apenas uma categorização das diferentes abordagens para os métodos de classificação. Uma das grandes divisões encontradas é a definição dos grupos: classificadores estatísticos e classificadores sintáticos (DUDA *et al.*, 2000). As subdivisões dependem das propriedades que se deseja enfatizar e uma categorização adotada não apresenta abordagens necessariamente independentes e, de acordo com as propriedades destacadas um método pode apresentar diferentes interpretações. Em Kuncheva (2000) os métodos são divididos em: (A) aproximadores de PDF's (função densidade de probabilidade) e (B) aproximadores de fronteiras de classificação ou funções discriminantes; (A) se subdivide em classificadores paramétricos, semi paramétricos e não paramétricos; e (B) em funcionais lineares e não lineares, e estruturais. Já em Jain *et al.* (2000), três abordagens são identificadas: classificadores baseados em similaridade, classificadores probabilísticos e classificadores geométricos. E em Leondes (1998) se discute abordagens neurais e os métodos estatísticos clássicos. Destacam-se os modelos de classificação que envolvem lógica difusa e teoria de possibilidades, que associam graus de pertinência aos elementos com relação a cada classe envolvida no problema.

Sabe-se que a teoria dos conjuntos difusos¹ foi primeiramente aplicada ao estudo de problemas relacionados à classificação de padrões em Bellman *et al.* (1966). O reconhecimento de padrões é um importante aspecto da percepção humana, à qual se associa naturalmente informação de caráter vago ou impreciso. Assim, é natural a referência ao reconhecimento de padrões difuso, que na literatura é identificado no agrupamento difuso, nos sistemas de regras difusas *se-então* usados como classificadores ou, numa visão mais geral, em qualquer paradigma de classificação que envolva conjuntos difusos. A teoria destes conjuntos, introduzida por Zadeh (1965), lida com informação de caráter vago ou impreciso. Os classificadores difusos são assim definidos por empregarem conjuntos difusos durante sua fase de treinamento e/ou fase de operação. Kuncheva (2000) apresenta uma revisão dos classificadores difusos definidos na literatura.

Uma característica que merece ser enfatizada é a capacidade de um classificador apresentar como saída uma distribuição de possibilidade associada a cada elemento. Esta distribuição pode ser obtida diretamente, através de classificadores possibilistas,

¹O termo em inglês *fuzzy* é traduzido para o português como difuso ou nebuloso.

ou indiretamente, através da transformação de distribuições de probabilidade em possibilidades (DRUMMOND; SANDRI, 2006). Para um classificador supervisionado, define-se como local, a abordagem que associa a cada elemento da amostra uma distribuição de possibilidade, e como global aquele modelo que gera distribuições associadas a todo elemento dentro do domínio do problema.

Abordagens difusas e possibilistas definidas na literatura podem ser caracterizadas como locais ou globais no que se refere às distribuições associadas a cada elemento. O classificador FKNN, definido por Keller *et al.* (1985), gera distribuições com restrições probabilísticas no domínio do conjunto de amostras. A classificação de um elemento cuja classe é desconhecida é feita com base na distribuição associada aos k vizinhos mais próximos. O classificador híbrido de Malinski (2000), pode ser definido como global, com distribuições de possibilidade derivadas no domínio total do problema.

O emprego de uma abordagem de classificação, seja ela difusa ou não, se apresenta adequada dependendo das características do conjunto de dados. Um dos problemas apontados na classificação de dados é justamente a escolha da técnica que apresente melhor desempenho para um determinado conjunto. Assim, avaliar o desempenho de um classificador é essencial, uma vez que o objetivo é encontrar uma classificação que atenda aos requisitos do problema abordado. Num processo supervisionado, o classificador é submetido aos dados de treinamento e verificado quanto à capacidade de generalização a partir dos dados de teste. A análise estatística dos elementos corretos e erroneamente classificados fornece um valor de desempenho da classificação para uma determinada aplicação (KUNCHEVA, 2000).

Este trabalho de tese aborda o estudo do desempenho de uma classificação imprecisa. Obtida a distribuição de possibilidade referente à classificação de cada elemento de um conjunto, a classificação imprecisa a qual o trabalho se refere é caracterizada pela associação de um subconjunto do conjunto das possíveis classes a cada elemento do conjunto (MALINSKI, 2000). A partir de uma classificação possibilista é gerada uma família parametrizada de classificadores imprecisos, que variam do mais correto e impreciso ao preciso, porém com maior taxa de erro (DRUMMOND; SANDRI, 2006).

É proposta uma estratégia de classificação de dados definidos no espaço multidimensional baseada em teoria de possibilidades, o algoritmo AGRUP. Trata-se de um modelo supervisionado baseado em agrupamento, que retorna distribuições de possibilidade, indicando a compatibilidade do elemento com relação a cada classe envolvida no problema.

O algoritmo definido se baseia em protótipos, onde, para cada classe de um dado conjunto de amostras, são gerados centros através de um algoritmo de agrupamento. Estes centros ou protótipos são transformados em protótipos difusos pela aplicação de relações de similaridade, transformando o espaço de atributos formado pelo conjunto de amostras de um dado problema. A partir destes conjuntos são derivadas as distribuições de possibilidade para cada um dos elementos do conjunto de dados, e uma heurística de classificação é definida para obtenção da classificação final.

Uma coleção de índices é proposta para medir a acurácia, confiança e utilidade de uma classificação imprecisa para uma determinada aplicação. Os índices de acurácia são derivados das partições do conjunto em elementos bem classificados, mal classificados, precisos, parcialmente imprecisos e completamente imprecisos. Estas partições são obtidas em função da cardinalidade associada a cada elemento, definida como sendo o número de classes distintas associadas a cada elemento. Deste modo, preciso é o elemento com uma única classe associado a ele, e completamente impreciso é o elemento que tem associado a ele todas as classes. Já os índices de confiança se baseiam nos valores de possibilidade da distribuição associada a cada elemento. A utilidade é a medida de desempenho da classificação imprecisa derivada dos índices de acurácia, que fornecem a quantidade de acerto preciso e impreciso para os elementos associados a cada uma das classes.

Diferentes aplicações são utilizadas na avaliação da classificação imprecisa. É feita a classificação de imagens médicas e de sensoriamento remoto, dados que representam o registro de queda na distribuição de energia em subestações elétricas e dados relacionados à distribuição da esquistossomose no estado de Minas Gerais. O banco de imagens médicas é composto por imagens de RM reais e sintéticas do banco de dados BrainWeb, disponíveis em (<http://www.bic.mni.mcgill.ca/brainweb/>) BrainWeb (2006). As imagens de sensoriamento remoto são do tipo CBERS (2004) disponíveis em (<http://www.cbers.inpe.br>) e LANDSAT (2004) disponíveis em (<http://www.sat.cnpm.embrapa.br/satelite/landsat.html>). Os dados das redes elétricas são registros de queda na distribuição de energia elétrica, também estudados em Llanos *et al.* (2003b) e Llanos *et al.* (2003a); e os dados relacionados à esquistossomose são provenientes do Projeto GeoSchisto (2007), mais informações podem ser encontradas em (<http://www.dpi.inpe.br/geoschisto/>). Trabalha-se com imagens em 2 e 3 bandas, ou seja, seus pixels são caracterizadas por vetores de atributos. No caso das bases de dados, os elementos são também caracterizados por diferentes atributos que formam espaços multidimensionais.

Diante da importância da atividade de classificação em diferentes sistemas e diferentes áreas de pesquisa, entre questões como quando e como empregar os classificadores difusos e ainda do fato de que a imprecisão é característica intrínseca a qualquer conjunto de informação, é pertinente a proposta de novas estratégias, além da combinação de técnicas existentes na obtenção da melhor classificação para uma aplicação. É ainda de grande valia o estudo do desempenho da classificação imprecisa através de índices que levam em conta uma classificação correta, porém imprecisa, em lugar do risco de uma classificação precisa e incorreta.

Tanto a estratégia de classificação proposta, quanto a abordagem de avaliação da classificação imprecisa, se mostram aplicáveis a diferentes conjuntos de dados, e com resultados competitivos quando comparados a abordagens existentes na literatura. Espera-se contribuir com novas abordagens que sejam aplicáveis tanto ao desenvolvimento de sistemas reais, como módulos de um sistema de diagnóstico e acompanhamento médico, que envolvam diretamente a etapa de classificação; quanto na melhoria de sistemas existentes, fornecendo uma ferramenta que viabilize o estudo da imprecisão proveniente de um classificador difuso.

A obtenção da classificação imprecisa a partir de uma classificação possibilista causa perda de informação, porém, facilita a combinação de classificadores. A agregação direta de distribuições pode não ser favorável enquanto tiverem significados distintos, o que é resolvido pela noção de ordem imposta pela distribuição que caracteriza o classificador impreciso.

Este documento é organizado da seguinte maneira:

- No Capítulo 2 são apresentadas algumas das características dos conjuntos difusos e da teoria de possibilidades, enfocando-se particularmente os aspectos mais relevantes para este trabalho.
- O Capítulo 3 descreve os principais métodos de classificação abordados na literatura. São apresentadas as abordagens difusas e não difusas, destacando-se ainda os algoritmos de agrupamento. Além dos métodos em si, são feitas as considerações sobre medidas de desempenho de um classificador e como é realizada a combinação de classificadores, estratégia que busca por classificadores com melhor desempenho através da combinação de diferentes técnicas.
- O Capítulo 4 apresenta estratégias de classificação possibilistas. É descrita a proposta do modelo baseado em agrupamento. Além disso, são definidas

também as abordagens do meta-agrupamento, as relações de similaridade empregadas, e seus parâmetros.

- O Capítulo 5 refere-se ao estudo da avaliação de classificadores imprecisos através de uma coleção de índices de classificação, e gráficos de incerteza e imprecisão, analisados para uma família de classificadores imprecisos. É definida ainda a combinação de classificadores efetuada pela agregação de distribuições provenientes de diferentes técnicas; e as heurísticas de decisão implementadas.
- No Capítulo 6 são feitas as considerações sobre os experimentos realizados, com a análise de desempenho da abordagem proposta e, para algumas aplicações, é feita a comparação com outro método difuso definido na literatura.
- Ao final, no Capítulo 7, são apresentadas as principais conclusões e propostas de trabalhos futuros.

CAPÍTULO 2

ELEMENTOS DA TEORIA DOS CONJUNTOS DIFUSOS

O objetivo deste capítulo é apresentar os fundamentos básicos da teoria dos conjuntos difusos e os aspectos relevantes para este trabalho.

O conceito de conjunto difuso e algumas idéias básicas sobre operadores destes conjuntos foram primeiramente introduzidos por Lotfi Zadeh em 1965 (ZADEH, 1965). A teoria dos conjuntos difusos lida com informação caracterizada pela imprecisão. O emprego desta teoria é claramente adequada para modelar classes ou conjuntos de objetos do mundo real, que são caracterizados por atributos definidos de forma incerta, imprecisa ou por outro tipo de imperfeição.

A inexatidão dessas classes é a base dos conjuntos difusos. Um conjunto difuso é uma classe de objetos com graus de pertinência contínuos. Na teoria clássica dos conjuntos, um elemento tem pertinência binária (0 ou 1), ou seja, pertence ou não a um determinado conjunto. Já um conjunto difuso considera a possibilidade de pertinência parcial. Conseqüentemente, conjuntos difusos podem ser vistos como uma generalização de conjuntos clássicos para situações em que os limites da classe não podem ser definidos.

2.1 Conjunto difuso e função de pertinência

Um conjunto difuso A , em um universo U , é definido por uma função de pertinência $A(x) : U \rightarrow [0, 1]$ mapeando os valores do domínio no intervalo dos reais em $[0, 1]$. A função de pertinência $A(x)$ indica o grau de compatibilidade entre x e o conceito expresso por A :

- $A(x) = 1$, x é completamente compatível com A ;
- $A(x) = 0$, x é completamente incompatível com A ;
- $0 < A(x) < 1$, x é parcialmente compatível com A , com grau $A(x)$;

Algumas das funções de pertinência mais usadas são apresentadas na Figura 2.1 e suas fórmulas são dadas pelas Equações 2.1, 2.2 e 2.3, respectivamente, para os conjuntos difusos $A_{Triangular}$, $A_{Trapezoidal}$ e $A_{Gaussiano}$:

$$A_{Triangular}(x) = \begin{cases} \frac{x-a}{b-a}, & \text{se } x \in (a, b) \\ \frac{d-x}{d-b}, & \text{se } x \in (b, d) \\ 0, & \text{senão} \end{cases} \quad (2.1)$$

$$A_{Trapezoidal}(x) = \begin{cases} \frac{x-a}{b-a}, & \text{se } x \in (a, b) \\ 1, & \text{se } x \in (b, c) \\ \frac{d-x}{d-c}, & \text{se } x \in (c, d) \\ 0, & \text{senão} \end{cases} \quad (2.2)$$

$$A_{Gaussiano}(x) = \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}. \quad (2.3)$$

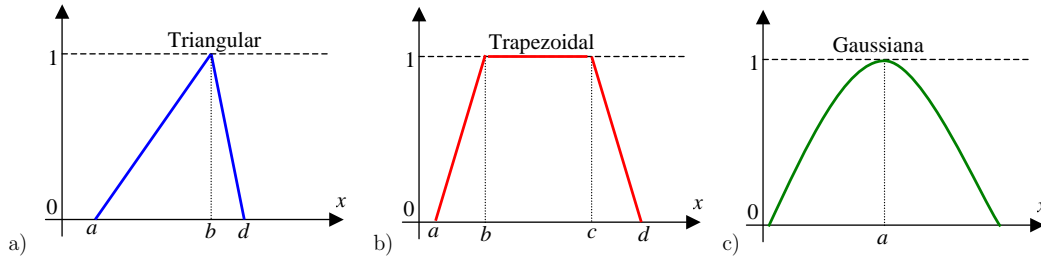


Figura 2.1 - Funções de pertinência: a) Triangular; b) Trapezoidal; c) Gaussiana.

2.2 Termos básicos

Seja um conjunto difuso A definido em U , com função de pertinência $A(x) : U \rightarrow [0, 1]$:

- O corte de nível α de um conjunto A é o conjunto clássico A_α . Este conjunto contém todos os elementos do conjunto universo com graus de pertinência em A maiores ou iguais a um valor especificado $\alpha \in [0, 1]$, conforme a equação:

$$A_\alpha = \{x \in U / A(x) \geq \alpha\} \quad (2.4)$$

- O suporte de A , $Su(A)$ agrupa elementos de U que são de alguma forma compatíveis com o conceito expresso por A :

$$Su(A) = \{x \in U / A(x) > 0\} = \lim_{\alpha \rightarrow 0} A_\alpha \quad (2.5)$$

- O núcleo de A , corresponde ao conjunto $Nu(A)$ formado pelos elementos que tem grau de pertinência igual a 1, ou seja, pelos elementos de U que são completamente compatíveis com o conceito expresso por A :

$$Nu(A) = \{x \in U / A(x) = 1\} = A_1 \quad (2.6)$$

- A altura de A é definida como o supremo da função $A(x)$:

$$Al(A) = \sup_{x \in U} A(x) \quad (2.7)$$

E se $Al(A) = 1$, tem-se um conjunto difuso *normalizado*.

- A cardinalidade do conjunto difuso A é expressa como:

– Para o universo de discurso discreto:

$$card(A) = \sum_{x \in U} A(x) \quad (2.8)$$

– Para o universo de discurso contínuo:

$$card(A) = \int_U A(x) \quad (2.9)$$

Um conjunto difuso A é *convexo* se, e somente se,

$$\forall x, y \in X \quad \forall \lambda \in [0, 1] : A(\lambda \cdot x + (1 - \lambda) \cdot y) \geq \min(A(x), A(y)) \quad (2.10)$$

Como exemplo de aplicação, os conjuntos convexos são normalmente utilizados para representar números difusos.

Na maior parte das aplicações, um conjunto difuso A pode ser representado utilizando a notação $\langle a, b, c, d \rangle$, onde $Su(A) = [a, d]$ e $Nu(A) = [b, c]$ e as funções entre a e b e d e c são estritamente monotônicas (crescente e decrescente, respectivamente). Este tipo de conjunto será chamado aqui de *conjunto difuso padrão*. Quando a função de pertinência é linear por partes, o conjunto difuso é chamado de trapezoidal, como o conjunto representado na Figura 2.1b. Um conjunto difuso triangular é denotado simplifiadamente por $\langle a, b, d \rangle = \langle a, b, b, d \rangle$.

2.3 Operações

A conjunção e a disjunção de conjuntos difusos são obtidas utilizando-se os conceitos de t-norma e t-conorma, respectivamente. As t-normas constituem uma família de operadores $\top : [0, 1] \times [0, 1] \rightarrow [0, 1]$ tais que, para quaisquer $x, y, z, w \in [0, 1]$, \top satisfaz as seguintes propriedades:

- **Comutatividade:** $x \top y = y \top x$
- **Associatividade:** $(x \top y) \top z = x \top (y \top z)$
- **Monotonicidade:** $x \top w \leq y \top z$, se $x \leq y, w \leq z$
- **Elemento neutro = 1:** $x \top 1 = x$

As t-conormas \perp , da mesma forma, são funções $\perp : [0, 1] \times [0, 1] \rightarrow [0, 1]$ comutativas, associativas e monotônicas. A diferença é que neste caso o elemento neutro é igual a 0, ou seja $x \perp 0 = x$. Em consequência destas propriedades tem-se $x \top 0 = 0$ e $x \perp 1 = 1$ (elemento absorvente).

Estes operadores são utilizados para implementar a união e a intersecção de conjuntos difusos. Por exemplo, pode-se fazer a união de dois conjuntos A e B utilizando a t-conorma max, $A \cup B(x) = \max(A(x), B(x))$ e a intersecção utilizando a t-norma min, $A \cap B(x) = \min(A(x), B(x))$.

Uma t-norma \top e uma t-conorma \perp são duais em relação a uma operação de negação $\neg : [0, 1] \rightarrow [0, 1]$ se elas satisfazem as relações de De Morgan, dadas por, para $\forall a, b \in [0, 1]$:

$$\begin{aligned} \neg(\top(a, b)) &= \perp(\neg a, \neg b) \\ \neg(\perp(a, b)) &= \top(\neg a, \neg b) \end{aligned} \tag{2.11}$$

O principal operador de negação é dado por $\neg a = 1 - a$, mas outros operadores podem ser encontrados na literatura (KLIR; FOLGER, 1988).

As t-normas e t-conormas mais utilizadas são mostradas na Tabela 2.1. Estes operadores satisfazem as leis de De Morgan em relação ao operador de negação $\neg a = 1 - a$.

Tabela 2.1 - Principais t-normas e t-conormas.

t-norma	t-conorma	nome
$\min\{a, b\}$	$\max\{a, b\}$	min / max
ab	$a + b - ab$	produto / soma probabilística
$\max\{0, a + b - 1\}$	$\min\{1, a + b\}$	diferença limitada / soma limitada
$\begin{cases} a, & \text{se } b = 1 \\ b, & \text{se } a = 1 \\ 0, & \text{senão} \end{cases}$	$\begin{cases} a, & \text{se } b = 0 \\ b, & \text{se } a = 0 \\ 1, & \text{senão} \end{cases}$	produto drástico / soma drástica

Outros operadores caracterizam operações de agregação na teoria dos conjuntos difusos. Os operadores que obedecem à propriedade da comutividade são chamados simétricos e consistem na classe mais importante das t-normas (normas triangulares). A média não ponderada é um operador deste tipo. Os operadores assimétricos são também definidos, sendo a média ponderada, o exemplo mais conhecido.

Dentre os operadores simétricos são definidos os operadores de média, que segundo Dubois e Prade (1988), são mapeamentos $m : [0, 1]^2 \rightarrow [0, 1]$, tais que:

- $\min(x, y) \leq m(x, y) \leq \max(x, y), \forall x, y, m \neq \{\min, \max\}$
- $m(x, y) = m(y, x)$
- m é não decrescente em cada argumento

Em particular, toda média estritamente crescente é não associativa. Diferentes famílias de médias são definidas, dentre elas, aquela em que se definem as médias harmônica, geométrica e aritmética e a família das medianas. Destaca-se neste trabalho uma outra família de médias, criada por Yager (1993), onde estão definidos os operadores OWA (*Ordered Weighted Average*). Um operador OWA de dimensão n é um mapeamento $f : R^n \rightarrow R$ associado a um vetor $w = [w_1, \dots, w_n]^T$ que é tal que $\sum_{i=1}^n w_i = 1, w_i \in [0, 1]$ e $f(a_1, \dots, a_n) = \sum_{j=1}^n w_j b_j$, onde b_j é o j -ésimo maior a_i . Ou seja, w_i não é necessariamente associado ao argumento x_i de f , mas ao i -ésimo maior dos argumentos de f , o que faz com que a operação de reordenação seja fundamental.

Os operadores OWA podem ser vistos como uma família parametrizada de médias, onde o parâmetro é um vetor de valores de R . Alguns vetores representativos são:

- Para $w^* = [1 \ 0 \ 0 \ \dots \ 0]^T, F^*(a_1, \dots, a_n) = \max_i a_i$
- Para $w^* = [0 \ 0 \ 0 \ \dots \ 1]^T, F^*(a_1, \dots, a_n) = \min_i a_i$

- Para $w_m = [1/n \ 1/n \ \dots \ 1/n]^T$, $F_m(a_1, \dots, a_n) = 1/n \sum_i a_i$
- Para $w^* = [0 \ 1/(n-2) \ \dots \ 1/(n-2) \ 0]^T$, $F^*(a_1, \dots, a_n) =$ média utilizada nas competições olímpicas, onde as primeiras maior e menor notas são descartadas.

Os operadores OWA são comutativos, monotônicos e, de maneira geral, não são associativos nem bissimétricos.

2.4 Relações difusas

As relações difusas são conjuntos difusos definidos em um universo multidimensional ($X_1 \times \dots \times X_n$). Como o próprio nome indica, uma relação implica na presença ou ausência de associação entre elementos de diferentes conjuntos. Na lógica difusa, o valor da função de pertinência das relações difusas indica o grau de associação entre os elementos (DUBOIS *et al.*, 1998).

Uma relação difusa R pode ser definida como um conjunto difuso, onde a cada ênupla (x_1, \dots, x_n) está associado um grau de pertinência entre 0 e 1. R é uma relação binária se definida em $X_1 \times X_2$. Tais relações são classificadas com base nas propriedades que apresentam.

Se R_1 e R_2 são relações difusas, respectivamente em $X \times Y$ e em $Y \times Z$, a composição de R_1 e R_2 , denotada por $R_1 \circ R_2$, resulta em uma nova relação que associa diretamente X a Z . A composição $sup\star$ pode ser definida da seguinte forma:

$$R_1 \circ R_2(x, z) = sup_{(y \in Y)} [R_1(x, y) \star R_2(y, z)] \quad (2.12)$$

onde $x \in X$, $z \in Z$ e \star é uma t-norma. A composição mais usual é a *sup-min*.

2.4.1 Relações de similaridade

As relações de similaridade, um tipo particular de relações difusas, foram originalmente introduzidas por Zadeh (1971) como uma generalização da definição clássica de relações de equivalência.

Basicamente, pode-se dizer que uma relação binária difusa S no universo U é uma relação de similaridade Ruspini *et al.* (1998), Dubois *et al.* (1998) em U se as seguintes propriedades forem satisfeitas:

$$S(x, x) = 1 \quad \text{reflexividade} \quad (1)$$

$$S(x, y) = S(y, x) \quad \text{simetria} \quad (2)$$

$$S(x, y) \odot S(y, z) \leq S(x, z) \quad \text{transitividade} \quad (3)$$

para todo $x, y, z \in X$, onde \odot é uma t-norma.

Quando somente a reflexividade e a simetria são obedecidas, a relação é chamada *relação de proximidade*. No entanto, na literatura frequentemente se mantém o termo *relação de similaridade*, mesmo que a transitividade não seja satisfeita. Neste trabalho, emprega-se o termo relação de similaridade sem a propriedade de transitividade.

Sejam S_1 e S_2 duas relações de similaridade em um domínio U . S_1 é dita ser maior ou igual que S_2 se, para qualquer a, b em U , $S_1(a, b) \geq S_2(a, b)$.

Tem-se que um conjunto de relações de similaridade em U formam um reticulado ordenado. O topo deste reticulado é a maior similaridade, S_{\top} , que faz todos os elementos serem completamente similares: $S_{\top}(v, v') = 1$ para todo $v, v' \in U$, não sendo possível distinguir um elemento do outro. A base deste reticulado S_{\perp} compreende a clássica relação de igualdade: $S_{\perp}(v, v') = 1$ se $v = v'$, senão, $S_{\perp}(v, v') = 0$, ou seja, é a relação que faz com que um elemento tenha grau de similaridade 1 consigo mesmo e 0 com os demais elementos (DUBOIS *et al.*, 1998).

2.5 Teoria de possibilidades

2.5.1 Fundamentos

A teoria de possibilidades é baseada na teoria dos conjuntos difusos e foi também proposta por Zadeh (YAGER *et al.*, 1987). É um modelo que permite trabalhar com conhecimento incerto ou impreciso. A informação fornecida por uma fonte de conhecimento sobre o verdadeiro valor de uma variável ω em um universo de discurso Ω é codificado sob a forma de uma distribuição de possibilidade $\pi : \Omega \rightarrow [0, 1]$. Para qualquer valor de ω_i em Ω , $\pi(\omega_i)$ reflete até que ponto é possível que $\omega = \omega_i$. Estes valores são supostamente mutuamente exclusivos, pois ω assume somente um valor (seu valor verdadeiro), que pertence a um conjunto universo Ω dado. Normalmente, supõe-se também que existe ao menos um valor considerado como completamente possível de ser o verdadeiro valor ω . Isto se traduz pela condição de normalização: $\exists \omega_i \in \Omega, \pi(\omega_i) = 1$.

A teoria dos conjuntos difusos e a teoria de possibilidades são intimamente ligadas. Isto é muito importante no sentido de que é possível se tratar tanto a imprecisão quanto a incerteza de um conjunto de informações em um único ambiente formal. De fato, a

maior parte do tempo não é necessário fazer a distinção entre um conjunto difuso e uma distribuição de possibilidade. Assim a intersecção ou união de duas distribuições de possibilidade é feita como entre conjuntos difusos, empregando as t-normas e t-conormas, como na Seção 2.3.

As medidas de possibilidade e necessidade, denotadas respectivamente por Π e N , são definidas da seguinte maneira:

$$\begin{aligned}\Pi(A) &= \sup_{\omega \in A} \pi(\omega) \\ N(A) &= \inf_{\omega \notin A} (1 - \pi(\omega))\end{aligned}\tag{2.13}$$

onde $\neg A$ é o complemento de A em Ω . A medida de possibilidade Π quantifica o quanto a evidência disponível não contradiz a hipótese de que A contém o verdadeiro valor de ω , e a necessidade N quantifica o quanto a evidência suporta esta hipótese. No caso em que π é dado pela função característica de um conjunto clássico B , um evento A é possível se, e somente se, $A \cap B \neq \emptyset$, e certo se $B \subseteq A$; e neste caso tem-se $\Pi(A) = N(A) = 1$.

A partir das expressões apresentadas na Equação 2.13 tem-se que $N(A) = 1 - \Pi(\neg A)$, o que exprime que A é tanto mais certo quanto mais $\neg A$ é impossível.

Para um conjunto difuso A , as medidas de possibilidade e necessidade, dada uma distribuição de possibilidade π , são dadas por:

$$\begin{aligned}\Pi(A) &= \sup_{\omega \in \Omega} \min(A(\omega), \pi(\omega)) \\ N(A) &= \inf_{\omega \in \Omega} \max(A(\omega), 1 - \pi(\omega))\end{aligned}\tag{2.14}$$

Dada uma distribuição de possibilidade π em Ω , para todos os conjuntos difusos A e B em Ω , tem-se:

$$\begin{aligned}
\Pi(\Omega) &= N(\Omega) = 1 \\
\Pi(\emptyset) &= N(\emptyset) = 0 \\
\Pi(A \cup B) &= \max(\Pi(A), \Pi(B)) \\
\Pi(A \cap B) &= \min(N(A), N(B))
\end{aligned}
\tag{2.15}$$

Uma distribuição de possibilidade pode ser vista como um conjunto finito de subconjuntos (focais) aninhados $\{A_1, A_2, \dots, A_n\}$, desde que $\{\pi(\omega) | \omega \in \Omega\}$ seja finito. Formalmente, existe um conjunto de pesos p_1, p_2, \dots, p_n , cuja soma é igual a 1, tal que (DUBOIS; PRADE, 1988):

$$\forall \omega, \pi(\omega) = \sum p_i / \{\omega \in A_i\} \tag{2.16}$$

Se o conjunto de possibilidades é $\{\alpha_1 = 1, \alpha_2 \geq \alpha_3 \dots \geq \alpha_n\}$, fazendo $\alpha_{n+1} = 0$ obtém-se:

$$A_i = \{\omega | \pi(\omega) \geq \alpha_i\}; p_i = \alpha_i - \alpha_{i+1}, \forall_i \tag{2.17}$$

Os intervalos A_i , são cortes de nível α_i , como definido na Seção 2.2.

2.5.2 Transformação da distribuição de possibilidade em probabilidade e vice-versa

A transformação de medidas de possibilidade em medidas de probabilidade, destacadas por Dubois *et al.* (1993) e Dubois e Prade (1988), se apresenta útil em problemas que envolvem dados heterogêneos com certo grau de incerteza e imprecisão. Entretanto, os modelos probabilistas e possibilistas consistem em formas distintas de representação da informação imperfeita. A representação possibilista é mais fraca devido à maneira explícita para manipulação da imprecisão (como, por exemplo, o conhecimento incompleto), além da sua representação ser baseada numa seqüência estruturada e não aditiva como na representação probabilista. Converter uma distribuição de possibilidade em probabilidade pode ser interessante no escopo da tomada de decisão, enquanto que a transformação de probabilidades em possibilidades pode ser útil quando a fonte de informação ou a computação com possibilidade são escassas.

Segundo Dubois *et al.* (1993), as transformações devem ser guiadas por dois diferentes

princípios de informação: o princípio da razão insuficiente da possibilidade para probabilidade, e o princípio da especificidade máxima da probabilidade para possibilidade. O primeiro garante encontrar uma medida de probabilidade que preserve a incerteza na escolha dos resultados, enquanto que o segundo garante encontrar mais informação na distribuição de possibilidade, baseado nos limites ditados pelo princípio da consistência entre probabilidade/possibilidade de Zadeh (1978), expresso por:

$$\forall A, N(A) \leq P(A) \leq \Pi(A) \quad (2.18)$$

As transformações empregadas neste trabalho são apresentadas por Dubois *et al.* (1993). No caso discreto, seja π_i o valor da distribuição no ponto x_i , $i = 1, n$, ou seja, $\pi_i = \pi(x_i)$. Denota-se por p_i o valor de probabilidade obtida para o ponto x_i com a aplicação da transformação. Supondo, sem perda de generalidade, que os valores da distribuição de possibilidade estão ordenados em ordem decrescente, sendo pois $\pi_1 = 1$. Note que se a distribuição não estiver ordenada basta ordená-la convenientemente. Sejam $\pi_{i+1} = 0$ e $p_{i+1} = 0$. Tem-se a seguinte transformação possibilidade-probabilidade:

$$p_i = p_{i+1} + (\pi_i - \pi_{i+1}) / |A_i| \quad (2.19)$$

onde $A_i = \{x \mid \pi(x) \geq \pi_i\}$.

Para a transformação probabilidade-possibilidade, seguindo as mesmas considerações, a transformação de valores de probabilidade em valores de possibilidade é tal que p_i é o valor da distribuição no ponto x_i , $i = 1, n$. Se $X = \{x_1, x_2, \dots, x_n\}$, $p_i = P(\{x_i\})$, $\pi_i = \Pi(\{x_i\})$ e $p_1 \geq p_2 \geq \dots \geq p_n$, então:

$$\forall i = 1, \dots, n, \pi_i = \sum_{j=i}^n p_j \quad (2.20)$$

Como conseqüência do método, para o menor valor tem-se $\pi(x) = p(x)$ e para o maior $\pi(x) = 1$.

CAPÍTULO 3

A TAREFA DE CLASSIFICAÇÃO

O reconhecimento de padrões lida com problemas de classificação em que se busca por estruturas dentro de um conjunto de objetos, onde cada objeto é descrito por um conjunto ρ de atributos e pode ser entendido como um ponto num espaço ρ -dimensional. Segundo Kuncheva e Whitaker (2005), um classificador é uma fórmula, algoritmo ou técnica que aplicado a um conjunto de valores de entrada, atribui uma classe a cada valor como saída. A saída de um classificador pode ser tanto a seleção discreta de uma das classes pré-definidas, ou um vetor de valores expressando a semelhança do dado original com os padrões correspondentes.

Um classificador é definido como um mapeamento $Cl : R^n \rightarrow \mathbf{C}$ (KUNCHEVA, 2004), onde \mathbf{C} representa o conjunto de classes. No modelo canônico, existem m funções discriminantes $g_i : R^n \rightarrow R$, cada qual resultando na pontuação referente à classe $c_i \in \mathbf{C}$. Geralmente, $\forall z \in R^n, Cl(z) = \arg \max_i g_i(z)$, ou seja, o classificador atribui a z a classe com maior pontuação.

Existem essencialmente duas abordagens na classificação: supervisionada e não supervisionada, distintas em função da presença ou ausência de uma fase de treinamento, onde amostras orientam o processo de classificação. Na classificação supervisionada, tem-se de antemão algum conhecimento sobre as classes presentes no conjunto de dados considerado, ou seja, são definidas as amostras das classes e o classificador é treinado, o que torna possível que este seja capaz de localizar todos os demais elementos pertencentes às classes pré-definidas. Já a técnica não supervisionada é utilizada quando não se tem nenhum tipo de conhecimento prévio e baseia-se numa análise de agrupamento onde são identificados os grupos ou classes formando padrões similares no espaço de atributos. Os classificadores não supervisionados são também referenciados como algoritmos de agrupamento (DUDA *et al.*, 2000; LEONDES, 1998).

Nas seções que se seguem, são descritos os principais métodos de classificação apresentados na literatura, bem como a combinação de classificadores e a avaliação do desempenho de um modelo de classificação.

3.1 Definições

Diante dos inúmeros métodos e aplicações apresentados na literatura, é natural que a terminologia seja igualmente vasta e muitas vezes contraditória. Os termos a seguir são empregados na definição dos métodos de classificação apresentados neste trabalho:

- **Dado, elemento, objeto, ponto ou padrão:** denotado por z , é um vetor de tamanho ρ cujos valores z_i na i -ésima posição estão relacionados a um atributo A_i definido no domínio Ω_i .
- **Espaço de atributo:** exprimido por Ψ com dimensão igual a ρ , que é o tamanho de cada vetor que compõe o espaço.
- **Classe:** denotada por c_i .
- **Conjunto de classes:** denotado por $\mathbf{C} = \{c_1, \dots, c_m\}$ com m classes distintas;
- **Conjunto de dados:** denotado por $Z = \{z_1, \dots, z_n\}$, onde $z_i \in R^\rho$.
- **Classe de um dado:** a classe de z é um dos elementos do conjunto \mathbf{C} , denotado por $c^*(z) \in \mathbf{C}$ ou, quando possível, simplesmente por $c^* \in \mathbf{C}$.

Exemplos de padrões são medidas feitas a partir de um paciente com o objetivo de diagnosticar uma doença; medidas sobre variáveis temporais para previsão do tempo ou pixels de uma imagem digital para reconhecimento de caracteres, lesões em imagens médicas ou áreas de desmatamento em imagens de sensoriamento remoto. Já os atributos são as características como a temperatura e a pressão de um paciente, que em determinadas faixas de valores evidenciam uma doença.

3.2 Modelos de classificação

Diferentes taxonomias podem ser identificadas, não havendo um consenso quanto a qual grupo um método se insere. Não se pretende neste trabalho adotar ou criar uma taxonomia de classificadores. Opta-se por apresentar os métodos principais e relevantes para o estudo, detalhando suas características e aplicações de maior impacto.

Vale ressaltar os modelos de classificação que empregam a lógica difusa. O conceito de conjunto difuso foi primeiramente estudado em problemas relacionados à classificação de padrões em Bellman *et al.* (1966). Em Leondes (1998) é empregado o termo *percepção artificial* para reconhecimento de padrões, sendo um processo que se apresenta como um importante aspecto da percepção humana e está naturalmente ligado à imprecisão presente na informação.

Diferentes definições de classificadores difusos são apresentadas por Kuncheva (2000). A primeira delas diz que um classificador difuso é qualquer classificador que emprega conjuntos difusos na sua fase de treinamento ou na sua fase de operação. Para a segunda definição, um classificador difuso ou probabilista é qualquer classificador possibilista

para o qual o somatório dos graus de pertinência do elemento para cada classe é igual a 1. E por fim, a terceira definição assume que um classificador difuso é um sistema de inferência difuso *se-então* que atribui a cada elemento uma classe pura ou composta.

A seguir são descritos os modelos de classificação mais empregados na literatura, e aqueles nos quais se baseiam as técnicas propostas neste trabalho de tese.

3.2.1 Algoritmos de agrupamento

Um algoritmo de agrupamento busca classificar uma massa de dados em um número determinado de grupos onde os elementos de cada grupo deveriam ser similares, sendo dissimilares aos elementos dos outros grupos. Isto implica na existência de uma medida de distância ou similaridade entre os dados, sendo o número de grupos pré-definido ou tido por consequência de alguma restrição imposta sobre os dados. Como saída, um algoritmo de agrupamento apresenta cada elemento pertencendo a um ou mais grupos, dependendo dos fundamentos que regem o algoritmo empregado. Além disso, a forma dos grupos difere de um método para outro. Assim, diferentes técnicas de agrupamento são tradicionalmente dispostas em duas categorias: agrupamento hierárquico e agrupamento por partição. Além destas, técnicas como redes neurais e algoritmos evolutivos vem sendo empregadas no agrupamento de dados. Já o emprego dos conjuntos difusos é tido como uma extensão da abordagem tradicional, sendo comum na literatura a divisão dos algoritmos de agrupamento em rígidos e difusos (JAIN *et al.*, 1999; SILVA, 2003).

Seja um conjunto de n objetos $Z = \{z_1, z_2, \dots, z_n\}$, onde cada $z_i \in \Psi^\rho$ é um vetor de atributos composto de ρ componentes que descrevem o objeto. Na abordagem rígida, o problema de agrupamento consiste em atribuir cada objeto z a unicamente um dos grupos em $G = \{G_1, G_2, \dots, G_k\}$, onde $G_i \neq \emptyset$ e $k \neq 0$, é um agrupamento qualquer e k é o número de grupos. Os grupos são disjuntos e a união de todos os grupos em G resulta em Z , ou seja, $G_1 \cup G_2 \cup \dots \cup G_k = Z$, $G_i \cap G_j = \emptyset$ para $i \neq j$. Em geral, cada grupo G é representado por um centro C que pode corresponder ou não a um elemento de Z .

Uma vez que os objetos devem ser agrupados de acordo com a similaridade entre eles, é necessário definir métricas através das quais é possível medir quão similares são dois objetos ou como seus valores podem ser comparados. Uma medida empregada é a distância entre os objetos, que na verdade quantifica a dissimilaridade, uma vez que quanto maior a distância entre os objetos, menor a semelhança entre eles.

A medida de distância habitualmente empregada em aplicações é a distância Euclidiana, que mede a distância em linha reta entre dois pontos, z_i e z_j , que representam os objetos:

$$d(z_i, z_j) = \sqrt{\sum_{l=1}^{\rho} (z_{i,l} - z_{j,l})^2} \quad (3.1)$$

Esta métrica funciona bem quando o conjunto de dados apresenta grupos compactos ou isolados, pois assume que os valores dos atributos do ponto não são correlacionados uns com os outros. Dependendo da aplicação esta pode se tornar uma característica indesejável.

Uma alternativa é o emprego da distância de Mahalanobis:

$$d(z_i, z_j) = \sqrt{(z_i - z_j)^T A (z_i - z_j)} \quad (3.2)$$

onde $A = C^{-1}$ é a matriz de covariância. Se $A = I$, tem-se a distância Euclidiana, já definida.

Diferentes modelos de agrupamento de dados são definidos na literatura e segundo suas características podem ser divididos em categorias. As características consideradas mais relevantes são descritas a seguir:

- Definição do número de grupos a priori.
- Estratégia do algoritmo que descreve o modelo. Uma abordagem pode ser aglomerativa, partindo de um conjunto onde cada elemento pertence ao seu grupo e unindo os grupos até que um critério de parada seja satisfeito. Ou divisiva, onde o método inicia com todos os elementos pertencentes a um único grupo e se subdivide sucessivamente até um critério de parada (JAIN *et al.*, 1999).
- Pertinência de um elemento a um grupo. Um algoritmo de agrupamento difuso considera que um elemento pode pertencer a mais de um grupo, porém com diferentes graus de pertinência, gerando grupos que se sobrepõem. Já num agrupamento rígido não existe sobreposição de grupos pois cada elemento pertence a um único grupo.

- Os algoritmos que envolvem a otimização de uma função objetivo podem ser executados usando técnicas tradicionais ou através da busca aleatória num espaço composto de todas as possibilidades de agrupamento.

No agrupamento hierárquico, os elementos são classificados em etapas, de modo hierárquico. Como resultado obtém-se uma árvore de grupos chamada dendrograma (Figura 3.1), que representa o agrupamento dos objetos. Os métodos hierárquicos podem ser subdivididos em aglomerativos e divisivos. No primeiro caso inicia-se o processo com cada objeto como sendo um grupo, que sofrem fusões dependendo da similaridade. A cada estágio os mais similares são agrupados, e o processo se repete até que um único grupo seja formado. O método divisivo trabalha de maneira oposta, iniciando com um grupo que contém todos os objetos. O processo de divisão ocorre até que cada objeto tenha o seu grupo. Neste processo hierárquico os níveis da hierarquia podem indicar o número correto de grupos para determinada aplicação.

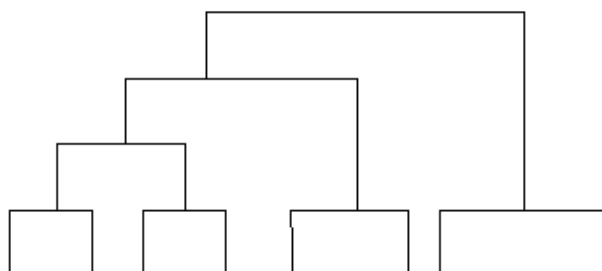


Figura 3.1 - Exemplo de um dendrograma.

As técnicas hierárquicas aglomerativas são ainda subdivididas em: distância mínima (single-linkage), distância máxima (complete-linkage) e distância média (average-linkage) que caracterizam a similaridade entre pares de grupos para que estes sejam combinados em um único grupo. Na distância mínima, a distância entre dois grupos corresponde à menor distância entre todos os pares de objetos entre os dois grupos considerados. Na distância máxima, a maior distância é considerada. E no terceiro caso é calculada uma média de todas as distâncias entre dois grupos (JAIN *et al.*, 1999; COLE, 1998).

O agrupamento por partição obtém diretamente uma partição do conjunto de dados através da otimização de funções objetivo, que podem ser definidas local ou globalmente, ou seja, sobre um subconjunto dos objetos ou sobre todos os objetos, respectivamente. Encontrar a solução ótima apresenta um custo computacional

proibitivo, pois requer uma busca exaustiva no conjunto de todas as partições possíveis. Na prática, os algoritmos são executados várias vezes empregando diferentes condições iniciais e a melhor configuração obtida é dada como partição de saída (JAIN *et al.*, 1999).

O modelo por partição, também chamado de modelo de otimização por se tratar de uma técnica de otimização, tem como função objetivo mais comumente empregada o erro quadrático.

3.2.1.1 O algoritmo k-médias

O método k-médias (*k-means*) proposto por McQuenn (1967) é um método de agrupamento usado em inúmeras aplicações. Este algoritmo se baseia na minimização da função do erro quadrático e o número k de grupos desejados deve ser fornecido na inicialização. Os passos do algoritmo são descritos a seguir:

- a) Determinar os k centros iniciais segundo algum critério, por exemplo, aleatoriamente ou a partir de uma heurística;
- b) Calcular a distância entre cada elemento e o centro de cada grupo;
- c) Atribuir cada elemento do conjunto ao grupo cujo centro está mais próximo, ou seja, de menor distância;
- d) Recalcular os centros de cada grupo a partir dos elementos realocados;
- e) Repetir os passos 2 a 4 até que algum critério de parada seja satisfeito.

É um algoritmo simples, iterativo, que sempre caminha no sentido da melhor solução dentro da vizinhança, mas que não apresenta garantia de que a solução encontrada seja ótima. Além disso, é altamente sensível às condições iniciais.

Variações deste algoritmo são encontradas na literatura, como Ball e Hall (1967), Diday (1973), Symon (1981) e mais recentemente Hamerly e Elkan (2002), Likas *et al.* (2003), algumas delas buscando a melhor inicialização do algoritmo com o propósito de se aproximar da solução ótima.

3.2.1.2 Fuzzy C-Means (FCM)

Os algoritmos de agrupamento difusos clássicos são baseados em uma função objetivo. A abordagem proposta por Ruspini (1969), onde graus de pertinência com relação a cada grupo são atribuídos aos objetos, apresenta vantagens notáveis sobre a representação

convencional de agrupamento. Os objetos, ou pontos do conjunto de dados que se encontram no centro de algum dos grupos apresentam grau de pertinência 1 com relação àquele grupo, enquanto que os pontos nas fronteiras apresentam um grau de pertinência que varia de acordo com a distância para o centro dos grupos.

Um exemplo simples de um agrupamento difuso é apresentado por Ruspini (1969) e está representado na Figura 3.2. Pode-se observar que o centro ou núcleo de cada um dos dois conjuntos é formado pelos pontos das colunas nas extremidades do conjunto. O primeiro valor dentro dos parênteses indica o grau de pertinência do ponto com relação ao grupo 1 e o segundo valor com relação ao grupo 2. Os graus de pertinência variam de acordo com a distância dos pontos aos centros de cada grupo.

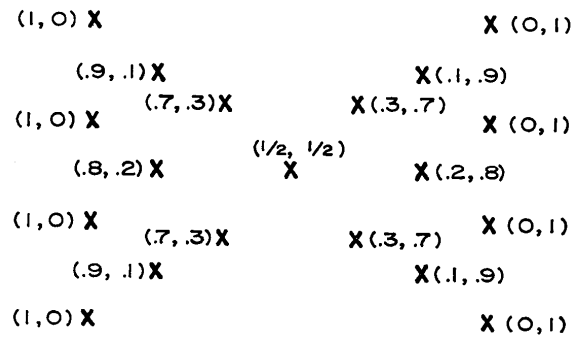


Figura 3.2 - Conjunto de dados classificado em dois grupos por agrupamento difuso.

Em termos matemáticos, o problema pode ser expresso como a partição de um conjunto Z , em G subconjuntos difusos (grupos). Esta partição pode ser representada pela matriz U , cujos elementos $u_{ik} \in [0, 1]$ representam os graus de pertinência de z_k com relação ao i -ésimo grupo.

Os algoritmos de agrupamento difuso visam usualmente minimizar uma função objetivo do tipo:

$$J(Z; U, T) = \sum_{i=1}^G \sum_{g=1}^N (u_{ig})^M d^2(z_g, t_i) \quad (3.3)$$

onde $T = \{t_1, \dots, t_N\}$ são os centros dos grupos, ou seja, existem tantos vetores t_i quantos grupos forem gerados, e $M \in (1, \infty)$ é o chamado coeficiente difuso, que vai ponderar o quanto o grau de pertinência influencia na métrica de distância d empregada.

A matriz U dos valores de pertinência apresenta as seguintes restrições:

$$\sum_{i=1}^G u_{ig} = 1, \forall g. \quad (3.4)$$

$$0 < \sum_{g=1}^N u_{ig} < N, \forall i. \quad (3.5)$$

Essas restrições fazem com que a soma de cada uma das colunas da matriz U seja 1. Diferentes métricas de distância dão origem a distintos algoritmos de agrupamento difuso, influenciando na forma dos grupos obtidos.

O algoritmo FCM de Bezdek *et al.* (1984) adota a distância euclidiana como medida de similaridade entre o ponto e o centro do grupo, e apresenta melhor desempenho quando se trabalha com grupos de forma esférica, separáveis e aproximadamente de tamanhos iguais. Basicamente, constitui uma extensão do algoritmo k-médias descrito na seção 3.2.1.1.

Segundo Bezdek *et al.* (1984) não existem evidências teóricas ou computacionais que definam o melhor valor do coeficiente difuso M , na Equação 3.3. Os valores utilizados estão no intervalo $[1, 30]$ e apenas experimentalmente é possível defini-lo. Para a maioria dos dados, $1.5 \leq M \leq 3$ apresenta bons resultados.

É possível observar que no algoritmo k-médias tem-se cada elemento pertencendo a apenas um grupo, resultando, portanto, em grupos disjuntos. Já no FCM cada elemento apresenta um grau de pertinência, segundo uma função, com relação a cada grupo. Desta forma, a saída do algoritmo é um agrupamento difuso, onde cada grupo é um conjunto difuso de todos os elementos.

3.2.1.3 O algoritmo GK - Gustafson Kessel

Para conjuntos de dados em que a forma dos grupos que se deseja obter é conhecida, é ideal adaptar a medida de distância aos dados, como no algoritmo Gustafson-Kessel (GK) (GUSTAFSON; KESSEL, 1978).

Fazendo $d^2(z_g, t_i) = (z_g - t_i)^T A_i (z_g - t_i)$ na Equação 3.3, tem-se $A_i = I$ para a distância euclidiana e conseqüentemente o algoritmo FCM, discutido na seção 3.2.1.2. Porém fazendo $A_i = C^{-1}$, onde C é uma matriz de covariância, tem-se a aplicação da distância de Mahalanobis e, desta maneira, o algoritmo GK se mostra capaz de determinar classes

elipsoidais de tamanhos e orientações diferentes.

3.2.1.4 O algoritmo FCV - *Fuzzy C-Varieties*

Originalmente, os algoritmos de agrupamento são aplicados a um conjunto de dados para detectar estruturas hiperesféricas. O algoritmo FCV (*Fuzzy C-Varieties*), proposto por Bezdek (1981), é um método de agrupamento desenvolvido para reconhecer linhas, planos e hiperplanos. Ou seja, ideal para agrupar conjuntos de dados cujos grupos apresentam formas não esféricas. A formulação do FCV se apresenta como a Equação 3.3, exceto pela distância d , que não é calculada para os centros c_i , mas para os planos C , que são variações lineares r -dimensionais, com $0 \leq r \leq p - 1$. Assim, é interessante observar que para o valor de $r = 0$ tem-se o algoritmo FCM (seção 3.2.1.2), para $r = 1$, tem-se o chamado FCL (*Fuzzy C-Lines*), para $r = 2$ tem-se o FCP (*Fuzzy C-Planes*) e para $r = p - 1$, o FCH (*Fuzzy C-Hyperplanes*), como em Bezdek (1981).

Grupos em forma de elipsóides podem ser detectados pelo FCE (*Fuzzy C-Elliptotypes*), em Bezdek *et al.* (1981), que consiste numa extensão do FCV, usando a combinação da função objetivo do FCV com a função objetivo do FCM.

3.2.1.5 Algoritmos de agrupamento difuso estendidos

Propostos por Kaymak e Setnes (2000), os algoritmos de agrupamento difuso estendidos, chamados aqui de algoritmos estendidos, a título de simplificação, se baseiam na extensão da função objetivo do agrupamento difuso (Equação 3.3), passando a considerar o volume dos grupos, além da união de grupos por similaridade. O objetivo da extensão é reduzir a sensibilidade do algoritmo com relação à distribuição dos dados e determinar o número de grupos automaticamente.

Considerando o volume, os pontos do espaço de agrupamento se transformam em regiões, e pode ser definido como sendo um subespaço n -dimensional, convexo e compacto deste espaço. O volume é limitado por um raio, que pode ser definido pelo usuário ou estimado a partir dos dados. A medida de distância empregada neste tipo de algoritmo é uma versão modificada da distância usada nos algoritmos de agrupamento. Primeiro calcula-se a distância d_{ik} do ponto x_k ao centro v_i do grupo. Em seguida, a distância \tilde{d}_{ik} para o volume \tilde{v}_i é determinada considerando o raio r_i , $\tilde{d}_{ik} = \max(0, d_{ik} - r_i)$. Assim, os pontos x_k que se encontram dentro da hiperesfera, $d(x_k, v_i) \leq r_i$ são elementos do volume \tilde{v}_i e apresentam por definição pertinência 1 com relação ao grupo. A forma de se definir os raios $r_i, i = 1, \dots, M$ é relacioná-los com o tamanho do grupo e isso pode ser obtido a partir da matriz de covariância.

O número de grupos é determinado inicializando o algoritmo com um número máximo de grupos e em seguida avaliando a similaridade entre os grupos. Grupos similares são unidos. Essa similaridade é avaliada considerando os grupos difusos no espaço de dados; se a similaridade é mais alta do que um limiar $\alpha \in [0, 1]$, os grupos mais similares são unidos em cada iteração do algoritmo.

Dentre os algoritmos estendidos destacam-se o E-FCM (*Extended Fuzzy C-Means*) e o E-GK (*Extended Gustafson-Kessel*) que consistem das versões estendidas do FCM e do GK, respectivamente. Estes algoritmos são descritos em Kaymak e Setnes (2000) juntamente com exemplos que ilustram suas propriedades.

Nos algoritmos estendidos a influência de diferentes distribuições e densidade dos dados no resultado do agrupamento é reduzida. Além disso, a inicialização do algoritmo com um número máximo de grupos, e posterior união dos grupos similares, reduz a dependência do algoritmo quanto à condição inicial aleatória.

3.2.1.6 Outros algoritmos de agrupamento

Silva (2003) apresenta um algoritmo de agrupamento difuso não supervisionado baseado em aprendizagem participativa que é capaz de determinar o número adequado de grupos e seus respectivos centros para um conjunto de dados Z . O algoritmo utiliza a distância de Mahalanobis como medida de similaridade e parâmetros de controle para o agrupamento não supervisionado de dados.

Em Belacel *et al.* (2002) é apresentada uma nova heurística de busca local, chamada FJM (*Fuzzy J-Means*) onde a vizinhança é definida através de todas as possíveis realocações dos centros. Esta solução “crisp” é transformada em difusa encontrando centros e graus de pertinência para todos os elementos e grupos. É um método que, assim como o FCM, pode ficar preso em mínimos locais, possivelmente distante do valor desejado.

Dentre as muitas técnicas apresentadas na literatura, cada qual com a sua particularidade, os algoritmos são em sua maioria, variações do FCM, que buscam suprir desvantagens com relação à sensibilidade a ruído, por exemplo, como o ε -FCM (*ε -insensitive Fuzzy C-Means*) proposto por Leski (2003). Além deste, Wu e Yang (2002) apresentam o AFCM (*Alternative Fuzzy C-means*) que se mostra mais robusto que o FCM e é usado com sucesso na segmentação de imagem de ressonância magnética de oftalmologia. Também baseado no FCM, Pedrycz (2002) apresenta uma nova arquitetura para agrupamento onde subconjuntos de padrões podem ser processados

de maneira simultânea e uma função objetivo busca pela estrutura comum, onde o algoritmo trabalha sobre subconjuntos de dados separados que colaboram entre si trocando informações. Em Yang e Wang (2004) são apresentadas abordagens competitivas para agrupamento de dados com ruído, que buscam solucionar problemas apresentados pelo algoritmo LVQ em determinadas aplicações.

Como é possível observar na literatura, uma grande variedade de algoritmos de agrupamento difuso de dados vêm sendo desenvolvidos, por se mostrarem simples e eficientes. Neste trabalho foram mostradas algumas das principais técnicas seguidas de citações de alguns algoritmos recentemente estudados.

3.2.2 Classificadores possibilistas

3.2.2.1 Abordagem possibilista para agrupamento

Em Krishnapuram e Keller (1993) é descrita uma abordagem de agrupamento cuja partição obtida para o conjunto de dados pode ser interpretada como uma partição possibilista, e os valores de pertinência associados podem ser interpretados como graus de possibilidade dos elementos para com suas classes. A diferença reside na construção da função objetivo cujo mínimo caracteriza uma boa partição do conjunto.

O algoritmo FCM, já descrito neste capítulo (vide seção 3.2.1.2), e a maioria dos métodos difusos de agrupamento, que são derivados do FCM, usa restrições probabilistas, sendo que a distribuição associada a um elemento tem soma igual a 1. Entende-se por distribuição associada a um elemento os graus de pertinência deste elemento com relação a cada uma das classes envolvidas no problema de classificação. As restrições probabilistas podem ser relaxadas para uma interpretação possibilista, ou as funções objetivo usadas no FCM podem ser modificadas, caracterizando uma família de algoritmos de agrupamento possibilista. Da Equação 3.3 com a restrição dada pela Equação 3.4, Krishnapuram e Keller (1993) definem a função:

$$J(Z; U, T) = \sum_{i=1}^G \sum_{g=1}^N (u_{ig})^M d^2(z_g, t_i) + \sum_{i=1}^G \eta_i \sum_{g=1}^N (1 - u_{ig})^M \quad (3.6)$$

onde η_i são números positivos adequados. O primeiro termo busca pela menor distância entre elemento e protótipo da classe, enquanto que o segundo termo faz com que u_{ig} seja o maior possível, impedindo a solução trivial. Na prática a definição que se segue é uma boa definição para η :

$$\eta_i = K \frac{\sum_{g=1}^N (u_{ig})^M d^2(z_g, t_i)}{\sum_{g=1}^N (u_{ig})^M} \quad (3.7)$$

Geralmente o valor 1 é escolhido para K . O valor de η_i pode ser fixo para todas iterações ou variar em cada uma delas. Krishnapuram e Keller (1993) mostram que o agrupamento final é praticamente insensível a grandes variações de η_i . Outras regras são empregadas e estudadas e as variações apresentam particularidades para aplicações específicas.

3.2.2.2 Classificador possibilista usando algoritmo genético

Uma abordagem supervisionada proposta para classificação de imagens digitais é apresentada em Malinski (2000). O classificador utiliza teoria de possibilidades para modelar a incerteza derivada da classificação dos pixels das amostras da imagem. Primeiramente é feita uma classificação não-determinística da imagem, para em seguida refinar esta classificação usando heurísticas contextuais de pós-processamento, que levam em consideração a classificação dos vizinhos de cada pixel. As distribuições de possibilidade são obtidas a partir dos histogramas das amostras das classes, que indicam o quanto cada nível de cinza é compatível com uma dada classe. Trata-se de um modelo híbrido, uma vez obtidas as distribuições de possibilidade, um algoritmo genético é aplicado sobre as amostras para determinar um limiar (grau de possibilidade) para cada classe, que é tal que se um nível de cinza tem grau de possibilidade, em relação àquela classe, superior ou igual ao limiar da classe, ele é considerado como possivelmente pertencente àquela classe. A classificação que se obtém empregando os limiares fornecidos pelo algoritmo genético é uma classificação em termos de classes puras e compostas, ou seja, uma classificação imprecisa. A classificação precisa é obtida submetendo-se o resultado obtido a um pós-processamento que leva em conta a vizinhança de cada pixel visando uma classificação final em termos de classes puras. O pós-processamento compreende duas fases, uma para eliminação das classes compostas e outra para refinamento.

Numa primeira instância, o modelo proposto faz uma classificação não-determinística dos pixels da imagem, sendo cada pixel atribuído não necessariamente a uma classe pura, mas eventualmente a uma classe composta. Basicamente o algoritmo de pós-processamento investigado por Malinski (2000) consiste em utilizar uma função que contabilize o número de classes puras contidas nas classes (eventualmente compostas) dos vizinhos de um dado pixel. A cada passo, o algoritmo reclassifica o pixel com classe igual ou maior do que um nível de indefinição ϕ , inicializado com a maior cardinalidade

presente na vizinhança. Para um nível de indefinição ϕ qualquer, o processo é repetido até que nenhum elemento tenha sua classe substituída. Um segundo algoritmo é definido para resolver os pixels com indefinição. Em última instância, se o pixel ainda pertencer a uma classe composta, é reclassificado drasticamente de forma aleatória.

No caso de classificação multibandas, duas estratégias são descritas em Malinski *et al.* (2001). Uma delas consiste em gerar as classificações imprecisas de forma independente para cada uma das dimensões, obtendo-se para cada distribuição de possibilidade uma classificação imprecisa. Estas classificações são então agregadas gerando uma classificação eventualmente imprecisa em uma única dimensão. A outra abordagem é gerar as distribuições de possibilidade em cada dimensão, agregá-las e só então gerar a classificação imprecisa.

3.2.3 Classificador Bayesiano

O emprego do formalismo bayesiano no projeto de um classificador fornece um método bem fundamentado na área de reconhecimento de padrões. A abordagem define o problema de decisão em termos probabilísticos em que todas as probabilidades relevantes são conhecidas. Supondo que as probabilidades *a priori* para cada classe são conhecidas e dadas por $P(c_i), i = 1, \dots, m$, a probabilidade *a posteriori* de um ponto z estar numa classe c_i é calculada pela regra de Bayes (DUDA *et al.*, 2000):

$$p(c_i|z) = \frac{p(z|c_i)P(c_i)}{\sum_{j=1}^m p(z|c_j)P(c_j)} \quad (3.8)$$

onde $p(z|c_i)$ é a probabilidade condicional de z , dada a classe c_i . A regra de Bayes mostra que através da observação de um valor z , é possível converter a probabilidade *a priori* $P(c_i)$ para a probabilidade *a posteriori* $P(c_i|z)$.

A forma mais usual de se classificar um objeto é considerar todas as probabilidades *a posteriori* e atribuir ao elemento a classe cujo valor de probabilidade é maior. Desta maneira, as probabilidades *a posteriori* são as funções discriminantes, também definidas como funções de decisão ou mesmo funções de classificação. Uma regra de decisão particiona o espaço de atributos em m regiões $C_i, i = 1, \dots, m$. Se um dado observado está em C_i então a classe deste dado é c_i .

A grande dificuldade do uso do classificador de Bayes é a necessidade de se conhecer antecipadamente as probabilidades relevantes, o que nem sempre é possível.

3.2.4 Classificador por Máxima Verossimilhança (MAXVER)

O classificador por Máxima Verossimilhança (MAXVER) é do tipo supervisionado. A partir do conjunto de amostras é definido um vetor média e a matriz de covariância que consiste na base para realizar o processo de classificação. Segundo Richards (1986), o MAXVER adota a modelagem das classes por uma distribuição de probabilidade no espaço multiespectral, usualmente descrita por parâmetros de uma distribuição gaussiana. Neste caso, cada classe é descrita por uma distribuição de probabilidade normal.

A classificação feita através do MAXVER pode ser descrita em duas etapas distintas: o treinamento, onde são selecionadas as amostras representativas de cada classe e então extraídos os parâmetros estatísticos (média e matriz de covariância, no caso gaussiano). Estes parâmetros são usados pelo classificador para criar uma regra de decisão que irá determinar a qual classe pertencerá um dado elemento dentro da área de estudo e fora das áreas amostradas; e a etapa de classificação onde é calculada a probabilidade de um dado elemento pertencer a cada uma das classes fornecidas ao sistema na fase de treinamento. O elemento será associado à classe com maior probabilidade de pertencer.

3.2.5 Classificador dos k vizinhos mais próximos (KNN)

A técnica do vizinho mais próximo atribui a cada elemento z do conjunto de dados Z o mesmo grupo ao qual pertence o seu vizinho mais próximo, de acordo com alguma métrica de distância ou similaridade. O processo é flexível, porém de alto custo computacional, devido à verificação da vizinhança. O algoritmo usa apenas as amostras de treinamento. Numa abordagem geral, a distância de um determinado elemento z para todos os elementos das amostras é calculada e z passa a pertencer ao mesmo grupo do ponto amostral que lhe é mais próximo. Encontrar a métrica de distância ou similaridade adequada para um dado conjunto de dados é um problema inerente à técnica.

Uma boa classificação pode ser feita examinando os k vizinhos mais próximos de cada elemento. O número k de vizinhos é dependente do número de classes que se pode detectar dentro da vizinhança. Por exemplo, num espaço de apenas duas classes, considerando a verificação de 3 vizinhos para cada elemento, será atribuída ao elemento a classe que ocorrer em 2 dos 3 vizinhos verificados. Caso existam mais de duas classes, o número de combinações aumenta e, conseqüentemente, a complexidade do problema (SHAPIRO; STOCKMAN, 2001).

Primeiramente aplicada ao reconhecimento de padrões em Cover e Hart (1967), a regra do vizinho mais próximo é uma das técnicas de classificação mais simples. Entretanto, o tempo computacional consumido na busca pelo vizinho mais próximo, principalmente nos casos em que o conjunto de treinamento é muito grande, pode tornar a técnica inviável. Assim, as variações do método buscam por técnicas que otimizem seu processo de busca. Em Dasarathy (1991) são apresentados os conceitos do vizinho mais próximo e variações do método. Técnicas mais recentes baseadas neste algoritmo são a versão de Zhang e Srihari (2004), o KNN rápido, que utiliza um algoritmo baseado em árvore para agrupamento e outros destaques, como Gao e Wang (2007) que busca por técnicas mais robustas seguindo a concepção original do método do vizinho mais próximo.

3.2.5.1 Vizinho mais próximo difuso (FKNN)

O algoritmo do vizinho mais próximo difuso, em inglês *fuzzy k-nearest neighbor algorithm*, definido como FKNN (KELLER *et al.*, 1985), produz regras de classificação difusa a partir dos graus de pertinência que podem ser associados aos elementos do conjunto de amostras. Keller *et al.* (1985) propõem três diferentes métodos de atribuição dos graus de pertinência aos elementos do conjunto de treinamento para uma dada aplicação.

Na versão original do classificador do vizinho mais próximo é possível verificar que as classes, representadas pelas respectivas amostras, são igualmente consideradas quando da atribuição de uma classe a um elemento. Esta é uma característica do método que pode dificultar a classificação no caso de haver sobreposição entre as amostras. Um outro problema identificado é que, uma vez atribuída uma classe a um elemento, não existe um grau de associação que possa medir a real pertinência do elemento à classe. São estes dois problemas que o algoritmo difuso apresentado por Keller *et al.* (1985) busca solucionar.

A versão não-difusa do classificador do vizinho mais próximo, descrita anteriormente, tem por base a regra que atribui a um elemento, cuja classe é desconhecida, a classe da maioria de seus k vizinhos mais próximos. A partir do momento em que o valor de k considerado é maior do que 1, existe o risco de empate na decisão da classe do elemento. Uma forma de lidar com este problema é estabelecendo restrições para o valor de k . Como por exemplo, pode-se estabelecer valores ímpares para k num problema que envolve duas classes, para que o empate não aconteça, o que não tem validade nos casos em que um número maior de classes esteja envolvida no problema. Uma outra forma de tratar o empate é, para cada uma das classes em que ocorre o empate, as distâncias para os vizinhos são somadas e a classe final atribuída ao elemento é aquela para a

qual a soma das distâncias é mínima. Ainda sim, um empate pode ocorrer e a definição da classe pode ser feita aleatoriamente. Existem casos em que a atribuição aleatória é feita, independentemente dos procedimentos contidos no algoritmo de classificação.

Um algoritmo que considera os k vizinhos mais próximos na sua versão difusa, trabalha com valores de pertinência associados a cada um dos vizinhos do elemento. Tais valores de pertinência fornecem um nível de confiança associado ao resultado da classificação. Por exemplo, se um elemento tem grau de pertinência 0,90 com relação à classe 1 e 0,05 com relação a outras duas classes, existe um grau de certeza muito alto de qual é a classe do elemento. Por outro lado, se o grau de pertinência para a classe 1 é 0,55; para a classe 2 é 0,44; e 0,01 para a classe 3, então o nível de indecisão é maior, apesar da certeza de que o elemento não pertence à classe 3.

A base do algoritmo FKNN são os graus de pertinência obtidos como função da distância dos k vizinhos mais próximos. O algoritmo se assemelha à sua versão não difusa na fase de obtenção dos k vizinhos, porém se diferencia consideravelmente no restante do procedimento. A Equação 3.9 descreve a função que gera o grau de pertinência $u_i(z)$ para um dado elemento z :

$$u_i(z) = \frac{\sum_{j=1}^k u_{ij}(1/\|z - z_j\|^{2/(M-1)})}{\sum_{j=1}^k (1/\|z - z_j\|^{2/(M-1)})} \quad (3.9)$$

Observa-se que os graus de pertinência associados a z são influenciados pelo inverso das distâncias para os vizinhos mais próximos e seus graus de pertinência u_{ij} , que indicam a pertinência do j -ésimo elemento da amostra na i -ésima classe. O efeito do inverso da distância é dar mais peso ao mais próximo, e conseqüentemente, um grau menor de pertinência é associado aos vizinhos mais afastados. A obtenção dos graus de pertinência associados às amostras pode ser feita de várias maneiras. A primeira delas é associar grau 1 aos elementos da própria classe e 0 aos demais. Outra alternativa é associar graus em função da distância do elemento da sua classe para as demais classes consideradas numa aplicação. A variável M determina o quanto a distância tem influência no cálculo da contribuição de cada vizinho no valor de pertinência dado por $u_i(z)$. Se o valor de M é 2, então a contribuição de cada vizinho é ponderada pela distância ao elemento a ser classificado. À medida que o valor de M cresce, os vizinhos têm mais peso e o efeito de suas distâncias relativas tende a diminuir. Se M se aproxima de 0, então é reduzido o número de pontos que contribuem na classe a ser definida, ou seja, os elementos mais próximos são os que realmente têm influência.

As três técnicas definidas por Keller *et al.* (1985) para definição dos graus de pertinência são como se seguem. A primeira delas atribui grau 1 para a própria classe do elemento e 0 para o restante; a segunda técnica, detalhada em Keller e Hunt (1985), é específica para aplicações em que apenas duas classes estão envolvidas e os graus de pertinência são calculados com base na distância média entre as classes. E por fim, a terceira técnica atribui graus de pertinência aos elementos da amostra de acordo com uma regra dos K vizinhos mais próximos. Os K vizinhos mais próximos de cada elemento z da classe i são encontrados e os graus de pertinência são definidos de acordo com a seguinte Equação:

$$u_j(z) = \begin{cases} 0,51 + (n_j/K) * 0,49; & \text{se } j = i \\ (n_j/K) * 0,49; & \text{se } j \neq i \end{cases} \quad (3.10)$$

É importante observar que o K considerado na Equação 3.10 não é o mesmo k do classificador, que define o número de vizinhos considerados no momento da classificação de um elemento. O valor n_j indica o número de vizinhos que pertencem à j -ésima classe. Este método tende a tornar mais difusas as pertinências dos elementos que se encontram em regiões de intersecção entre as classes da amostra, fazendo com que grau 1 seja atribuído aos elementos que estão completamente fora dessas regiões.

3.2.5.2 O classificador do protótipo mais próximo

O classificador do protótipo mais próximo é semelhante à abordagem do vizinho mais próximo. A diferença é que no tradicional algoritmo do vizinho mais próximo trabalha-se diretamente com as amostras rotuladas e, no classificador do protótipo mais próximo, um conjunto de protótipos é gerado a partir da amostra, para então ser aplicada a regra do vizinho mais próximo, sendo o método igualmente estendido para os k protótipos mais próximos e também definidas as versões difusa e não-difusa.

A técnica do protótipo mais próximo é uma variação simples da regra do KNN, sendo gerado um protótipo para cada classe e então verificada a proximidade com relação ao elemento que se deseja classificar. Em Kuncheva e Bezdek (2001) é apresentado um estudo da abordagem em que diferentes métodos de geração de protótipos são descritos e comparados, além da análise de grupos de classificadores caracterizados como pertencentes à família do classificador do protótipo mais próximo generalizado. Modelos de agrupamento, do vizinho mais próximo e mesmo o aprendizado por vetor de quantização e funções de base radial podem ser assim caracterizados.

3.2.6 Árvores de decisão

As árvores de decisão são empregadas na tarefa de classificação, e podem ser chamadas de árvores de classificação. Basicamente, uma árvore de decisão consiste de uma hierarquia de nós internos e externos que são conectados por ramos. O nó interno é a unidade de tomada de decisão que avalia através de teste lógico qual será o próximo nó descendente ou filho. Um nó externo, que não tem nó descendente, também conhecido como folha ou nó terminal, está associado a um rótulo ou a um valor. Em geral, apresenta-se um conjunto de dados associado ao nó inicial (ou nó raiz que também é um nó interno) da árvore; dependendo do resultado do teste lógico usado pelo nó, a árvore ramifica-se para um dos nós filhos e este procedimento é repetido até que um nó terminal é alcançado. No caso das árvores de decisão binária, cada nó intermediário divide-se exatamente em dois nós descendentes e uma decisão é sempre interpretada como verdadeira ou falsa.

A classificação de um padrão se inicia da raiz, no topo da árvore, que busca pelo valor de uma propriedade em particular para o padrão. As diferentes ligações que existem a partir do nó raiz correspondem aos possíveis valores assumidos pela propriedade em questão e a escolha de um deles define o próximo nó raiz. Estes valores devem ser mutuamente exclusivos, fazendo com que apenas um caminho seja seguido. O processo segue até que um nó folha seja alcançado e sua classe é atribuída ao padrão apresentado à árvore.

O uso de árvores de decisão em classificação de imagens digitais é abordado em Clarke e Bittencourt (2003). Em Duda *et al.* (2000) as árvores de decisão são discutidas dentre as técnicas de reconhecimento de padrões sintático, em que os elementos não são descritos por atributos numéricos e uma medida natural de distância pode ser definida entre eles, mas sim descritos por atributos nominais, que compõem a entrada de um classificador.

3.2.7 Redes Neurais Artificiais (RNA)

As redes neurais artificiais, também conhecidas como *paradigma conexionista*, são sistemas paralelos distribuídos compostos por unidades de processamento simples que calculam determinadas funções matemáticas (BRAGA *et al.*, 2000; HAYKIN, 2001). Para solucionar um problema empregando as redes neurais, inicialmente, tem-se uma fase de aprendizagem, onde um conjunto de exemplos é apresentado para a rede, a qual extrai automaticamente as características necessárias para representar a informação fornecida. Estas características são utilizadas posteriormente para gerar respostas para o problema. As principais características das redes neurais são a capacidade de

aprendizado e generalização. A generalização é a habilidade de produzir respostas para padrões de entrada que são similares, mas não idênticos, aos padrões apresentados à rede durante o seu treinamento.

A capacidade de generalização das redes neurais é o que garante o bom desempenho deste paradigma, possibilitando que sejam empregadas em uma grande variedade de aplicações. Entretanto, para determinadas aplicações é importante não só o desempenho, mas também a compreensão do resultado obtido por parte do usuário.

As redes neurais vem sendo empregadas com sucesso na tarefa de classificação (DUDA *et al.*, 2000). Pesquisas na área mostram que esta técnica apresenta características adequadas à sua aplicação na classificação de dados. Além de serem consideradas aproximadores universais de funções, são modelos não-lineares, o que as tornam altamente flexíveis na modelagem de problemas do mundo real. Sendo ainda métodos auto-adaptativos, capazes de se ajustarem ao dado apresentado sem que especificações funcionais sejam fornecidas ao modelo. Por fim, as redes neurais são capazes de estimar probabilidades a posteriori, fornecendo uma base para estabelecer a regra de classificação e a análise estatística do método. Bishop (1995) aborda as redes neurais para reconhecimento de padrões e, em Zhang (2000) é apresentada uma revisão da literatura com relação aos principais avanços dos classificadores neurais.

No caso da utilização das redes neurais em classificação, pode ser importante para o usuário entender porque um determinado padrão de entrada foi classificado em uma dada classe. Para problemas, como o diagnóstico médico, por exemplo, é essencial que exista um mecanismo que explique e justifique as decisões tomadas pela rede, uma vez que se trata de um problema em que a segurança na operação é um aspecto importante (BRAGA *et al.*, 2000).

Existem na literatura diferentes modelos de redes neurais que podem ser empregados em classificação. O perceptron é um classificador supervisionado para duas classes com apenas um neurônio, limitado a resolver problemas lineares. Já as redes perceptron de múltiplas camadas (MLP) tratam problemas de múltiplas classes, constituindo modelos simples que são capazes de resolver problemas não-lineares, como são, de fato, grande parte das aplicações do mundo real. Um dos algoritmos mais comumente utilizados no treinamento de uma MLP é o algoritmo de retropropagação do erro (*error backpropagation*), que consiste nos passos para frente, propagação e para trás, retropropagação, através das camadas da rede.

A arquitetura das redes função de base radial (RBF) e máquina de vetor suporte (SVM)

se assemelham no ponto em que buscam por uma transformação não-linear no espaço de características que faz com que as amostras apresentadas sejam linearmente mapeadas em suas respectivas classes. A idéia é encontrar um conjunto de protótipos que gere novas características. Estas redes se diferem na escolha da transformação não-linear, no cálculo dos protótipos e no método de aprendizado da rede.

Ainda dentre as redes neurais empregadas na classificação de dados estão os mapas auto-organizáveis (SOM - *Self organizing maps*) e as redes LVQ (*Learning Vector Quantization*), descritas por Kohonen (2001), e as redes baseadas na teoria da ressonância adaptativa (ART - *Adaptative resonance theory*), em Carpenter e Grossberg (1987b), Carpenter e Grossberg (1987a), Carpenter e Grossberg (1990). Estas redes apresentam arquitetura simples, com apenas uma camada, onde os padrões são apresentados na entrada e associados com os nós de saída.

As redes SOM vem sendo aplicadas a uma ampla variedade de problemas desde seu surgimento, destacando as potencialidades de análise de agrupamento e mineração de dados. Em virtude da propriedade de serem capazes de selecionar um conjunto das melhores características a partir dos dados do espaço de entrada, segundo Haykin (2001), são consideradas adequadas para a tarefa de agrupamento. As redes LVQ são algoritmos de aprendizado competitivo. As variações LVQ2 e LVQ3 são abordadas na literatura como procedimentos similares a abordagens clássicas de agrupamento, como o algoritmo k-médias, apresentado na seção 3.2.1.1.

As redes ART são modelos projetados para solucionar o dilema plasticidade-estabilidade, entretanto, o número e o tamanho dos grupos obtidos como resultado é dependente do valor escolhido para o parâmetro nomeado limiar de vigilância, que é utilizado para decidir quando um padrão apresentado para a rede será atribuído a um grupo já existente ou se dará início a um novo grupo no processo de agrupamento. Quanto à forma dos grupos detectados, tanto a rede ART quanto a SOM detectam apenas grupos hiperesféricos (JAIN *et al.*, 1999).

3.3 Combinação de classificadores

Com o objetivo de desenvolver sistemas de classificação que sejam capazes de alcançar o melhor desempenho para uma dada aplicação, diferentes esquemas são propostos. A combinação de classificadores vem sendo alvo de pesquisas, como pode ser observado em Kittler *et al.* (1998), Lam e Suen (1995), Rogova (1994). Kuncheva (2000) apresenta outras referências e, mas recentemente, Kuncheva (2004) aborda métodos e algoritmos referentes à questão.

Kuncheva (2004) define quatro abordagens distintas na combinação de classificadores. A primeira delas se refere ao projeto de combinadores, a segunda ao uso de diferentes classificadores, a terceira ao uso de diferentes conjuntos de atributos para os dados a serem classificados e a quarta abordagem se refere ao uso de diferentes conjuntos de amostras no projeto dos classificadores. Tais abordagens podem ser sintetizadas por níveis: (A) Nível de combinação, (B) Nível de classificadores, (C) Nível de atributos e (D) Nível de dados.

São consideradas duas principais estratégias de combinação de classificadores na literatura: fusão e seleção. A fusão assume que cada classificador tem conhecimento sobre todo o espaço de atributos e na seleção, que cada classificador é especialista em uma parte do espaço de atributos. Assim, pode-se dizer que a primeira abordagem é competitiva, enquanto a segunda é complementar.

As diferentes combinações podem também apresentar parâmetros que são definidos após a combinação, o que caracteriza as combinações treináveis. Para os casos em que todos os parâmetros são treinados nos classificadores de base, as combinações obtidas são testadas diretamente para uma dada aplicação.

Destaca-se neste trabalho a combinação de distribuições de possibilidade obtidas a partir dos classificadores de base. Kuncheva (2004) descreve a combinação de classificadores cujas saídas são contínuas e podem ser interpretadas de duas formas mais comuns: como valores de crença com relação às classes do problema e como estimativa das probabilidades a posteriori para as classes. Para cada elemento z da amostra, um classificador Cl produz um valor $p_{ij} \in [0, 1]$, indicando o suporte para a hipótese que z pertence à classe c_j . Dado um conjunto de classificadores $Cl_i, i = 1, \dots, \Delta$ associado, a combinação é dada por:

$$Cl(\mathbf{z}) = \begin{bmatrix} p_{1,1}(\mathbf{z}) & \cdots & p_{1,j}(\mathbf{z}) & \cdots & p_{1,c}(\mathbf{z}) \\ p_{i,1}(\mathbf{z}) & \cdots & p_{i,j}(\mathbf{z}) & \cdots & p_{i,c}(\mathbf{z}) \\ p_{\Delta,1}(\mathbf{z}) & \cdots & p_{\Delta,j}(\mathbf{z}) & \cdots & p_{\Delta,c}(\mathbf{z}) \end{bmatrix} \quad (3.11)$$

Esta combinação permite que vários classificadores sejam agregados obtendo um suporte final μ_j para cada classe, gerando métodos que se dividem em fusão consciente das classes e fusão indiferente às classes. No primeiro caso são exploradas as colunas de $Cl(z)$ e no segundo, os valores $p_{ij}(z)$ são tratados como características de um espaço intermediário, e usa-se um classificador extra para tomar a decisão final.

A Seção 3.2 deste capítulo descreve várias técnicas a partir das quais é possível derivar valores contínuos associados, ou seja, distribuições de probabilidade ou possibilidade associadas a cada elemento z , como os métodos difusos, que retornam graus de pertinência; e aqueles a partir dos quais se pode derivar as probabilidades a posteriori, como o classificador de Bayes, MAXVER e KNN.

Numa combinação não-treinável, calcula-se o suporte da seguinte maneira:

$$\mu_j(x) = F[p_{1,j}(x), \dots, p_{\Delta,j}(x)] \quad (3.12)$$

onde F é a função de combinação, que pode ser selecionada de várias maneiras:

- Média simples (F=mean)

$$\mu_j(x) = 1/L \sum_{i=1}^L p_{ij}(x) \quad (3.13)$$

- Mínimo / Máximo (F=min/max)

$$\mu_j(x) = \min_i \{p_{ij}(x)\} \quad (3.14)$$

- Produto (F=prod)

$$\mu_j(x) = \prod_{i=1}^L p_{ij}(x) \quad (3.15)$$

Outras funções e exemplos são apresentados em Kuncheva (2004).

3.4 O desempenho de um classificador

Segundo Kuncheva (2000), a avaliação do desempenho de um classificador envolve não só a acurácia da classificação, mas também outras questões como a taxa de rejeição e a precisão correspondente; habilidade de identificar pontos mal definidos no domínio do problema (*outliers*); o desempenho do método frente a dados com ruído ou dados incompletos e o desempenho com variáveis mistas como entrada.

Para medir acurácia deve-se sempre considerar que sua medida pode ser adquirida para um conjunto de treinamento e para um conjunto de teste. Estas medidas são diferentes, sendo comum que a qualidade do conjunto de treinamento seja praticamente perfeita e a qualidade de teste desfavorável. Define-se conjunto de treinamento como sendo aquele

utilizado para treinar o classificador, enquanto que o conjunto de teste é formado pelas situações desconhecidas, por elementos que jamais foram apresentados ao classificador. Geralmente, a acurácia medida no conjunto de teste para os casos em que a verdade é conhecida, é aquela que tem importância na prática.

O processo usual para estimar acurácia de um método de classificação supervisionado utiliza os elementos conhecidos, que compõem o conjunto verdade, que aqui é tido também como o conjunto de amostras das classes de um determinado problema. Primeiramente, o método é treinado usando parte dos dados, em seguida as regras geradas são testadas no restante do conjunto e o resultado comparado com a verdade. A proporção de pontos corretamente classificados é uma estimativa imparcial da qualidade da regra fornecida pelo treinamento, em que o conjunto de treinamento é formado aleatoriamente a partir da amostra conhecida dos dados.

Alguns métodos são sugeridos na literatura e seu uso depende das características da aplicação tornando um ou outro método mais adequado. A seguir são descritas as idéias de algumas abordagens utilizadas.

3.4.1 Taxa de acerto

Define-se por taxa de acerto de um classificador a relação entre o número de pontos bem classificados no conjunto e o número total de elementos considerados. Seja o conjunto de dados $Z = Z_{tr} \cup Z_{te}$ onde Z_{tr} é o conjunto de dados de treinamento e Z_{te} o conjunto de dados de teste.

A acurácia de um classificador preciso Cl em relação ao conjunto de dados Z é dada por:

$$acc(Cl, Z) = |\{z \in Z \mid l(z) = Cl(z)\}|/|Z| \quad (3.16)$$

A taxa de acerto é uma medida comumente empregada para estimar o desempenho da classificação obtida. A acurácia exata de um método é atingida caso seja possível treinar um classificador com todas as possíveis entradas. Na realidade, não é considerada esta possibilidade e são calculadas estimativas.

3.4.1.1 A matriz de confusão

Uma das maneiras usualmente empregadas para representar a exatidão de uma classificação é através da matriz de confusão, onde são representadas as quantidades

ou porcentagens de elementos de um conjunto de dados correta e incorretamente classificados em cada uma das classes de um problema.

Cada elemento x_{ij} na matriz de confusão corresponde ao número de elementos tidos como da j -ésima classe (tomando as amostras de referência), classificados pelo modelo de classificação como pertencentes à i -ésima classe. O número de acertos para cada classe localiza-se na diagonal principal da matriz, ou seja, cada x_{ii} indica o número de elementos classificados corretamente pelo modelo. Os demais elementos representam erros na classificação. Observa-se que esta abordagem, além de fornecer a exatidão global do método, permite avaliar o desempenho do classificador para uma classe individual.

3.4.1.2 O índice *kappa*

A partir da matriz de confusão é possível extrair o índice *kappa*, definido em Bishop *et al.* (1975) como sendo:

$$kappa = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} + x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} + x_{+i})} \quad (3.17)$$

onde r é o número de linhas ou colunas da matriz de confusão, x_{ii} é o número de observações na linha i e coluna i , x_{i+} é a soma dos valores na linha i , x_{+i} é a soma dos valores na coluna i e N é o número total de observações. O estimador *kappa* varia entre 0 e 1.

Uma das vantagens provenientes da utilização do índice *kappa* é que ele incorpora a informação dos elementos mal classificados, e não apenas dos bem classificados como exatidão global.

3.4.2 Conjuntos de treinamento e teste

Segundo Kuncheva (2004), para um dado conjunto o que se busca é treinar um classificador com o máximo de informação possível, ou seja, com a maior quantidade de dados. Da mesma maneira, para avaliar o desempenho do método requer que a maior quantidade de dados seja testada. Entretanto, usar todo o conjunto no treinamento e o mesmo para teste, pode levar o classificador a um super treinamento (*overfitting*). Para evitar este problema, são definidas na literatura técnicas que buscam pelo melhor uso do conjunto de dados.

A seguir são brevemente comentados alguns métodos para determinar conjuntos de

teste e treinamento. Kuncheva (2004) comenta ainda uma prática comum em que, em lugar de dois conjuntos, treinamento e teste, é empregado ainda um terceiro, o conjunto de validação. Este conjunto trabalha como um pseudoteste. O treinamento do classificador é feito até que sua melhoria não corresponda à uma melhoria avaliada no conjunto de validação. Uma desvantagem nítida neste caso é que nem todos os conjuntos de dados têm tamanho suficiente para que sejam subdivididos em três partes, o que acarreta conjuntos pequenos cujas estimativas de erro não correspondem à realidade.

3.4.2.1 Método H (Hold-out)

Este método, também definido como Teste/Treinamento em Michie *et al.* (1994), consiste na forma geral, comumente empregada, de avaliação de um método de classificação, em que o conjunto de dados Z é dividido em duas partes, uma para treinamento e outra para teste do classificador. A proporção em que os dois conjuntos são divididos pode variar e a medida de desempenho pode ser resultante da média de duas variações distintas. É um método mais empregado quando o conjunto de treinamento e teste tem tamanho grande, maior que 1000 elementos (MICHIE *et al.*, 1994). Segundo Kuncheva (2004) a estimativa do erro usando o método H é pessimista.

3.4.2.2 Validação cruzada

A validação cruzada, também chamada método de rotação, é mais utilizada para os problemas em que as amostras têm tamanho moderado. Na sua forma mais simples o método divide o conjunto de amostras em L subconjuntos. Cada um dos L subconjuntos é submetido ao classificador, cujas regras são obtidas do treinamento, utilizando a união dos $(L - 1)$ subconjuntos restantes. O procedimento é repetido L vezes escolhendo em cada uma delas um L diferente e, ao final, a taxa de erro (ou acerto) do classificador é dada pela média dos L experimentos executados. Uma versão muito empregada do método é aquela em que o número de subconjuntos é igual ao número de elementos que compõem as amostras, definido na literatura como o *leave-one-out*, o que define a validação cruzada completa.

Segundo Michie *et al.* (1994), uma dificuldade prática do uso da validação cruzada em métodos computacionais que têm fase de treinamento mais custosa, como as redes neurais, é justamente a repetição desta fase empregando L conjuntos distintos, o que pode demandar um grande esforço computacional.

3.4.2.3 Bootstrap

A idéia do método *bootstrap*, detalhado em Efron e Tibshirani (1993), é repetir a classificação de todo o conjunto de dados várias vezes e estimar quantidades, analisando repetidamente as subamostras do conjunto de dados em lugar dos subconjuntos do conjunto de dados, como feito no método anterior. Isto é feito pela geração aleatória de L conjuntos com cardinalidade N a partir do conjunto original Z , com reposição. Ou seja, uma subamostra é definida como sendo um subconjunto aleatório de tamanho N do conjunto de amostras, onde a parte omitida do conjunto é preenchida repetindo dados dentro do conjunto.

CAPÍTULO 4

ESTRATÉGIAS DE CLASSIFICAÇÃO POSSIBILISTA

Neste capítulo são propostas estratégias de classificação de dados em espaços multidimensionais que empregam teoria de possibilidades e relações de similaridade. Os modelos de classificação são supervisionados, trabalham no espaço de atributos e retornam distribuições de possibilidade associadas aos elementos da amostra ou a todos os elementos do conjunto de dados, caracterizando efetivamente as abordagens local e global.

Destacam-se dois trabalhos como base das abordagens definidas. O classificador possibilista proposto por Malinski (2000), descrito na Seção 3.2.2.2, e o trabalho de Keller *et al.* (1985), descrito na Seção 3.2.5.1. O classificador de Malinski (2000) deriva distribuições do tipo global, ou seja, para qualquer elemento é possível obter uma distribuição dos histogramas das amostras das classes, que indicam o quanto cada nível de cinza é compatível com uma dada classe. Nota-se que originalmente o método é proposto para a classificação de imagens, onde faz-se referência a histogramas de imagens e compatibilidade de níveis de cinza entre as classes. Keller *et al.* (1985) definem a técnica do vizinho mais próximo difuso (FKNN). Para esta técnica, observa-se que distribuições de probabilidade são obtidas para cada elemento da amostra e que a classificação de um elemento qualquer do conjunto é dada de acordo com os k vizinhos mais próximos. Seguindo a mesma linha, propõe-se aqui uma abordagem que difere na forma com a qual os graus de pertinência para cada elemento com relação a cada classe são definidos.

As estratégias do vizinho mais próximo e possibilista servem de motivação para os algoritmos propostos neste trabalho. A estratégia AGRUP, baseada em protótipos, gera centros representativos de cada classe, e a partir deles são derivados os parâmetros que definem as relações de similaridade para transformá-los em protótipos difusos. O caso mais geral desta abordagem é aquele em que cada elemento da amostra é tido como um centro, sendo as relações de similaridade aplicadas aos próprios elementos do conjunto de amostras. As relações de similaridade criam conjuntos difusos, a partir dos quais se derivam as distribuições de possibilidade para cada elemento do conjunto de dados no domínio total do problema.

Valendo-se do agrupamento gerado pela abordagem AGRUP, é definido ainda um classificador alternativo, em que os graus gerados pela aplicação de uma técnica de agrupamento difuso são reutilizados. A idéia é definir uma função para gerar os graus

de pertinência associados a cada elemento da amostra e utilizar a regra do vizinho mais próximo para obter a classificação.

Parte da metodologia proposta tem aplicação bem-sucedida com resultados apresentados em Drummond e Sandri (2004) empregando imagem de satélite. Para a etapa inicial do método AGRUP, que se atém à geração de protótipos, foi proposta uma maneira de otimizar a obtenção de centros para cada classe usando um índice de agrupamento. Uma abordagem de transformação global dada pelo índice de agrupamento está publicada em Drummond e Sandri (2005) com testes nos conjuntos de dados **cone-torus** (KUNCHEVA, 2000) e Iris, disponível em <http://www.ics.uci.edu/mlearn/MLRepository.html>.

O conjunto sintético **cone-torus**, empregado em Kuncheva (2000), é usado na literatura com propósito ilustrativo e didático para teste de modelos de classificação. Alguns dos exemplos contidos neste capítulo e no seguinte, usados para ilustrar as etapas das estratégias definidas, fazem uso deste conjunto ou de parte dele.

4.1 Definições

Sejam as seguintes definições:

- \mathbf{C} é o conjunto de m classes, Z é o conjunto de dados, $B \subset Z$ é o conjunto de amostras, ou seja, os elementos de B são pontos em \mathbf{R}^ρ (espaço ρ -dimensional nos reais), e $B_i \subseteq B$ é um conjunto de pontos amostrais associados à classe $c_i \in \mathbf{C}$.
- $d(b_1, b_2)$ é a distância entre dois pontos b_1 and $b_2 \in \mathbf{R}^\rho$, correspondente a uma dada métrica de distância, por exemplo, a distância euclidiana.
- t_{ij} é um dos centros (protótipo) da classe c_i e T_i é o conjunto de todos os centros associados à classe c_i .
- $B_{ij} \subseteq B_i$ é o conjunto de todos os elementos da amostra relativa à classe c_i que estão mais próximos do centro $t_{ij} \in T_i$ do que de qualquer outro centro de c_i , de acordo com a distância d . Formalmente, $\forall b \in B_i, b \in B_{ij}$ sse $\inf_w d(t_{iw}, b) = d(t_{ij}, b)$.

Para exemplificar, na Figura 4.1a é apresentado um conjunto sintético¹ em duas dimensões com duas classes. Na Figura 4.1b o mesmo conjunto é mostrado juntamente

¹Conjunto derivado da base de dados sintética **cone-torus**.

com os centros associados a cada classe. Para a classe c_1 , $T_1 = \{t_{11}\}$ e para a classe c_2 tem-se $T_2 = \{t_{21}, t_{22}, t_{23}, t_{24}, t_{25}, t_{26}, t_{27}, t_{28}\}$.

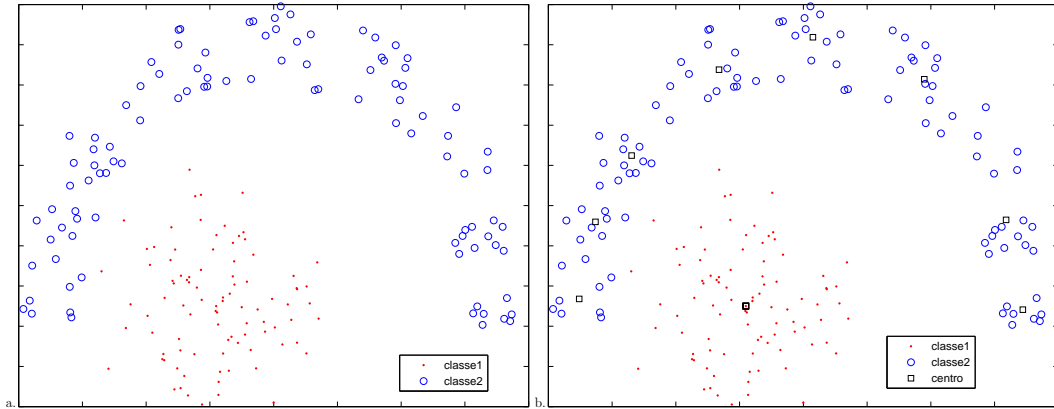


Figura 4.1 - a) Conjunto sintético b) Conjunto sintético com centros associados às classes.

4.2 O classificador AGRUP

No classificador difuso baseado em agrupamento, o AGRUP, um conjunto de protótipos T é inicialmente criado. O método busca, primeiramente, pela transformação do espaço de atributos que apresente a melhor separabilidade das classes gerando centros representativos de cada classe. Tais centros são chamados de protótipos, que numa segunda fase são transformados em conjuntos difusos pela aplicação de uma relação de similaridade. Como saída, o classificador apresenta uma distribuição de possibilidade para cada ponto z do conjunto de dados Z .

4.2.1 Transformação do espaço de atributos

A cada classe c_i está associado um conjunto de pontos amostrais $B_i \subset Z$, tal que $\forall i, j, i \neq j, B_i \cap B_j = \emptyset, B_i \cup B_j = B$, ou seja, a classificação correta dos elementos para cada c_i . Dado o conjunto de amostras B_i , são gerados centros para cada uma das classes empregando um algoritmo de agrupamento. É proposta aqui a utilização do algoritmo FCM (vide Seção 3.2.1.2), mas outros métodos poderiam ser empregados, como por exemplo, o clássico algoritmo k-médias, descrito na Seção 3.2.1.1.

Um dos problemas da utilização de uma abordagem baseada em agrupamento é definir o número de centros que deve ser empregado para cada classe. Um número ótimo nem sempre existe, mas busca-se um número que seja grande o suficiente para gerar uma classificação perfeita da amostra, e ainda pequeno o suficiente para garantir uma generalização aceitável. Uma solução para o problema é definir um número inicial

de centros para cada classe e, em seguida, aumentar ou diminuir estes números de acordo com alguma medida que verifique a qualidade dos grupos obtidos. Entretanto, a qualidade é geralmente medida verificando a classificação obtida a partir dos centros definidos. No caso em que o problema de classificação envolve um grande número de atributos, a verificação é muito custosa, mesmo para um pequeno número de centros, devido ao número exponencial de combinações possíveis que deve ser verificado. Uma solução alternativa é aplicar a cada classe um algoritmo de agrupamento que encontre o número ideal de centros. Este algoritmo pode iniciar ou com um pequeno número de grupos que são refinados sucessivamente, ou com um grande número de grupos que são unidos até que um número aceitável de grupos seja encontrado. Abordagens apresentadas na literatura com este propósito são discutidas na seção 3.2.1.

É evidente que fixar um número de centros é uma solução mais simples e mais eficiente do ponto de vista computacional. Por outro lado, determinar automaticamente o número de centros para cada classe não depende da qualidade da classificação propriamente dita. Encontrar o número adequado de centros para cada classe é um problema dependente das características do conjunto de dados. Em Drummond e Sandri (2005) é proposto um índice de agrupamento que considera o conjunto de dados como um todo na determinação do melhor número de centros para uma dada classe, e não somente os dados pertencentes a esta classe. O índice pode ser empregado para decidir de maneira automática quando uma classe deve ser reagrupada com um número maior de centros. Não se trata de uma técnica de agrupamento, mas do uso de técnicas existentes para a formação de um método de meta-agrupamento.

Desta maneira, a etapa de geração de protótipos e posterior aplicação das relações e similaridade a cada um deles, pode ser vista como a transformação do espaço de atributos, a partir do qual o algoritmo de classificação deriva a distribuição de possibilidade $\pi(z)$ para cada ponto z . A seguir são apresentadas medidas que podem ser derivadas do conjunto de amostras B e do conjunto de centros T , obtido a partir do algoritmo de agrupamento, no caso o FCM, empregando a distância euclidiana. O índice de agrupamento é calculado a partir destas medidas. São descritas as abordagens local e global do meta-agrupamento, e as relações de similaridade aplicadas ao conjunto de protótipos.

4.2.1.1 Parâmetros

O meta-agrupamento proposto é regido por uma coleção de índices calculados a partir do conjunto de protótipos T e do conjunto de amostras B . Estes índices são parâmetros que auxiliam na avaliação da qualidade do agrupamento obtido para cada classe, além

de serem utilizados nas relações de similaridade que definem os protótipos difusos. Esta última questão é discutida na seção seguinte. Os parâmetros são dados por:

- a) $d_{min}(t_{ij}) = \inf_{b \in B_{ij}} d(t_{ij}, b)$ (Raio mínimo)
- b) $d_{max}(t_{ij}) = \sup_{b \in B_{ij}} d(t_{ij}, b)$ (Raio máximo)
- c) $d_{ext}(t_{ij}) = \inf_{b \in B_l, l \neq i} d(t_{ij}, b)$ (Raio externo)
- d) $d_{prox}(t_{ij}) = \inf_{t \in T_l, l \neq i} d(t_{ij}, t)$ (Raio do centro mais próximo)
- e) $d_{in}(t_{ij}) = d_{ext}(t_{ij}) - d_{max}(t_{ij})$ (Raio interno: diferença de raios)

Os índices a) e b) são chamados locais, ou seja, as distâncias consideradas são aquelas calculadas entre os centros i da classe j e os elementos da amostra desta mesma classe j . Assim, d_{min} (respectivamente, d_{max}) determina a distância entre o centro t_{ij} para o elemento mais próximo (respectivamente, mais distante) de B_{ij} . Já os índices c) e d) podem ser vistos como globais, uma vez que levam em conta as distâncias entre centros e pontos das amostras de todas as classes. O índice d_{ext} representa a distância do centro t_{ij} ao elemento mais próximo que pertence a outra classe. E o índice d_{prox} indica a distância do centro t_{ij} ao centro mais próximo pertencente a uma outra classe.

Na Figura 4.2 são ilustrados os índices apresentados. Note que o raio mínimo é 0 neste exemplo, pois o centro t_{11} definido coincide com um ponto da classe 1.

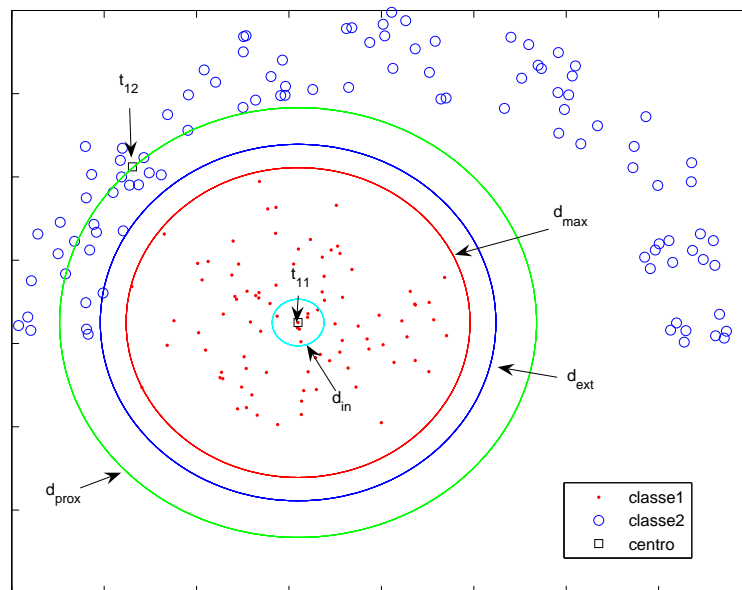


Figura 4.2 - Parâmetros: d_{max} , d_{ext} , d_{prox} e d_{in} .

4.2.1.2 Meta-agrupamento: abordagens global e local

A partir dos índices apresentados, verifica-se que num grupo com centro t_{ij} , se o seu raio máximo é maior do que seu raio externo, existe pelo menos um elemento da classe c_i cuja distância d é maior do que a distância aos pontos que não pertencem à classe c_i . Formalmente, para $d_{max}(t_{ij}) > d_{ext}(t_{ij}), \exists b \in B_{ij}, \exists b' \in B - B_i, d(t_{ij}, b) > d(t_{ij}, b')$. Entretanto, avaliar a qualidade de um agrupamento usando somente esta condição não é muito significativo, visto que se trata de uma condição do tipo sim ou não. Desta forma, através do índice de agrupamento mis , definido a seguir, é medida a qualidade de um agrupamento:

$$\begin{aligned} mis_{ij} &= d_{max}(t_{ij})/d_{ext}(t_{ij}) \\ mis_i &= \sup_j mis_{ij} \end{aligned} \quad (4.1)$$

Nota-se que $mis_{ij} > 1$ se e somente se $d_{max}(t_{ij}) > d_{ext}(t_{ij})$. Quanto maior é mis_{ij} , maior a possibilidade de pontos mal classificados para a classe c_i , já que o critério de classificação é baseado somente na menor distância do elemento ao centro de uma classe.

O meta-agrupamento global é tal que são gerados centros para cada classe c_i até que um limiar q seja alcançado, descrito pela condição: $mis_{ij} < q, q \geq 1$. Enquanto a condição é falsa, toda a classe é reagrupada, os centros da iteração anterior são descartados e um número de centros mais 1 é criado. Neste ponto vale observar que os centros podem ser gerados aleatoriamente dentro dos limites da aplicação ou podem ser os próprios pontos do conjunto de amostras. Os experimentos realizados seguem a geração aleatória de centros e o algoritmo pára se o número de centros atingir o valor do número de pontos pertencentes à classe. Na Figura 4.3 são apresentadas 3 gerações de centros, onde o meta-agrupamento global encontrou 8, 9 e 10 centros para a classe c_2 .

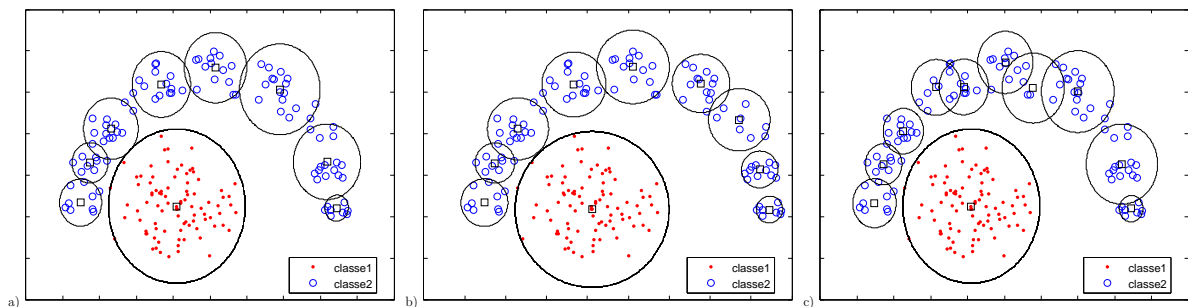


Figura 4.3 - Meta-agrupamento global com: a) 8 centros b) 9 centros e c) 10 centros.

O meta-agrupamento local emprega também o índice definido pela Equação 4.1, porém a cada iteração, se a condição é satisfeita para alguns dos centros criados, estes são guardados e os dados pertencentes a eles excluídos do reagrupamento. Nesta abordagem é possível, mas não garantido, que um número de centros menor seja encontrado. Entretanto, o que define o bom desempenho de uma ou outra abordagem são as características dos dados apresentados. A Figura 4.4 mostra as iterações do meta-agrupamento local.

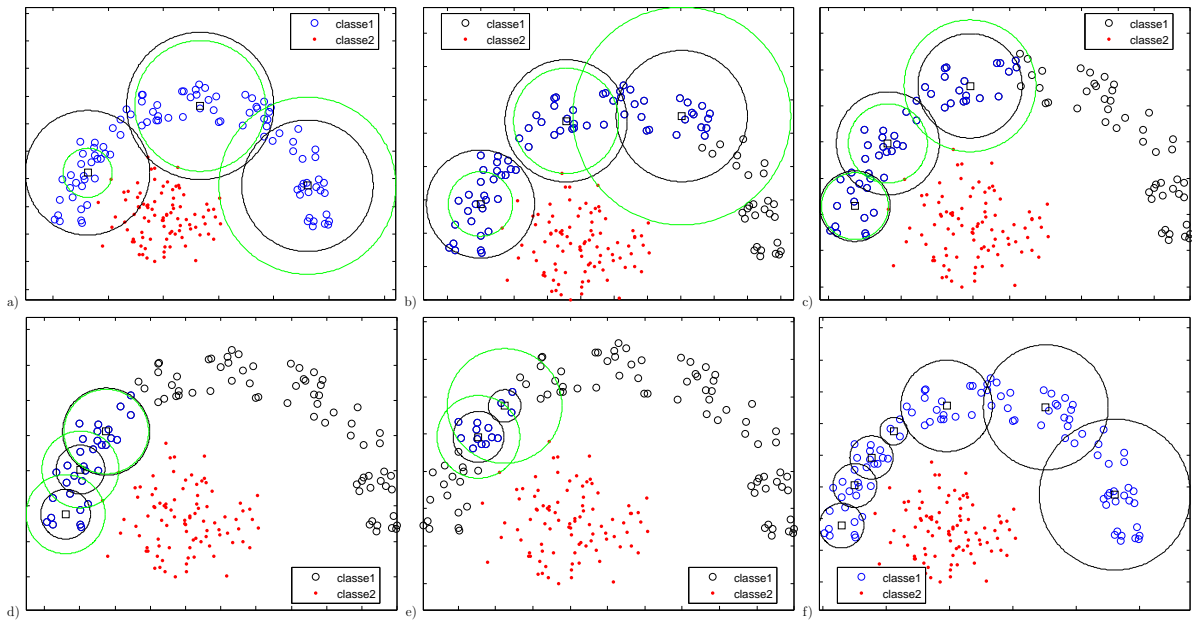


Figura 4.4 - Fases do meta-agrupamento local.

4.2.1.3 Relações de similaridade

Para transformar os protótipos obtidos para as classes em protótipos difusos são empregadas as relações de similaridade (vide seção 2.4.1). Neste trabalho são propostas relações de similaridade lineares e não-lineares para a criação dos conjuntos difusos.

A família de relações de similaridade definida é tal que uma hipersfera é criada ao redor de cada centro, onde o raio do núcleo é dado por γ e o raio do suporte por λ . Seja $S_{\gamma,\lambda}^f(t_{ij}) : P \rightarrow [0, 1]$ com $f : R^+ \rightarrow R^+$. Desta forma, são definidas:

- Relação de similaridade linear

$$\text{Seja } L_\delta(x) = \max(0, 1 - x/\delta).$$

$$S_{\gamma,\lambda}^L(t_{ij})(b) = \begin{cases} 1, & \text{se } d(t_{ij}, b) \leq \gamma \\ L_{\lambda-\gamma}(d(t_{ij}, b) - \gamma), & \text{senão} \end{cases} \quad (4.2)$$

• **Relação de similaridade gaussiana**

Seja $G_\sigma(x) = \exp(-x^2/(2 * \sigma^2))$.

$$S_{\gamma,\lambda}^G(t_{ij})(b) = \begin{cases} 1, & \text{se } d(t_{ij}, b) \leq \gamma \\ G_{(\lambda-\gamma)/3}(d(t_{ij}, b) - \gamma), & \text{senão} \end{cases} \quad (4.3)$$

Neste caso, o raio do suporte é infinito e λ é uma aproximação. A definição apresentada deve ser modificada para que o suporte seja finito.

Para a relação de similaridade trapezoidal tem-se duas funções γ e uma função λ :

$$\begin{aligned} \gamma_{tr1} &= nn * \min(d_{max}(t_{ij}), d_{ext}(t_{ij})), 0 < nn < 1. \\ \gamma_{tr2} &= \begin{cases} \min(d_{max}, d_{ext}), & \text{se } d_{in} > 0 \\ 0, & \text{senão} \end{cases} \end{aligned} \quad (4.4)$$

$$\lambda_{tr1} = mm * d_{max}(t_{ij}), mm > 1. \quad (4.5)$$

O núcleo γ_{tr1} é proporcional ao raio de um dado centro que limita a região associada a pontos corretamente classificados. O suporte λ_{tr1} é proporcional à distância entre o centro t_{ij} e o elemento mais distante associado a ele (d_{max}).

Para a relação de similaridade gaussiana o γ e λ são definidos pelas Equações 4.6 e 4.7, sendo o núcleo proporcional ao suporte:

$$\gamma_{g1} = (1/3) * \lambda_{g1}. \quad (4.6)$$

$$\lambda_{g1} = d_{prox}/2. \quad (4.7)$$

A aplicação de uma relação de similaridade S a um protótipo t_{ij} define um protótipo difuso, ou seja, a cada protótipo $t_{ij} \in T$ está associado um conjunto difuso em Z , com função de pertinência $\mu_{ij}(Z) = S(t_{ij}, z)$.

4.2.2 Algoritmo de classificação

O algoritmo de classificação propriamente dito é a etapa do método em que são derivadas as distribuições de possibilidade para cada elemento do conjunto de dados. O método verifica o grau de pertinência de cada elemento z para os protótipos difusos definidos para cada classe, sendo que a cada classe c_i está associado um conjunto de protótipos difusos. A possibilidade de cada elemento “pertencer” a um dos protótipos é dada formalmente por $\pi(z)(c_{ij}) = \mu_{ij}(z) = S(t_{ij}, z)$.

Dada a possível ocorrência de um número de centros maior do que 1, simplifica-se a obtenção de apenas um único valor associado a cada classe através da agregação dos valores obtidos com relação a cada protótipo. Utiliza-se a t-conorma *max*, o que define $\pi(z)(c_i) = \max_{t_{ij} \in T_i} \pi(z)(t_{ij})$. Nota-se que esta é uma simplificação sugerida, o que não impede que outras regras sejam empregadas, como a análise e ponderação do número de protótipos de cada classe cujo valor de pertinência do elemento é não nulo.

Desta maneira, cada elemento pertence a todas as classes com $\pi(z)(c_{ij}) \in [0, 1]$. A classe ou o subconjunto de classes associado a cada elemento é então definida por uma heurística de decisão ou critério de classificação. As heurísticas empregadas neste trabalho são apresentadas no Capítulo 5.

4.2.3 Amostras definidas como centros

Para o caso específico em que os centros são as próprias amostras, os parâmetros são obtidos com base nas distâncias entre os elementos da amostra, da mesma classe e de classes distintas. E os conjuntos difusos são criados a partir da própria amostra, aplicando-se a relação de similaridade a cada elemento pertencente ao conjunto. Na etapa de classificação é obtida, da mesma maneira, uma distribuição de possibilidade π para cada elemento z , e ao final, um critério de classificação é empregado para obtenção da classe de cada elemento.

A transformação do espaço de atributos identificada neste caso, consiste apenas em aplicar as relações de similaridade aos próprios elementos da amostras, criando conjuntos difusos. Os parâmetros para as relações de similaridade, definidos com base no vizinho mais próximo de cada ponto, da mesma classe e de outra classe, são apresentados a seguir:

a) $d_{proxin}(b_{ij}) = (\text{Ponto mais próximo da mesma classe})$

b) $d_{proxext}(b_{ij}) = (\text{Ponto mais próximo de outra classe})$

Assim a relação de similaridade trapezoidal definida pela Equação 4.2 pode ser dada empregando-se como núcleo:

$$\gamma_{tr1} = nn * \min(d_{proxin}(b_{ij}), d_{proxext}(b_{ij})), 0 < nn < 1. \quad (4.8)$$

e como suporte:

$$\lambda_{tr1} = mm * d_{proxin}(b_{ij}), mm > 1. \quad (4.9)$$

Nota-se que estas medidas são análogas àquelas empregadas no classificador baseado em agrupamento, porém com parâmetros adequados a cada abordagem.

4.3 Uma abordagem alternativa

Para a estratégia AGRUP, quando utilizado um algoritmo de agrupamento difuso na sua fase inicial, onde são gerados os centros, interessam apenas os protótipos definidos. Os graus de pertinência que a própria técnica produz são inutilizados. A abordagem alternativa em questão se vale do agrupamento difuso feito sobre as amostras para propor uma maneira de reaproveitar a distribuição associada a cada elemento e então, definir um novo método.

Esta abordagem se baseia na regra dos k vizinhos mais próximos e define uma função que atribui graus de pertinência aos elementos do conjunto de treinamento, gerando uma distribuição de possibilidade associada a cada um deles. Assim como é feito por Keller *et al.* (1985), obtidas as distribuições de possibilidade, a classificação de um elemento z qualquer do conjunto de dados depende dos k elementos da amostra mais próximos de z . Define-se aqui uma técnica de obtenção dos graus de pertinência que determinam a compatibilidade de cada elemento com relação a cada classe. Desta maneira, a fase inicial é um agrupamento difuso e o algoritmo de classificação empregado é a regra do KNN.

A proposta é fazer uso dos graus produzidos por um algoritmo de agrupamento difuso quando aplicado a cada uma das classes. Dentre estes algoritmos, é sugerida a utilização do FCM, ou mesmo do FCM possibilista definido por Krishnapuram e Keller (1993). Ambos produzem uma matriz de pertinência contendo os graus de compatibilidade do elemento para cada um dos centros criados. O número de centros é um parâmetro que deve ser fornecido ao FCM. Dada a amostra da classe c_i , o número n de centros

é tal que $1 < n < |B_i|$, onde uma das formas de se obter n é através do índice de agrupamento *mis* definido na Seção 4.2.1.2. Nesta etapa é extraído um vetor v de valores de pertinência associados a cada classe. Este vetor corresponde a uma distribuição de probabilidade (ou possibilidade), onde cada v_{ij} está associado a um protótipo t_{ij} criado.

Na etapa seguinte é definida a função utilizada para extrair do vetor v_{ij} um único valor associado à cada elemento da classe. Primeiramente, v é normalizado de maneira que se obtenha uma distribuição de possibilidade v' ; isto para o caso em que graus provenientes do agrupamento, como no FCM, apresentem restrições probabilistas. Em seguida obtém-se o conjunto difuso com função de pertinência μ_i , associado a cada classe c_i , como sendo a agregação das probabilidades normalizadas usando uma t-conorma ou um operador simétrico, como a média OWA (vide Seção 2.3). Assim, tem-se:

$$\mu_i(z) = \pi_z(c_i) = \max(v'_{i1}(z), \dots, v'_{in}(z)) \quad (4.10)$$

onde n é o número de protótipos da classe, como definido anteriormente.

Isto define, para cada classe de maneira independente, uma distribuição de possibilidade normalizada. Na seção seguinte são definidos os tipos de normalização empregadas neste trabalho.

4.4 Normalização

Dentre os tipos de normalização descritos na literatura, destaca-se a transformação definida pela equação 4.11:

$$f(\delta(z)) = \delta(z) / \max(\delta(z)) \quad (4.11)$$

onde $\delta(z)$ é uma distribuição de possibilidade ou de probabilidade, e $f(\delta(x))$ é a distribuição normalizada.

Outra normalização utilizada é a transformação de probabilidade em possibilidade definida pela Equação 2.20, na Seção 2.5.2 deste documento, reescrita a seguir:

$$\forall i = 1, \dots, n, \pi_i = \sum_{j=i}^n p_j, \text{ onde } p_j \geq p_{j+1} \quad (4.12)$$

A normalização empregada neste trabalho é a transformação linear dada por:

$$f(\delta(z)) = (\delta(z) - \min(d(z)))/(\max(\delta(z)) - \min(\delta(z))) \quad (4.13)$$

Um exemplo bem simples é tal que, dada a distribuição de probabilidade $p = [0,3 \ 0,4 \ 0,2 \ 0,1]$, a distribuição de possibilidade π obtida pela normalização linear é $\pi = [0,6667 \ 1 \ 0,3333 \ 0]$ e pela transformação 4.12 é $\pi = [0,6 \ 1 \ 0,3 \ 0,1]$. A transformação 4.13 tem a particularidade de descartar as classes com o menor valor de pertinência.

CAPÍTULO 5

CLASSIFICAÇÃO IMPRECISA

A incerteza ou a imprecisão pode ser considerada um problema na informação quando se quer tomar uma decisão. Num sistema médico, por exemplo, a incerteza pode impedir que o melhor tratamento seja ministrado a um paciente ou contribuir para que a terapia incorreta seja indicada. Todavia, a informação do mundo real é intrinsecamente caracterizada pela imprecisão e o ideal é tratar de forma adequada esta característica, ou buscar ferramentas que possam tratá-la da melhor maneira. No processo de classificação, os objetos podem apresentar atributos cuja interpretação possa ser ambígua, duvidosa ou inexata, assinalando informação incompleta, incerta e imprecisa, o que pode acarretar uma classificação incorreta.

Entretanto, em diversas aplicações reais, uma classificação incorreta pode causar prejuízos, e a obtenção de uma resposta completamente imprecisa em lugar de um resultado preciso, porém incorreto, é preferível. Na tarefa de classificação, um sistema que aceita como saída uma resposta imprecisa pode ser visto como um sistema capaz de detectar os possíveis elementos mal definidos (*outliers*) dentro do conjunto de amostras apresentado, ou que não existe evidência suficiente para que uma classe precisa seja tomada como resposta.

Várias teorias são encontradas na literatura para lidar com a informação que envolve incerteza, imprecisão ou outras imperfeições. Dentre elas estão a teoria clássica de probabilidades, além da probabilidade Bayesiana e da teoria de Shannon; e ainda a teoria da evidência de Dempster-Shafer e a teoria dos conjuntos difusos de Zadeh. A informação de carácter probabilista pode ser tratada pela teoria de probabilidades ou pela teoria da evidência de Dempster-Shafer (SMETS, 1994); enquanto que a informação imprecisa ou vaga pode ser tratada pelas teorias dos conjuntos difusos (vide Capítulo 2) ou dos conjuntos de aproximação descritos em Pawlak (1982); e ainda a informação de carácter possibilista pode ser tratada pela teoria de possibilidades. Assim, a incerteza presente na informação pode ser modelada pelos modelos difusos, probabilistas, possibilistas ou de evidência.

O interesse aqui é atentar para as diferentes formas em que se apresentam os sistemas e focar na classificação imprecisa obtida a partir de uma classificação possibilista. Neste capítulo é descrito como se obter um classificador impreciso a partir de uma distribuição de possibilidade. Os classificadores descritos no Capítulo 4 retornam distribuições de possibilidade, tal qual outras abordagens descritas na literatura e referenciadas no

Capítulo 3. Uma distribuição de possibilidade pode ser obtida também a partir de classificadores que produzem distribuições de probabilidade como saída, aplicando uma transformação. Os classificadores difusos que têm por base o algoritmo FCM retornam distribuições com restrições probabilistas. A Seção 4.4, no capítulo anterior, trata dos tipos de normalização na obtenção de graus de pertinência associados aos elementos, de maneira que eles possam ser interpretados como os graus de possibilidade com que os elementos pertencem às classes, ou seja, a compatibilidade do elemento com os protótipos das classes.

Além disso, são definidos indicadores de qualidade de uma classificação imprecisa, empregados para auxiliar na tomada de decisão: dada uma família parametrizada de classificadores, eger uma boa classificação, avaliando precisão e confiança. Drummond e Sandri (2006) mostram um estudo da família parametrizada de classificadores imprecisos.

A definição de classe pura e composta é tida como se segue. Considerando os classificadores que atribuem um *subconjunto* de \mathbf{C} a um dado elemento $z \in R^n$, um *classificador impreciso* é um mapeamento $Cl : R^n \rightarrow 2^{\mathbf{C}}$: quando $|Cl(z)| = 1$, é dito que a classe é *pura*, e quando $|Cl(z)| > 1$ tem-se classe *composta*¹. Em outras palavras um classificador impreciso é um classificador que associa como saída a cada elemento z um subconjunto do conjunto de classes \mathbf{C} .

5.1 Família parametrizada de classificadores imprecisos

Dada a distribuição de possibilidade normalizada para um elemento z do conjunto de dados Z , uma família parametrizada de classificadores é obtida variando o parâmetro $\alpha \in [0, 1]$ de uma distribuição de possibilidade. O parâmetro α cria uma série de classificadores imprecisos, que variam do classificador mais correto e impreciso ao mais preciso, porém com maior taxa de erro.

Dada uma distribuição de possibilidade $\pi(z)$ obtida a partir de um classificador Cl para um elemento $z \in Z$, um classificador impreciso $Cl_\alpha, \alpha \in]0, 1]$ é dado por:

$$Cl_\alpha(z) = \{c \in \mathbf{C} \mid \pi(z)(c) \geq \alpha\} \quad (5.1)$$

Desta forma, a classe definida como saída para o elemento z é imprecisa, ou seja, a classe de z é dada pelo corte de nível de grau α .

¹Em classificadores do tipo árvore, classes compostas são conhecidas como classes *impuras* (KUNCHEVA, 2004).

5.2 A matriz de confusão imprecisa

Seguindo a definição de matriz de confusão apresentada na Seção 3.4.1.1, para uma classificação imprecisa é necessário uma matriz de confusão imprecisa, que contém m linhas e 2^m colunas. Ou seja, número de linhas igual ao número de classes do problema e uma coluna para cada subconjunto do conjunto de classes \mathbf{C} , incluindo o conjunto vazio. A matriz correspondente ao classificador Cl_α pode ser representada como na Tabela 5.1, onde as entradas $a_{i,i}$, $i = 1, \dots, m$ contêm o número de elementos perfeitamente classificados, ou seja, têm classificação correta e precisa. A última coluna da matriz corresponde ao número de elementos cuja classificação é correta porém completamente imprecisa, ou seja, $Cl(z) = \mathbf{C}$. Já a primeira coluna, indicada por \emptyset , corresponde ao número de elementos para os quais o classificador não atribuiu nenhuma classificação. Trata-se de dois casos distintos, a primeira coluna modela conflito total e a última modela mesma evidência para todas as possibilidades.

Tabela 5.1 - Matriz de confusão imprecisa para o classificador Cl_α num problema com m classes

c^* / Cl_α	\emptyset	$\{c_1\}$...	$\{c_m\}$	$\{c_1, c_2\}$...	$\{c_{m-1}, c_m\}$...	$\{c_1, \dots, c_m\}$
c_1	$a_{1,0}$	$a_{1,1}$...	$a_{1,m}$	$a_{1,12}$...	$a_{1,[m-1]m}$...	$a_{1,1\dots m}$
...
c_m	$a_{m,0}$	$a_{m,1}$...	$a_{m,m}$	$a_{m,12}$...	$a_{m,[m-1]m}$...	$a_{m,1\dots m}$

Para ilustrar a matriz de confusão imprecisa, seja um problema com 3 classes $\mathbf{C} = \{c_1, c_2, c_3\}$. A imprecisão completa é dada pelo somatório da coluna c_{123} , na Tabela 5.2. E são ainda verificadas as seguintes medidas:

- Erro preciso: $a_{2,1} + a_{3,1} + a_{1,2} + a_{3,2} + a_{1,3} + a_{2,3}$
- Erro impreciso: $a_{3,12} + a_{2,13} + a_{1,23}$
- Acerto impreciso: $a_{1,12} + a_{2,12} + a_{1,13} + a_{3,13} + a_{2,23} + a_{3,23}$

Tabela 5.2 - Matriz de confusão imprecisa para o classificador Cl_α num problema com 3 classes

c^* / Cl_α	\emptyset	c_1	c_2	c_3	c_{12}	c_{13}	c_{23}	c_{123}
c_1	$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,12}$	$a_{1,13}$	$a_{1,23}$	$a_{1,123}$
c_2	$a_{2,0}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{2,12}$	$a_{2,13}$	$a_{2,23}$	$a_{2,123}$
c_3	$a_{3,0}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,12}$	$a_{3,13}$	$a_{3,23}$	$a_{3,123}$

Esta mesma avaliação de uma classificação imprecisa é feita através dos indicadores de qualidade definidos na próxima seção.

Um exemplo numérico é também apresentado, usando o conjunto de dados sintético **cone-torus**. A Figura 5.1 apresenta uma comparação entre dois classificadores imprecisos, Cl_1 e $Cl_{0,4}$. Os pontos correspondem aos elementos classificados como classe pura, ou seja, pontos precisos que têm associados apenas uma classe; os de cor verde são corretos e os azuis são incorretos; os círculos correspondem aos elementos cuja classe atribuída pelo classificador é composta; são imprecisos, porém não *completamente* imprecisos. Os triângulos azuis indicam imprecisão completa, ou seja, todas as classes são associadas ao elemento.

Foi empregado neste exemplo o classificador AGRUP com 5 centros associados à classe c_1 e c_2 , e 10 centros associados à classe c_3 . A matriz de confusão imprecisa que sintetiza os resultados para os classificadores Cl_1 e $Cl_{0,4}$ é apresentada na Tabela 5.3.

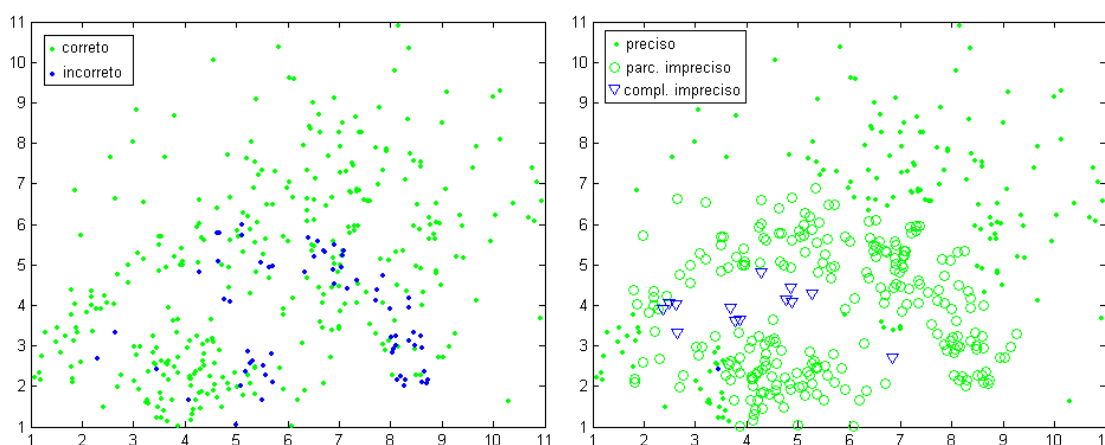


Figura 5.1 - Classificação precisa (Cl_1) e classificação imprecisa ($Cl_{0,4}$) para o conjunto de treinamento cone-torus (KUNCHEVA, 2000) (Número de centros associados a cada classe $\langle 5, 5, 10 \rangle$).

Tabela 5.3 - Matriz de confusão imprecisa para os classificadores a) preciso Cl_1 e b) impreciso $Cl_{0,4}$.

		c^* / Cl_1	\emptyset	c_1	c_2	c_3	c_{12}	c_{13}	c_{23}	c_{123}
a)	c_1	0	83	2	7	0	0	0	0	
	c_2	0	0	60	39	0	0	0	0	
	c_3	0	8	8	193	0	0	0	0	
		$c^* / Cl_{0,4}$	\emptyset	c_1	c_2	c_3	c_{12}	c_{13}	c_{23}	c_{123}
b)	c_1	0	19	0	0	6	63	0	4	
	c_2	0	0	14	0	4	0	78	3	
	c_3	0	1	0	115	0	12	75	6	

5.3 Índices de acurácia e confiança do classificador impreciso

Uma coleção de índices de qualidade é definida, para que, a partir de uma família parametrizada, seja possível eleger o classificador que apresente o nível de imprecisão e confiança mais adequado para uma dada aplicação. Os índices são derivados da partição do conjunto de dados Z segundo a precisão e a cardinalidade κ .

Seja $c^*(z)$ ou, simplesmente c^* a classe correta do elemento $z \in Z$. O conjunto $Z^{*\kappa}$ é definido como:

$$Z^{*\kappa} = \{z \in Z \mid c^* \in Cl(z) \wedge |Cl(z)| = \kappa\}. \quad (5.2)$$

e conseqüentemente, os elementos úteis do conjunto Z são dados por:

$$Z^* = \bigcup_{\kappa < |\mathbf{C}|} Z^{*\kappa}. \quad (5.3)$$

Os conjuntos Z_{err} e Z_{imp} contêm os elementos de Z classificados como incorretos e imprecisos, respectivamente. Note que os conjuntos Z^* , Z_{err} e Z_{imp} são mutuamente exclusivos.

Os *índices de qualidade* para uma classificação imprecisa seguem a concepção estabelecida para a acurácia precisa definida pela Equação 3.16, e são dados por:

$$qual^*(Cl, Z) = |Z^*|/|Z| \quad (5.4)$$

$$qual^{*\kappa}(Cl, Z) = |Z^{*\kappa}|/|Z| \quad (5.5)$$

$$err = |Z_{err}|/|Z| \quad (5.6)$$

$$imp = |Z_{imp}|/|Z| \quad (5.7)$$

Os índices err e imp exprimem as medidas de erro e imprecisão completa, respectiva-

mente. O índice $qual^{*\kappa}$ com $\kappa = 1$ define a medida de acerto preciso e a imprecisão parcial é dada pela soma dos $qual^{*\kappa}$, ou seja, $\sum_{\kappa>1}^{\kappa<m} qual^{*\kappa}$. Tem-se, então, que $qual^* + err + imp = 1$.

Medidas de incerteza podem também ser calculadas a partir de uma distribuição de possibilidade $\pi(\cdot)$. Supondo que $\pi(c), c \in \mathbf{C}$, modela a possibilidade de que a classe correta para o ponto z é c , ou seja, que $c^* = c$, então, para medir a confiança de que c^* pertence ao conjunto $A \subseteq \mathbf{C}$, tem-se as medidas:

$$\Pi(A) = \max_{u \in A} \pi(u) \quad (5.8)$$

$$N(A) = 1 - \Pi(\bar{A}) = \min_{u \notin A} 1 - \pi(u) \quad (5.9)$$

$$N\Pi(A) = (\Pi(A) + N(A))/2 \quad (5.10)$$

$$P(A) = \sum_{u \in A} p(u) \quad (5.11)$$

onde Π e N são as medidas de possibilidade e necessidade, respectivamente, apresentadas na seção 2.5, e $N\Pi$ é a média de N e Π ; P é a medida de probabilidade derivada da distribuição de probabilidade p . Cada p é obtido a partir de π , aplicando-se a transformação definida pela Equação 2.19. Fazendo-se $A = Cl(z)$ estas medidas podem ser vistas como medidas de confiança do classificador.

A acurácia individual da classificação para um ponto z , dada uma classificação imprecisa Cl_α , é definida como:

$$acc(Cl_\alpha, z) = \begin{cases} 1, & \text{se } c^* \in Cl_\alpha(z) \\ 0, & \text{senão} \end{cases} \quad (5.12)$$

Os índices individuais de acurácia baseados em incerteza Π_{acc} , N_{acc} , P_{acc} e $N\Pi_{acc}$, para $\pi(z)(\cdot)$, são definidos por:

$$f_{acc}(Cl_\alpha, z) = f(Cl(z)) \times acc(Cl_\alpha, z) \quad (5.13)$$

com f substituído por Π , N , P ou $N\Pi$

Da mesma maneira, para o conjunto de dados Z , os índices de acurácia baseados em incerteza Π -, N -, P - e $N\Pi$, que medem a *confiança* do classificador impreciso, são dados por:

$$f_{acc}(Cl_\alpha, Z) = \sum_{z \in Z} f_{acc}(Cl_\alpha, z) / |Z| \quad (5.14)$$

substituindo f por Π , N , P e $N\Pi$ respectivamente. Note que $N_{acc}(Cl_\alpha, Z) \leq P_{acc}(Cl_\alpha, Z)$, $N\Pi_{acc}(Cl_\alpha, Z) \leq \Pi_{acc}(Cl_\alpha, Z)$, $\forall \alpha$.

A partir dos índices definidos, são derivados parâmetros que servem de base para análise de uma classificação imprecisa considerando uma família de classificadores imprecisos Cl_α :

- **a**: indica o número de pontos perfeitamente classificados para $\alpha = 1$. Note que, para qualquer Z , $qual^*(Cl_1, Z) = qual^{*1}(Cl_1, Z)$ e $qual^*(Cl_1, Z) = 0, \forall \kappa \neq 1$.
- **b**: refere-se ao número de pontos perfeitamente classificados independente do valor de α , ou seja, pontos cuja distribuição de possibilidade é um conjunto difuso preciso.
- **c**: indica o menor valor de α para o qual todos os pontos corretamente classificados dentro do conjunto de dados são considerados úteis, ou seja, apresentam classificação correta e não completamente imprecisa.
- **d**: indica o valor de α que produz o maior valor para $qual^*$.
- **e**: indica o menor valor de α para o qual Cl_α produz classificação incorreta.
- **u**: refere-se a medidas de probabilidade obtidas para a classificação precisa, $P_{acc}(Cl_1, Z)$.

A Figura 5.2 mostra os gráficos de classificação baseados em imprecisão e em incerteza, indicando os parâmetros de base descritos acima. Os gráficos são obtidos para 101 valores de $\alpha \in [0, 1]$, e se referem ao exemplo descrito na Seção 5.2.

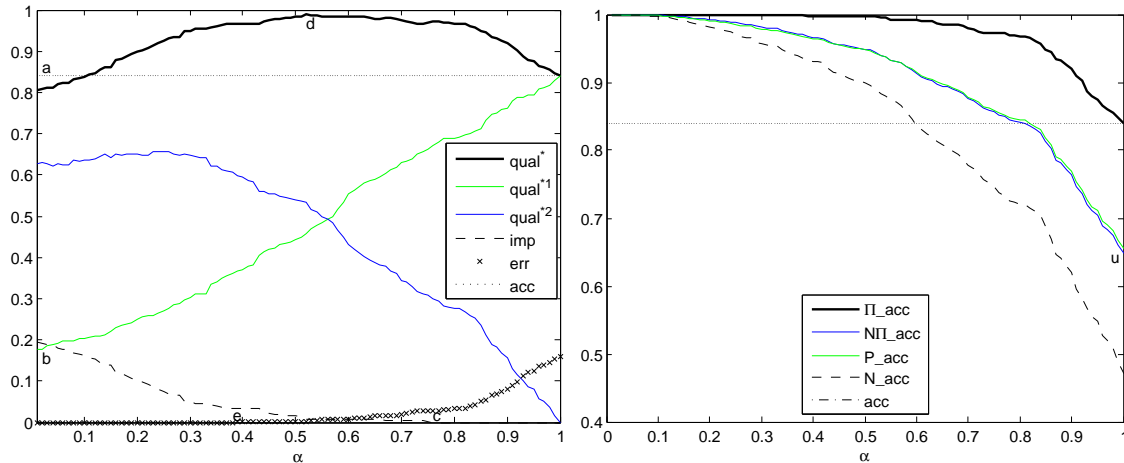


Figura 5.2 - Índices de classificação de Cl_α para o conjunto de treinamento (número de centros por classe: $\langle 5, 5, 10 \rangle$): a) baseados em imprecisão, b) baseados em incerteza.

5.4 Medida de utilidade de imprecisão

A utilidade de imprecisão é uma medida simples que avalia o quanto, além da precisão existente, o sistema ganha com a imprecisão correta. A imprecisão correta agrega tanto mais valor à medida quando menor for sua cardinalidade. A medida penaliza o erro preciso, e penaliza ainda mais o erro impreciso, fazendo com que os elementos classificados incorretamente tenham efeito tanto mais negativo quanto maior for sua imprecisão. Ou seja, quanto maior a cardinalidade de um elemento que tem classificação incorreta e imprecisa, maior é a penalidade aplicada à utilidade de imprecisão. A seguir, a equação que modela a medida:

$$utilidade = \sum_{\kappa=1}^m (qual^{*\kappa} \times ((m - \kappa)/(m - 1))) \quad (5.15)$$

onde κ é a cardinalidade, m o número de classes e $qual^{*\kappa}$ o índice de qualidade definido na Seção 5.3. Esta medida é apenas uma das formas de se ponderar a imprecisão.

5.5 Agregação de classificadores imprecisos

A agregação de dois ou mais classificadores, independentemente do modelo adotado, pode levar a resultados melhores do que aqueles obtidos pelos classificadores individuais. Uma breve descrição da combinação de classificadores é apresentada na Seção 3.3. A agregação pretendida neste trabalho pode ser classificada como no nível de classificadores, segundo as abordagens definidas por Kuncheva (2004), onde são

combinadas classificações imprecisas provenientes de diferentes modelos de classificação.

A agregação de dois classificadores parametrizados, $Cl_{\alpha_1}^1$ e $Cl_{\alpha_2}^2$ sobre o conjunto de dados Z é dada por:

$$Cl_{\alpha_1, \alpha_2}^{12}(z) = f(Cl_{\alpha_1}^1(z), Cl_{\alpha_2}^2(z)) \quad (5.16)$$

onde f é uma função de agregação e $\alpha \in [0, 1]; \alpha_1 \neq \alpha_2$. Uma função que pode ser empregada é $f = \cap$, sendo que se $Cl_{\alpha_1, \alpha_2}^{12}(z) = \emptyset$, o elemento z é classificado como completamente impreciso. A união, $f = \cup$, também pode ser utilizada, porém a classificação obtida é demasiadamente imprecisa. Uma outra função de agregação é a função híbrida, $f = \cap'$, que representa o compromisso entre a união e a intersecção, e é definida por:

$$\cap' (Cl_{\alpha_1, \alpha_2}^{12})(z) = \begin{cases} \cap(Cl_{\alpha_1}^1, Cl_{\alpha_2}^2)(z), & \text{se } \cap(Cl_{\alpha_1}^1, Cl_{\alpha_2}^2)(z) \neq \emptyset \\ \cup(Cl_{\alpha_1}^1, Cl_{\alpha_2}^2)(z), & \text{senão} \end{cases} \quad (5.17)$$

5.6 Heurísticas de decisão

As heurísticas de decisão definidas neste trabalho são regras ou métodos que levam à classificação final precisa de um conjunto de dados, dada a classificação imprecisa Cl derivada a partir de uma distribuição de possibilidade π . A classificação final é definida analisando a classificação imprecisa em si e, também, através da sua vizinhança no domínio do problema.

A análise da classificação imprecisa leva em conta a distribuição de possibilidade que a originou, trabalhando diretamente sobre os valores de pertinência associados. A classe pura para cada elemento é determinada verificando quais são as classes com máxima pertinência.

O refinamento pela vizinhança é uma abordagem descrita em Malinski (2000), que apresenta uma técnica de pós-processamento de imagens, através da qual as classes compostas, resultantes do processo de classificação, são substituídas por classes puras de acordo com a vizinhança de cada elemento no espaço do problema (neste caso, de cada pixel da imagem) (vide classificador possibilista descrito na Seção 3.2.2).

5.6.1 Heurística do valor máximo: n-MAX

O critério mais simples, porém muito dependente da distribuição de possibilidade derivada do processo de classificação, é o emprego do operador max, que associa a cada elemento z a classe para a qual o grau de pertinência é máximo. Logo, tem-se $\pi_{(z)}(c_i) = \max(\pi_{ij}(z))$, e que a classe do ponto é dada por $c(z) = \arg \max_{c_i \in \mathbf{C}} \pi_i(z)$.

Dois situações são destacadas: (i) o grau de pertinência de um elemento pode ser 0 para todas as classes, assim $\max(\pi_{ij}(z)) = 0$ e (ii) o grau de pertinência de um elemento para a classe c_1 pode ser igual àquele calculado para a classe c_2 . Na situação (i), a classe do elemento é dada como indefinida e pode caracterizar um ponto mal definido no domínio do problema (*outlier*), ou ainda um espaço de atributos que não reflete as características reais do problema. Já na segunda situação, que acontece quando o elemento dista igualmente de protótipos pertencentes a classes distintas, a classe do elemento é composta.

Para $n = 1$ a heurística 1-MAX é tal que a classe do elemento é *indefinida* se $\max(\pi_i(z)) = 0$, *pura* para o caso em que $|\arg \max(\pi_i(z))| = 1$ e *composta* quando $|\arg \max(\pi_i(z))| > 1 \wedge \max(\pi_i(z)) \neq 0$.

Para $n = 2$ a heurística 2-MAX, são considerados os dois graus máximos. Sejam $t_{a,b}$ e $t_{c,d}$, $1 \leq a, c \leq n$, $1 \leq b, d \leq \kappa$ os centros referentes às maiores possibilidades em $\pi(z)$, e que $\pi(z)(t_{a,b}) \geq \pi(z)(t_{c,d})$. Definidos os parâmetros arbitrários $\beta \in R$, $\beta > 1$ e $\delta \in]0, 1]$, o seguinte critério de classificação é definido:

Sejam os valores arbitrários $\beta \in R$, $\beta > 1$ e $\delta \in]0, 1]$

if $a = c$ **then**

$c(z) = \textit{pura} = c_a$

else if $\pi(z)(t_{c,d})(t_{a,b}) \geq \beta * \pi_z(t_{a,b})$ **then**

$c(z) = \textit{indefinida}$

else if $\pi_z(t_{c,d})/\pi_z(t_{a,b}) \geq \delta$ **then**

$c(z) = \textit{composta} = \{c_a, c_c\}$

else

$c(z) = \textit{pura} = c_a$

end if

5.6.2 Refinamento pela vizinhança

Basicamente, o algoritmo de pós-processamento investigado por Malinski (2000) consiste em utilizar uma função que contabilize o número de classes puras contidas nas classes (eventualmente compostas) dos vizinhos de um dado pixel. A vizinhança considerada é 8-conectada, ou seja, são verificadas as classes dos 8 vizinhos mais próximos de cada elemento. O nível de indefinição ϕ de um conjunto classificado é definido com o maior valor de cardinalidade dentre as classes compostas pertencentes a $Z_{\#}$ ($Z_{\#}$ indica o conjunto Z classificado em termos de classes puras e compostas). Formalmente, tem-se que $\phi(c) = \sup\{|c_j| \mid \exists c(z) \in Z_{\#}, c(z) = c_j\}$.

A técnica é tal que seja $\psi(c_j)$ o número de vizinhos do ponto z que estão classificados como c_j . Assim, para uma vizinhança 8-conectada, $\forall j, 0 \leq \psi(c_j) \leq 8$ e $\sum \psi(c_j) = 8$. A classe ϵ , candidata à reclassificação do elemento z , é dada por:

$$\epsilon = \cup\{c_j \in \mathbf{C} \mid \phi(c_j) = \sup_j \phi(c_j)\}$$

Sendo que a classe atual do elemento será substituída somente por classes de cardinalidade igual ou superior ao nível de indefinição. Para um nível de indefinição ϕ qualquer, o processo é repetido até que nenhum elemento tenha sua classe substituída. Um segundo algoritmo é definido para resolver os pixels com indefinição. Em última instância, se o pixel ainda pertencer a uma classe composta, é reclassificado drasticamente de forma aleatória.

Um outro emprego desta mesma abordagem seria tal que, o refinamento é aplicado diretamente às distribuições de possibilidade $\pi(z)$ obtidas para cada elemento. Desta maneira, o processo considera os graus de possibilidade da vizinhança, e não somente se o ponto pertence ou não a uma determinada classe, como descrito acima. A proposta se mostra interessante, porém a avaliação das distribuições da vizinhança de cada elemento, através de uma função que pondere as pertinências, pode ser muito custosa.

CAPÍTULO 6

APLICAÇÕES

Neste capítulo são apresentados os resultados obtidos com a aplicação de estratégias de classificação possibilista e difusa a diferentes conjuntos de dados. Dentre as aplicações, está a classificação de imagens médicas e de sensoriamento remoto, dados que representam o registro de queda na distribuição de energia em subestações elétricas e dados do Projeto GeoSchisto, relacionados à distribuição da esquistossomose no estado de Minas Gerais. Cada aplicação é detalhada em uma das subseções seguintes.

A metodologia seguida para avaliação dos modelos de classificação, bem como da utilização dos índices propostos para avaliação e combinação de classificadores pode ser generalizada na Figura 6.1.

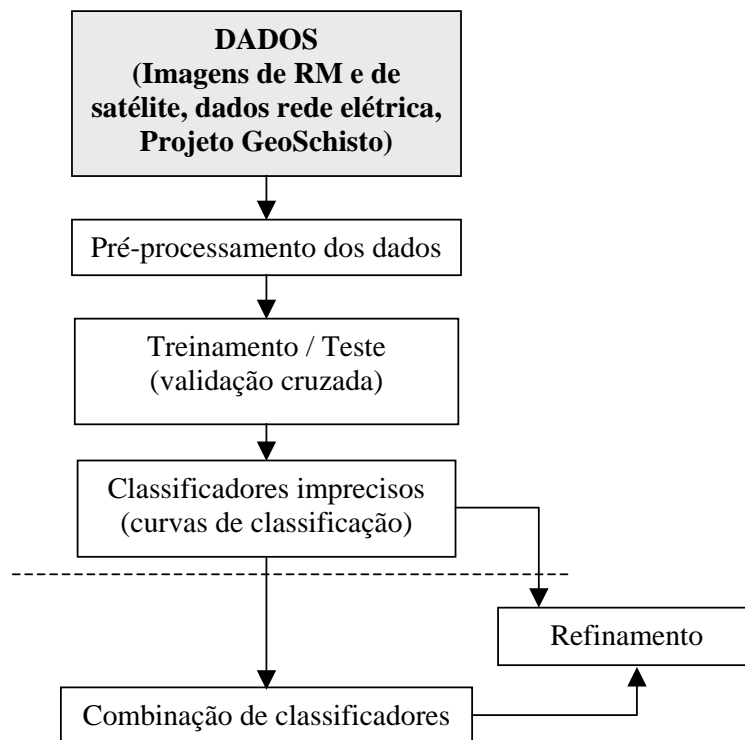


Figura 6.1 - Metodologia.

Na fase de pré-processamento dos dados, os conjuntos são normalizados para a faixa de valores $[0, 1]$, utilizando a transformação linear dada pela Equação 4.13.

É avaliado o modelo de classificação proposta AGRUP. O classificador FKNN (vide

seção 3.2.5.1) também é empregado, tanto para fins de comparação, quanto para agregação na etapa de combinação dos classificadores. As estratégias escolhidas para os experimentos são submetidas aos mesmos conjuntos de teste e treinamento, definidos pelo método de validação cruzada (vide seção 3.4.2.2). Ou seja, o conjunto total de amostras é dividido em L subamostras, sendo o classificador treinado L vezes nos $L - 1$ conjuntos, e testado no conjunto restante.

As curvas de classificação compõem os gráficos de classificação imprecisa, variando-se o parâmetro $\alpha \in [0, 1]$, obtidas a partir dos índices definidos na Seção 5.3. O estudo da classificação imprecisa é realizado também através de tabelas que sintetizam os resultados obtidos, apresentando os valores de precisão, erro, imprecisão parcial e completa, resultante da aplicação de cada classificador. O desempenho de cada técnica é mostrado com variações dos conjuntos de teste (Te) e treinamento (Tr), para 11 valores de α , indicados na primeira linha das tabelas.

A combinação de classificadores é feita com base na avaliação dos níveis de precisão, imprecisão parcial e completa, e erro obtido para a família parametrizada de classificadores imprecisos. A agregação de dois classificadores permite a criação das superfícies correspondentes aos índices de classificação imprecisa. Na etapa de refinamento, ao final, é empregada uma heurística que leva à classificação precisa, e seu desempenho é medido através da taxa de acerto do classificador, e também pelo cálculo do índice *kappa*, ambos discutidos na seção 3.4.

Basicamente, esta é a metodologia seguida para os experimentos realizados, tendo alguma variação dependendo do conjunto de dados considerado. As modificações são descritas nas seções correspondentes. Os procedimentos referentes ao projeto foram implementados em ambiente MATLAB.

A seção que se segue apresenta uma breve análise no que diz respeito à classificação de imagens, apresentando as características específicas das imagens médicas e de sensoriamento remoto. Logo após, é apresentada a seção de experimentos, subdividida por aplicações, onde são descritos os resultados relevantes na discussão e análise.

6.1 Classificação de imagens

Segundo Gonzalez e Woods (2000) a análise de imagens é um processo de descobrimento, de identificação e de entendimento de padrões que sejam relevantes à performance de uma tarefa baseada em imagens. Uma das principais metas de análise de imagens por computador é dotar a máquina com uma performance de

aproximar a capacidade similar dos seres humanos, como por exemplo, a leitura automática de imagens de documentos datilografados. Neste caso, os padrões de interesse são caracteres alfanuméricos, e se tem como meta fazer com que a precisão do reconhecimento de caracteres seja o mais próximo possível daquela que os seres humanos são capazes de realizar. Assim, um sistema de análise automática de imagens deveria ser capaz de exibir vários graus de inteligência, como a habilidade de extrair informações pertinentes a partir de um fundo de detalhes irrelevantes; a capacidade de aprender a partir de exemplos e de generalizar o conhecimento de maneira que ele possa ser aplicado em circunstâncias novas e diferentes; e a habilidade de fazer inferências a partir de informação incompleta.

Projetar e implementar sistemas de análise de imagens, providos de desempenho próximo da capacidade humana de realizar funções de análise de imagens, ainda é um problema em aberto. Assim, novas e promissoras teorias vêm sendo descobertas com as pesquisas em sistemas biológicos e computacionais, porém a maior parte da análise computadorizada de imagens é baseada em formulações heurísticas projetadas para resolver problemas específicos. Desta forma, existem vários métodos, técnicas e sistemas que possuem pouca ou nenhuma capacidade de generalização. Como existem limitações teóricas, as soluções tendem a ser dependente do problema.

Para Gonzalez e Woods (2000) a análise de imagens envolve: (i) aquisição e pré-processamento de imagem, (ii) extração e caracterização de componentes e (iii) reconhecimento e interpretação da imagem. Em (i) são consideradas as atividades de formação da imagem e compensações, como redução de ruído e borramento da imagem, sendo tarefas de baixo nível. Já em (ii) estão incluídas as atividades de segmentação e descrição, onde o comportamento inteligente deve estar presente em procedimentos de segmentação. E por fim, em (iii) trata-se de processamento de alto nível onde menos formulações teóricas são definidas para reconhecer e interpretar uma imagem, uma vez que estas atividades envolvem menor precisão e maior investigação. Os conceitos predominantes por trás das metodologias de interpretação e reconhecimento de imagens são a organização efetiva e uso de conhecimento sobre o domínio do problema, onde se encontram os sistemas especialistas, entre outros tipos de sistemas.

A classificação de uma imagem consiste em identificar diferentes feições que apresentam padrões espectrais similares, e atribuí-los a uma determinada classe. Várias técnicas podem ser usadas para realizar a classificação, como heurísticas, estatísticas, redes neurais, morfologia matemática, agrupamento, entre outras.

6.1.1 Imagens de RM

Imagem de ressonância magnética (MRI - *Magnetic Resonance Image*), em Marques (1998), Jain (1989), Shapiro e Stockman (2001), constitui uma tecnologia não invasiva que adquire imagens bidimensionais e tridimensionais, que permitem visualizar o interior do corpo humano, bem como os tecidos orgânicos, entre outras aplicações. O princípio fundamental do método baseia-se em impor condições que causam o fenômeno físico da ressonância magnética sobre a região que se deseja analisar. Esta área é submetida a um campo magnético específico, que se mantém constante, e sinais de rádio frequência (RF) com a frequência específica associada. Introduce-se também gradientes de campos magnéticos para codificar espacialmente as unidades de volume da área em estudo. A ressonância magnética implica a absorção de energia eletromagnética de um sinal de RF e a sua conseqüente dissipação, sendo este o sinal de interesse.

A imagem obtida é construída a partir de vários cortes e conseqüente processamento de imagens, quando se trata de um volume. Geralmente, estas imagens são tridimensionais representadas pelo vetor $I[s, r, c]$, onde s indica a fatia do corpo e r e c são as coordenadas da imagem.

As imagens de RM consistem numa importante ferramenta para diagnóstico em várias áreas da medicina, como a neurologia e a oncologia. Embora um sistema de RM seja capaz de distinguir diferentes tipos de moléculas nos tecidos, os sistemas têm sido empregados para imagear a distribuição de moléculas de água (H_2O), que constituem mais de 80% do peso total do corpo humano. As moléculas de H_2O estão orientadas aleatoriamente no tecido. Sob a influência de um campo magnético forte, todas as moléculas de H_2O se orientam na direção do campo magnético e giram em uma frequência angular específica. Tal frequência, conhecida como frequência angular de Larmor, é diretamente proporcional à força do campo magnético. O campo magnético requerido para diagnóstico nos sistemas de RM é de 1 Tesla, que corresponde a 10.000 vezes o campo magnético da Terra. Uma vez que as moléculas estão alinhadas pelo campo magnético, energia adicional é introduzida através de um transmissor de RF. A onda eletromagnética gerada pelo transmissor é polarizada e ajustada às frequências de ressonância das moléculas. Uma pequena explosão da onda eletromagnética de RF é enviada, empurrando as moléculas de água para fora do seu eixo original. Enquanto as moléculas retornam para sua orientação, energia é liberada também na forma de sinal de RF, sendo a magnitude do sinal recebido proporcional à quantidade de moléculas de água. O tempo necessário para que as moléculas possam recuperar sua orientação é a constante T_1 , que pode ser extraída de uma seqüência especial de pulsos. O sinal

de RF recebido mostra um decaimento exponencial e essa constante de decaimento, chamada de T_2 está relacionada à fase. Enquanto as moléculas de água gradualmente recuperam sua orientação, energia de RF é liberada, resultando no cancelamento do sinal recebido. O decaimento T_2 está relacionado à não homogeneidade das moléculas de água dentro do ambiente que as envolve.

A informação espacial na imagem de RM é codificada aplicando um gradiente de campo magnético pequeno através do plano da imagem. Cada ponto do plano tem um único campo magnético correspondendo a uma única frequência de ressonância. Assim, as imagens de RM podem ser reconstruídas a partir da transformada de Fourier do sinal de RF recebido.

São definidos três parâmetros de controle no sistema de RM: (i) TR que corresponde ao tempo de repetição da seqüência de pulsos de RF; (ii) TE que é o tempo de decaimento (“spin-echo”) e (iii) TI que é tempo de inversão. A variação destes parâmetros gera diferentes imagens como as especificadas a seguir:

- a) Piloto: tempos pequenos para aquisição rápida de imagens piloto.
- b) T1: adquirida quando se emprega TR pequeno (<800ms) e TE pequeno (<20ms). Devido aos tempos de aquisição serem pequenos, as imagens ponderadas em T1 tendem a conter pouco ruído.
- c) T2: adquirida quando se emprega TR grande (>1500ms) e TE grande (>70ms). As imagens ponderadas em T2 tendem a ser mais ruidosas que as ponderadas em T1 e PD.
- d) PD: combina os tempos TR e TE, TR relativamente grande e TE pequeno.
- e) Flair: neste caso, TE e TR são fixados como em T2, porém um tempo de inversão TI não nulo é empregado. Este tipo de imagem é usada para destacar tecidos com determinadas propriedades. Por exemplo, um tempo de inversão de aproximadamente 2000 faz com que a água seja imageada na cor escura, o que se opõe à imagem ponderada em T2, onde o líquido é claro.

A Tabela 6.1 mostra valores de parâmetros usualmente empregados para gerar as diferentes imagens num sistema de RM.

Considerando a aquisição de imagens do cérebro, são definidas três visões (Figura 6.2): axial (da base para o topo, a metade direita da imagem corresponde ao lado esquerdo da

Tabela 6.1 - Parâmetros usualmente empregados num sistema de RM

	Piloto	T1			T2			PD		Flair	
TE	15	6	14	20	80	96	100	20	32	80	6
TR	35	500	550	800	2000	3555	3826	2000	2100	6677	2321
TI	0	0	0	0	0	0	0	0	0	2000	600



Figura 6.2 - Cortes de uma imagem de RM. *Adaptado de Pleet e Hicks (2004)*

cabeça do paciente); sagital (corresponde aos cortes de lado) e coronal (corresponde aos cortes da frente para a parte de trás). Essas imagens são chamadas de fatias, ou cortes, que tem uma espessura definida no momento da obtenção da imagem. São necessárias várias fatias para se obter a imagem em 3 dimensões, existindo *gaps* entre os cortes definidos.

Foram empregados dois conjuntos de imagens nos testes realizados, uma base de dados sintética e uma real, ambas de imagens do cérebro contendo lesões de esclerose múltipla (EM).

6.1.2 Imagens de sensoriamento remoto

Imagens de sensoriamento remoto têm servido de fonte de dados para estudos geológicos, ambientais, agrícolas, cartográficos, florestais, urbanos, oceanográficos, etc. Representam, atualmente, uma das poucas formas viáveis de monitoramento ambiental em escalas locais e globais devido à rapidez, eficiência, periodicidade e visão sinóptica que as caracterizam.

As características registradas em imagens de sensoriamento remoto estão vinculadas ao espectro eletromagnético usado pelos sensores, que podem ser classificados de acordo

com a região do espectro em que operam. Destacam-se dois grupos de sensores: passivos (óticos) ou ativos. Os passivos são aqueles que necessitam de uma fonte externa de iluminação e operam na faixa visível do infravermelho e os ativos produzem sua própria energia e operam na faixa de microondas.

Destacam-se aqui as imagens óticas, como aquelas produzidas pelo sensor CCD do satélite CBERS (CBERS, 2004) e pelo sensor TM do satélite LANDSAT5 (LANDSAT, 2004).

6.2 Experimentos

Os resultados apresentados a seguir são gerados a partir da aplicação das estratégias AGRUP e FKNN. Para o algoritmo AGRUP foram empregadas as relações de similaridade $S_{\gamma_{tr1}, \lambda_{tr1}}^L$, $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ e $S_{\gamma_{g1}, \lambda_{g1}}^G$, apresentadas na Seção 4.2.1.3. Para as relações lineares existem dois parâmetros de desempenho que podem ser ajustados: mm para o suporte λ_{tr1} e nn para o núcleo γ_{tr1} . Para todos os experimentos realizados foi empregado $mm = 1,5$ e $nn = 0,8$. Na fase inicial deste algoritmo, para a maioria das aplicações, foi empregado o meta-agrupamento global, definido na Seção 4.2.1.2. O número de centros para cada classe, referido como configuração inicial, em cada um dos experimentos executados, é fornecido no seguinte formato: $[T_{expL} < |t_{ij}| >]$, onde L é um dos 5 experimentos e $|t_{ij}|$ é o número de centros i da classe j .

Além da avaliação individual, para alguns casos, é feito um estudo da classificação obtida a partir da agregação de dois classificadores. E, para os experimentos com as imagens de satélite, o refinamento é empregado como forma de transformar classes compostas em classes puras e obter a classificação precisa final.

A partir dos dados normalizados, a validação cruzada é feita com 5 conjuntos, ou seja, para as aplicações que são submetidas a esta avaliação, o seu conjunto de amostras é dividido em 5 subconjuntos, sendo feitas 5 execuções, cada qual com 4 conjuntos no treinamento e 1 no teste.

Em cada uma das tabelas que se seguem, a primeira linha indica os valores de α , que corresponde ao parâmetro da família de classificadores imprecisos. E as linhas subseqüentes, Tr1 e Te1, correspondem às medidas referentes ao conjunto de treinamento e teste, respectivamente, para o experimento 1. Da mesma maneira para os experimentos 2, 3, 4 e 5. As tabelas sintetizam: medidas de precisão, que correspondem ao índice $qual^{*1}$; medidas de erro (índice err); medidas de imprecisão parcial (índice $qual^{*\kappa}$) e medidas de imprecisão completa (índice imp), definidos como índices de

qualidade na Seção 5.3.

Os gráficos de classificação têm suas curvas traçadas para 101 valores do parâmetro de corte $\alpha \in [0, 1]$, variando no eixo x. As curvas (no eixo y) representam os índices de qualidade.

6.2.1 Imagens de RM do cérebro com EM - Banco de dados sintético

O banco de dados simulado utilizado neste trabalho é o SBD (*Simulated Brain Database*) disponível em BrainWeb (2006). Os volumes de RM do cérebro apresentados são produzidos por um simulador de imagem de RM, baseado em dois modelos anatômicos: um cérebro normal e um cérebro com esclerose múltipla (EM). Os volumes de dados em 3D são simulados em três seqüências T1, T2 e PD, sendo possível variar a espessura de cada corte, o nível de ruído aditivo e o nível de ruído de intensidade não uniforme, que é definida como uma variação indesejada do sinal ao longo da imagem. Os volumes são gerados a partir de um modelo anatômico, que pode ser observado na Figura 6.3a, contendo 11 classes, representadas como: 0=fundo, 1=CSF (*Cerebral Spinal Fluid*), 2=substância cinza, 3=substância branca, 4=gordura, 5=músculo/pele, 6=pele, 7=caixa craniana, 8=substância glial, 9=conexão e 10=lesão de EM.

As estratégias de classificação foram aplicadas a alguns dos cortes pertencentes aos volumes simulados. Das 11 classes presentes no modelo são consideradas apenas 5: fundo, CSF, substâncias cinza e branca e a lesão de EM (vide Figura 6.3b). Os cortes utilizados são provenientes de volumes em seqüências do tipo T1 e T2. A espessura dos cortes foi configurada para 1mm, o nível de ruído aditivo variou de 3% a 9% e a intensidade de ruído não uniforme se apresenta nula ou em 20%.

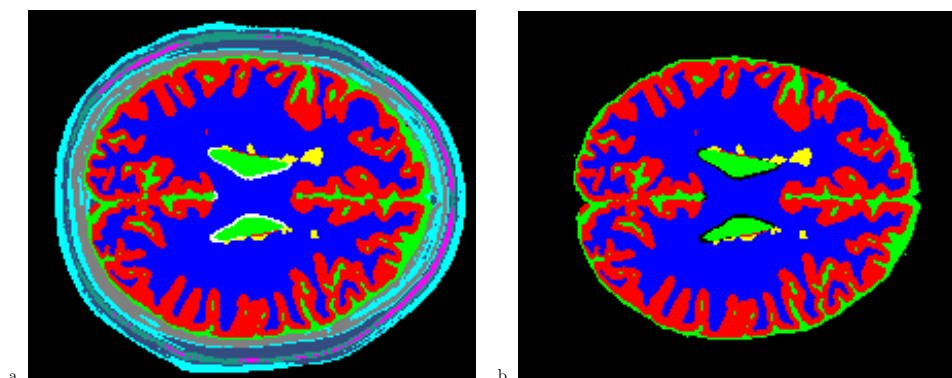


Figura 6.3 - Modelos anatômicos: a) 11 classes e b) 5 classes

6.2.1.1 Espaço de atributos

Os espaços de atributos são obtidos a partir dos conjuntos de treinamento e teste. Nesta aplicação, os espaços bidimensionais são formados pelos valores de intensidade de pixel provenientes de imagens do tipo T1 e T2, mais precisamente, do mesmo corte axial de número 100, oriundos dos volumes T1 e T2. Foram extraídos 100 pontos para cada classe, totalizando um conjunto de amostras com 500 pontos. Os cortes T1 e T2 podem ser observados na Figura 6.4

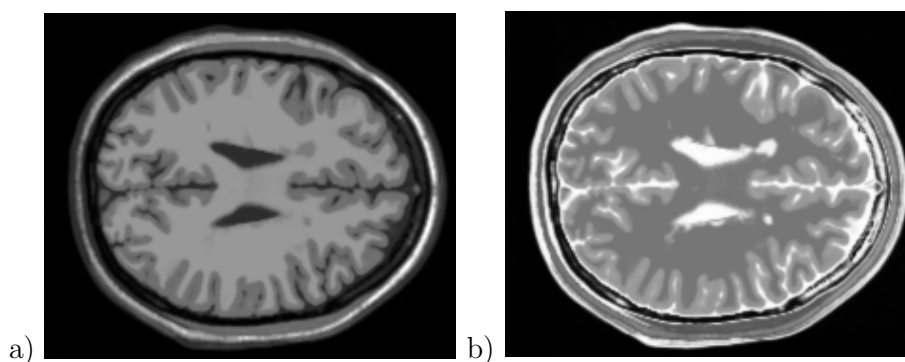


Figura 6.4 - Cortes axiais: a) tipo T1 e b) tipo T2.

A validação foi feita subdividindo o conjunto em 5 partes. Para cada um dos experimentos realizados, o conjunto de treinamento contém 400 pontos e o conjunto de teste 100 pontos, o que corresponde a 80 e 20 pontos para cada classe, respectivamente. Os espaços de atributos apresentados na Figura 6.5 são correspondentes aos conjuntos de amostras para as imagens sem ruído, com 3% e com 9% de ruído aditivo.

6.2.1.2 Classificação imprecisa

Os conjuntos de tabelas apresentados a seguir são referentes aos experimentos realizados com a mesma imagem de RM sem ruído, com 3% e com 9% de ruído. A configuração inicial, ou seja, o número de centros calculados para cada classe, em cada um dos experimentos, para a imagem sem ruído é dada por: $[T_{exp_1} < 1; 20; 14; 9; 6 >$, $T_{exp_2} < 1; 26; 16; 12; 1 >$, $T_{exp_3} < 1; 14; 7; 3; 6 >$, $T_{exp_4} < 1; 14; 46; 9; 7 >$, $T_{exp_5} < 1; 21; 33; 9; 1 >]$. Para a imagem com 3% de ruído tem-se: $[T_{exp_1} < 1; 24; 59; 37; 28 >$, $T_{exp_2} < 1; 28; 39; 41; 2 >$, $T_{exp_3} < 1; 19; 41; 43; 42 >$, $T_{exp_4} < 1; 25; 60; 29; 30 >$, $T_{exp_5} < 1; 17; 22; 29; 29 >]$; e para 9%, $[T_{exp_1} < 1; 67; 78; 71; 57 >$, $T_{exp_2} < 1; 73; 76; 73; 54 >$, $T_{exp_3} < 1; 72; 74; 57; 55 >$, $T_{exp_4} < 1; 69; 73; 72; 54 >$, $T_{exp_5} < 1; 63; 71; 69; 49 >]$.

No primeiro caso, sem ruído, para o modelo AGRUP, a relação de similaridade que

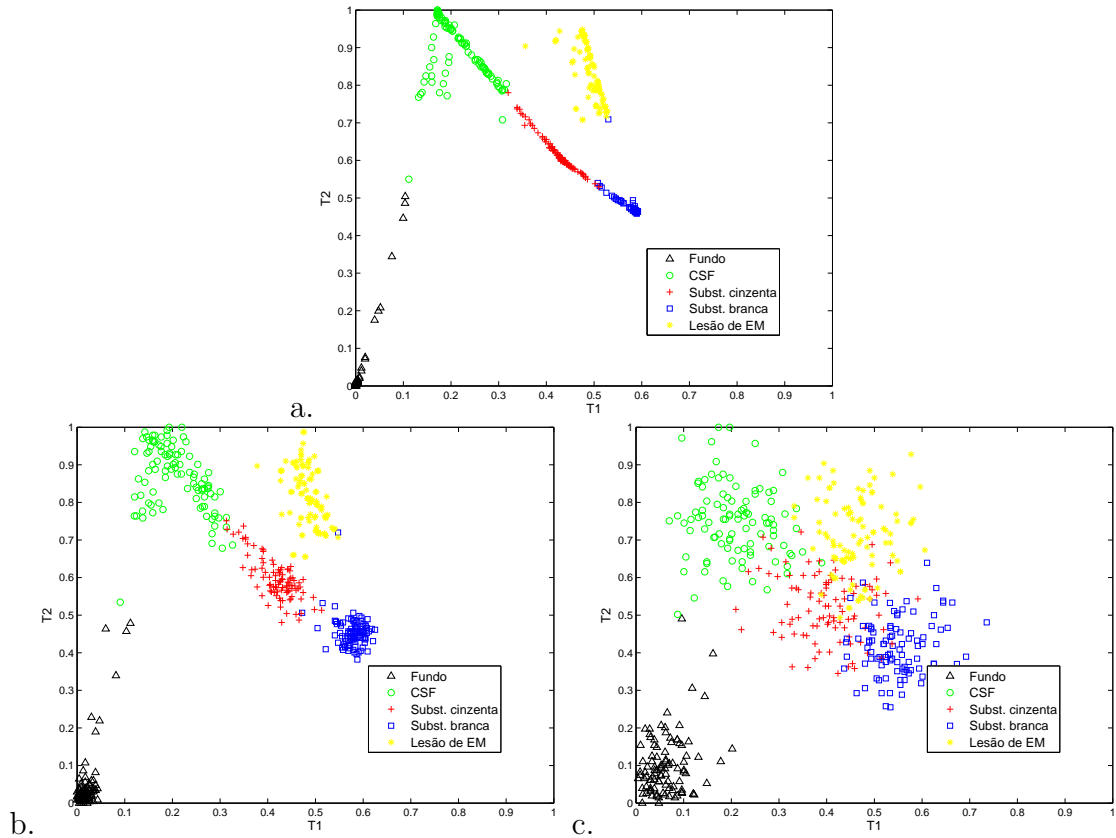


Figura 6.5 - Espaço de atributos formado pelas amostras das imagens T1 e T2: a) sem ruído, b) com 3% de ruído e c) com 9% de ruído.

apresentou melhor resultado foi o trapézio empregando γ_{tr2} e λ_{tr1} , ou seja, a relação $S_{\gamma_{tr2}, \lambda_{tr1}}^L$. Observando as Tabelas 6.2 e 6.3, é possível comparar as medidas de precisão para as famílias parametrizadas de classificadores geradas pelos modelos AGRUP e FKNN com $k = 5$, ou seja, usando 5 vizinhos mais próximos. Para valores de α acima de 0,5; os modelos se apresentam muito similares tanto no teste, quanto no treinamento, e em todos os experimentos.

Tabela 6.2 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética sem ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,2425	0,3850	0,4975	0,5725	0,825	0,9875	0,9925	0,9925	0,9925	0,9975
Te1	0,0000	0,2700	0,4100	0,4700	0,5100	0,6200	0,9800	0,9800	0,9800	0,9800	0,9800
Tr2	0,0000	0,5025	0,8475	0,8850	0,9300	0,9575	0,9825	0,9900	0,9925	0,9925	0,9925
Te2	0,0000	0,5200	0,8100	0,8800	0,9300	0,9500	0,9800	0,9900	0,9900	0,9900	0,9900
Tr3	0,0000	0,4575	0,6950	0,8925	0,9425	0,9775	0,9875	0,9925	0,9925	0,9925	0,9975
Te3	0,0000	0,4500	0,6200	0,9200	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800	0,9900
Tr4	0,0000	0,4475	0,7250	0,8875	0,9350	0,9900	0,9925	1,0000	1,0000	1,0000	1,0000
Te4	0,0000	0,4300	0,7400	0,8600	0,9000	0,9600	0,9600	0,9600	0,9700	0,9800	0,9800
Tr5	0,0000	0,4750	0,7675	0,9550	0,9675	0,9875	0,9875	0,9900	0,9925	0,9925	0,9950
Te5	0,0000	0,4800	0,8100	0,9300	0,9600	0,9700	0,9800	0,9800	0,9800	0,9800	0,9800

Tabela 6.3 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética sem ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,9600	0,9725	0,9800	0,9800	0,9850	0,9850	0,9850	0,9850	0,9875	0,9950
Te1	0,0000	0,9600	0,9700	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800
Tr2	0,0000	0,9650	0,9700	0,9775	0,9775	0,9850	0,9850	0,9875	0,9925	0,9925	0,9975
Te2	0,0000	0,9800	0,9800	0,9800	0,9900	0,9900	0,9900	0,9900	0,9900	0,9900	0,9900
Tr3	0,0000	0,9650	0,9775	0,9800	0,9800	0,9875	0,9875	0,9875	0,9875	0,9875	0,9975
Te3	0,0000	0,9700	0,9700	0,9700	0,9700	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800
Tr4	0,0000	0,9675	0,9775	0,9850	0,9850	0,9925	0,9925	0,9950	0,9950	0,9950	1,0000
Te4	0,0000	0,9300	0,9600	0,9600	0,9700	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800
Tr5	0,0000	0,9625	0,9725	0,9800	0,9800	0,9850	0,9850	0,9875	0,9925	0,9925	0,9975
Te5	0,0000	0,9600	0,9700	0,9700	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800

Tabela 6.4 - Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética sem ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,7575	0,6150	0,5025	0,4275	0,1750	0,0125	0,0075	0,0075	0,0075	0,0000
Te1	0,0000	0,7300	0,5900	0,5200	0,4800	0,3700	0,0100	0,0100	0,0100	0,0000	0,0000
Tr2	0,0000	0,4975	0,1525	0,1150	0,0700	0,0425	0,0175	0,0100	0,0075	0,0075	0,0075
Te2	0,0000	0,4700	0,1800	0,1100	0,0600	0,0400	0,0100	0,0000	0,0000	0,0000	0,0000
Tr3	0,0000	0,5425	0,3050	0,1075	0,0575	0,0225	0,0125	0,0075	0,0075	0,0075	0,0000
Te3	0,0000	0,5500	0,3800	0,0800	0,0200	0,0100	0,0100	0,0100	0,0100	0,0100	0,0000
Tr4	0,0000	0,5525	0,2750	0,1125	0,0650	0,0100	0,0075	0,0000	0,0000	0,0000	0,0000
Te4	0,0000	0,5700	0,2600	0,1400	0,1000	0,0300	0,0300	0,0300	0,0200	0,0000	0,0000
Tr5	0,0000	0,5250	0,2325	0,0450	0,0325	0,0125	0,0125	0,0100	0,0075	0,0075	0,0000
Te5	0,0000	0,5200	0,1900	0,0600	0,0300	0,0200	0,0100	0,0100	0,0100	0,0000	0,0000

Tabela 6.5 - Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética sem ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0400	0,0275	0,0200	0,0200	0,0150	0,0150	0,0150	0,0150	0,0075	0,0000
Te1	0,0000	0,0400	0,0300	0,0200	0,0200	0,0100	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	0,0000	0,0350	0,0300	0,0225	0,0225	0,0150	0,0150	0,0125	0,0050	0,0050	0,0000
Te2	0,0000	0,0100	0,0100	0,0100	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr3	0,0000	0,0350	0,0225	0,0200	0,0200	0,0125	0,0125	0,0125	0,0125	0,0125	0,0000
Te3	0,0000	0,0300	0,0300	0,0300	0,0300	0,0100	0,0100	0,0100	0,0100	0,0100	0,0000
Tr4	0,0000	0,0325	0,0225	0,0150	0,0150	0,0075	0,0075	0,0050	0,005	0,005	0,0000
Te4	0,0000	0,0600	0,0300	0,0300	0,0200	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr5	0,0000	0,0375	0,0275	0,0200	0,0200	0,0150	0,0150	0,0125	0,0050	0,0050	0,0000
Te5	0,0000	0,0300	0,0200	0,0200	0,0100	0,0100	0,0100	0,0100	0,0000	0,0000	0,0000

Já para valores baixos de α , é interessante observar que a imprecisão parcial para o modelo FKNN (Tabela 6.5) é muito baixa quando comparada àquela apresentada pelo modelo AGRUP (Tabela 6.4). Verifica-se que os elementos tidos como imprecisos no modelo AGRUP, são corretamente classificados no modelo FKNN, uma vez que as medidas de erro, apresentadas nas Tabelas 6.6 e 6.7, são praticamente similares nos dois modelos. A diferença mais nítida é que a estratégia AGRUP é tão mais imprecisa quanto menor o valor de α , acompanhando a família parametrizada. Os elementos

são corretamente classificados, porém como precisos no modelo FKNN e imprecisos no modelo AGRUP. A vantagem na imprecisão está na flexibilidade do sistema, um modelo mais rígido, como se mostra o FKNN, pode apresentar erro prematuramente.

Tabela 6.6 - Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética sem ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025
Te1	0,0000	0,0000	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0200	0,0200
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025
Te3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0100	0,0100	0,0100	0,0100	0,0200	0,0200
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0050
Te5	0,0000	0,0000	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0200	0,0200

Tabela 6.7 - Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética sem ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0050	0,0050
Te1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0100	0,0200	0,0200	0,0200	0,0200	0,0200
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025
Te2	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025
Te3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100	0,0200
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	0,0000	0,0100	0,0100	0,0100	0,0100	0,0200	0,0200	0,0200	0,0200	0,0200	0,0200
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025
Te5	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0200	0,0200	0,0200

Para os níveis de ruído 3% e 9%, a característica de modelos mais ou menos rígidos se repete. Para a imagem com 3% de ruído, os resultados para precisão, erro e imprecisão parcial, provenientes dos modelos AGRUP e FKNN, estão apresentados, respectivamente, nas Tabelas 6.8, 6.9, 6.10, 6.11, 6.12 e 6.13.

Comparando dados mais ou menos ruidosos, verifica-se que as características do próprio espaço de atributos (vide Figura 6.5) justificam, em parte, o comportamento dos modelos, quando criadas as famílias parametrizadas. Dados menos ruidosos caracterizam classes bem comportadas, no sentido de que se apresentam mais separadas. Comparando os extremos, como a imagem sem ruído e aquela que tem 9% de ruído, para as Tabelas de imprecisão parcial 6.18 e 6.19 para os dados mais ruidosos, observa-se que a diferença entre os modelos não é tão grande para valores baixos de α como constatado no caso da imagem sem ruído. Naturalmente, a distribuição de probabilidade derivada da avaliação dos 5 vizinhos, no momento da classificação, pelo FKNN, tende

Tabela 6.8 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM sintética com 3% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,3725	0,4800	0,8000	0,9675	0,9825	0,9900	0,9925	0,9975	0,9975	0,9975
Te1	0,0000	0,3800	0,4700	0,7200	0,9300	0,9700	0,9800	0,9800	0,9800	0,9800	0,9800
Tr2	0,0000	0,6825	0,8375	0,8975	0,9550	0,9850	0,9900	0,9925	0,9975	0,9975	0,9975
Te2	0,0000	0,6700	0,7900	0,8800	0,9200	0,9400	0,9400	0,9500	0,9600	0,9700	0,9700
Tr3	0,0000	0,6300	0,8450	0,9075	0,9475	0,9800	0,9825	0,9850	0,9950	0,9975	0,9975
Te3	0,0000	0,6300	0,8400	0,8900	0,9400	0,9600	0,9800	0,9900	0,9900	0,9900	0,9900
Tr4	0,0000	0,6375	0,8375	0,9175	0,9550	0,9875	0,9925	0,9925	0,9975	0,9975	0,9975
Te4	0,0000	0,7100	0,8400	0,9000	0,9400	0,9400	0,9400	0,9400	0,9500	0,9500	0,9500
Tr5	0,0000	0,5450	0,7900	0,9150	0,9675	0,9825	0,9825	0,9925	0,9975	0,9975	0,9975
Te5	0,0000	0,5400	0,7800	0,9100	0,9500	0,9700	0,9700	0,9800	0,9800	0,9800	0,9800

Tabela 6.9 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 3% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,9200	0,9675	0,9725	0,9725	0,9850	0,9850	0,9925	0,9925	0,9925	0,9975
Te1	0,0000	0,9200	0,9500	0,9600	0,9700	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800
Tr2	0,0000	0,9275	0,9625	0,9775	0,9775	0,9825	0,9825	0,9975	0,9975	0,9975	1,0000
Te2	0,0000	0,9100	0,9400	0,9400	0,9600	0,9600	0,9600	0,9800	0,9800	0,9800	0,9800
Tr3	0,0000	0,9225	0,9650	0,9825	0,9825	0,9900	0,9900	0,9950	0,9950	0,9950	0,9975
Te3	0,0000	0,9500	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800	0,9900	0,9900	0,9900
Tr4	0,0000	0,9350	0,9625	0,9800	0,9800	0,9900	0,9900	0,9925	0,9925	0,9925	0,9975
Te4	0,0000	0,9000	0,9200	0,9300	0,9300	0,9500	0,9600	0,9600	0,9600	0,9600	0,9600
Tr5	0,0000	0,9300	0,9700	0,9750	0,9750	0,9925	0,9925	0,9975	0,9975	0,9975	1,0000
Te5	0,0000	0,9600	0,9700	0,9700	0,9700	0,9800	0,9800	0,9800	0,9800	0,9800	0,9800

Tabela 6.10 - Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM sintética com 3% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025	0,0025
Te1	0,0000	0,0000	0,0000	0,0000	0,0200	0,0200	0,0200	0,0200	0,0200	0,0200	0,0200
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025
Te2	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100	0,0200	0,0200	0,0200	0,0200	0,0300
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025
Te3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025
Te4	0,0000	0,0000	0,0000	0,0000	0,0100	0,0200	0,0300	0,0300	0,0300	0,0400	0,0500
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025
Te5	0,0000	0,0000	0,0000	0,0000	0,0100	0,0200	0,0200	0,0200	0,0200	0,0200	0,0200

a apresentar valores de pertinência não nulos para mais classes. O comportamento do modelo AGRUP no que se refere à imprecisão, independente da presença ou ausência de ruído, está também relacionada ao modelo em si, que, através da relação de similaridade, cria um conjunto difuso em torno de cada protótipo, definindo a sobreposição entre as classes. Os parâmetros destas relações determinam a imprecisão.

Medidas de precisão e erro para a imagem com 9% de ruído podem ser observadas nas Tabelas 6.14, 6.15, 6.16 e 6.17. A imprecisão completa é nula para os três casos

Tabela 6.11 - Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 3% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025	0,0025	0,0025	0,0025
Te1	0,0000	0,0000	0,0000	0,0100	0,0100	0,0100	0,0100	0,0200	0,0200	0,0200	0,0200
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0200
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025	0,0025	0,0025	0,0025
Te3	0,0000	0,0000	0,0000	0,0000	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100	0,0100
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025	0,0025	0,0025	0,0025
Te4	0,0000	0,0200	0,0200	0,0200	0,0200	0,0200	0,0300	0,0300	0,0400	0,0400	0,0400
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,0100	0,0100	0,0200	0,0200	0,0200	0,0200	0,0200	0,0200	0,0200	0,0200

Tabela 6.12 - Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM sintética com 3% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,6275	0,5200	0,2000	0,0325	0,0175	0,0100	0,0050	0,0000	0,0000	0,0000
Te1	0,0000	0,6200	0,5300	0,2800	0,0500	0,0100	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	0,0000	0,3175	0,1625	0,1025	0,045	0,0150	0,0100	0,0075	0,0000	0,0000	0,0000
Te2	0,0000	0,3200	0,2000	0,1100	0,0700	0,0500	0,0400	0,0300	0,0200	0,0100	0,0000
Tr3	0,0000	0,3700	0,1550	0,0925	0,0525	0,0200	0,0175	0,015	0,0025	0,0000	0,0000
Te3	0,0000	0,3700	0,1600	0,1100	0,0600	0,0400	0,0100	0,0000	0,0000	0,0000	0,0000
Tr4	0,0000	0,3625	0,1625	0,0825	0,0450	0,0125	0,0075	0,0075	0,0000	0,0000	0,0000
Te4	0,0000	0,2900	0,1600	0,1000	0,0500	0,0400	0,0300	0,0300	0,0200	0,0100	0,0000
Tr5	0,0000	0,4550	0,2100	0,0850	0,0325	0,0175	0,0125	0,0075	0,0000	0,0000	0,0000
Te5	0,0000	0,4600	0,2200	0,0900	0,0400	0,0100	0,0100	0,0000	0,0000	0,0000	0,0000

Tabela 6.13 - Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 3% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0800	0,0325	0,0275	0,0275	0,0125	0,0125	0,0050	0,0050	0,0050	0,0000
Te1	0,0000	0,0800	0,0500	0,0300	0,0200	0,0100	0,0100	0,0000	0,0000	0,0000	0,0000
Tr2	0,0000	0,0725	0,0375	0,0225	0,0225	0,0175	0,0175	0,0025	0,0025	0,0025	0,0000
Te2	0,0000	0,0800	0,0500	0,0500	0,0300	0,0300	0,0300	0,0100	0,0100	0,0100	0,0000
Tr3	0,0000	0,0775	0,0350	0,0175	0,0175	0,0075	0,0075	0,0025	0,0025	0,0025	0,0000
Te3	0,0000	0,0500	0,0200	0,0200	0,0100	0,0100	0,0100	0,0100	0,0000	0,0000	0,0000
Tr4	0,0000	0,0650	0,0375	0,0200	0,0200	0,0075	0,0075	0,0050	0,0050	0,0050	0,0000
Te4	0,0000	0,0800	0,0600	0,0500	0,0500	0,0300	0,0100	0,0100	0,0000	0,0000	0,0000
Tr5	0,0000	0,0700	0,0300	0,0250	0,025	0,0075	0,0075	0,0025	0,0025	0,0025	0,0000
Te5	0,0000	0,0300	0,0200	0,0100	0,0100	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

analisados (0%, 3% e 9% ruído), para $\alpha > 0$.

Na Figura 6.6 são apresentados alguns gráficos de classificação. Os índices que indicam a imprecisão parcial são apresentados graficamente de forma mais detalhada, separados por cardinalidade. Existem elementos que tem até 4 classes associadas. A curva $qual^{*\kappa}$, $1 < \kappa < m$, é resultante da soma dos índices correspondentes aos elementos classificados como corretos, que apresentam cardinalidade maior que 1 e menor que o número de classes (m). Os gráficos apresentados referem-se: ao experimento 2 para a imagem sem ruído, modelo AGRUP na Figura 6.6a e modelo FKNN na Figura 6.6b; ao experimento 1 para a imagem com 3% de ruído, modelo AGRUP na Figura

Tabela 6.14 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética com 9% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,6100	0,6925	0,7600	0,9650	0,9800	0,9875	0,9900	0,9950	0,9950	0,9950
Te1	0,0000	0,3900	0,5100	0,5800	0,6400	0,6700	0,6800	0,7100	0,7100	0,7400	0,8000
Tr2	0,0000	0,6025	0,6950	0,7625	0,9475	0,9625	0,9700	0,9775	0,9800	0,9800	0,9825
Te2	0,0000	0,6500	0,7400	0,7500	0,8000	0,8300	0,8500	0,8500	0,8500	0,8600	0,8600
Tr3	0,0000	0,6050	0,6625	0,7325	0,9225	0,9550	0,9700	0,9775	0,9800	0,9800	0,9875
Te3	0,0000	0,6600	0,6700	0,6900	0,7000	0,7000	0,7300	0,7500	0,7700	0,7900	0,7900
Tr4	0,0000	0,6100	0,7050	0,7800	0,9550	0,9775	0,9850	0,9900	0,9900	0,9900	0,9900
Te4	0,0000	0,5300	0,6600	0,7200	0,7300	0,7500	0,7600	0,7700	0,7700	0,7900	0,7900
Tr5	0,0000	0,5375	0,6325	0,7350	0,9225	0,9725	0,9900	0,9925	0,9950	0,9950	0,9950
Te5	0,0000	0,5400	0,6300	0,7100	0,7600	0,7600	0,7800	0,8100	0,8200	0,8400	0,8700

Tabela 6.15 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 9% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,675	0,8075	0,8875	0,8900	0,9350	0,9350	0,9700	0,9750	0,9775	0,9975
Te1	0,0000	0,6500	0,7100	0,7700	0,7900	0,8100	0,8200	0,8600	0,8700	0,8800	0,8900
Tr2	0,0000	0,6300	0,7975	0,8825	0,8850	0,9075	0,9125	0,9550	0,9600	0,9600	0,9900
Te2	0,0000	0,7100	0,8000	0,8200	0,8400	0,8500	0,8600	0,8900	0,8900	0,8900	0,8900
Tr3	0,0000	0,6625	0,8100	0,8700	0,8750	0,9225	0,9250	0,9600	0,9625	0,9625	0,9900
Te3	0,0000	0,6300	0,7400	0,8100	0,8400	0,8500	0,8500	0,8900	0,8900	0,8900	0,8900
Tr4	0,0000	0,6675	0,8225	0,8800	0,8825	0,9375	0,9425	0,9675	0,9675	0,9675	0,995
Te4	0,0000	0,6400	0,7200	0,8000	0,8000	0,8300	0,8400	0,8400	0,8700	0,8700	0,9100
Tr5	0,0000	0,6500	0,8300	0,8975	0,8975	0,9350	0,9375	0,9600	0,9650	0,9650	0,9975
Te5	0,0000	0,6300	0,7200	0,7800	0,8000	0,8000	0,8200	0,8200	0,8500	0,8500	0,8700

Tabela 6.16 - Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética com 9% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	0,0000	0,0500	0,0700	0,0900	0,0900	0,1200	0,1500	0,1600	0,1800	0,1900	0,2000
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025
Te2	0,0000	0,0100	0,1100	0,1100	0,1200	0,1200	0,1200	0,1200	0,1200	0,1400	0,1400
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0075
Te3	0,0000	0,1100	0,1400	0,1700	0,1900	0,1900	0,1900	0,1900	0,1900	0,2000	0,2100
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	0,0000	0,0600	0,0900	0,1200	0,1500	0,1500	0,1600	0,1800	0,2000	0,2000	0,2100
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,0700	0,0900	0,1000	0,1100	0,1100	0,1100	0,1100	0,1100	0,1100	0,1300

6.6c e modelo FKNN na Figura 6.6d; e ao experimento 5 para a imagem com 9% de ruído, modelo AGRUP na Figura 6.6e e modelo FKNN na Figura 6.6f. Todas as curvas mostradas são para os dados de teste (Te).

Além das medidas já comentadas para as tabelas, estes gráficos mostram ainda a medida de acurácia (acc) do classificador preciso, ou seja, para $\alpha = 1$, e também a utilidade de imprecisão (vide Seção 5.4), que agrega à precisão de um classificador impreciso dentro da família, o ganho da sua imprecisão. Esta medida é apenas uma das formas

Tabela 6.17 - Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 9% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025
Te1	0,0000	0,0300	0,0300	0,0600	0,0600	0,0600	0,0600	0,0800	0,0900	0,1100	0,1100
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0075	0,0075	0,0075	0,0100
Te2	0,0000	0,0100	0,0300	0,0600	0,0600	0,0700	0,0800	0,1000	0,1000	0,1100	0,1100
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0075	0,0075	0,0075	0,0100
Te3	0,0000	0,0100	0,0200	0,0200	0,0700	0,0900	0,0900	0,1000	0,1100	0,1100	0,1100
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,005
Te4	0,0000	0,0500	0,0800	0,0800	0,0800	0,0900	0,0900	0,0900	0,0900	0,0900	0,0900
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0025	0,0025	0,0025	0,0025	0,0025	0,0025
Te5	0,0000	0,0200	0,0400	0,0600	0,0800	0,0800	0,1000	0,1100	0,1200	0,1200	0,1300

Tabela 6.18 - Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - imagem de RM sintética com 9% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,3900	0,3075	0,2400	0,0350	0,0200	0,0125	0,0100	0,0050	0,0050	0,0050
Te1	0,0000	0,5600	0,4200	0,3300	0,2700	0,2100	0,1700	0,1300	0,1100	0,0700	0,0000
Tr2	0,0000	0,3975	0,3050	0,2375	0,0525	0,0375	0,0300	0,0225	0,0200	0,0200	0,0150
Te2	0,0000	0,2500	0,1500	0,1400	0,0800	0,0500	0,0300	0,0300	0,0300	0,0000	0,0000
Tr3	0,0000	0,395	0,3375	0,2675	0,0775	0,0450	0,0300	0,0225	0,0200	0,0200	0,0050
Te3	0,0000	0,2300	0,1900	0,1400	0,1100	0,1100	0,0800	0,0600	0,0400	0,0100	0,0000
Tr4	0,0000	0,3900	0,295	0,2200	0,0450	0,0225	0,0150	0,0100	0,0100	0,0100	0,0100
Te4	0,0000	0,4100	0,2500	0,1600	0,1200	0,1000	0,0800	0,0500	0,0300	0,0100	0,0000
Tr5	0,0000	0,4625	0,3675	0,2650	0,0775	0,0275	0,0100	0,0075	0,0050	0,0050	0,0050
Te5	0,0000	0,3900	0,2800	0,2000	0,1400	0,1300	0,1100	0,0800	0,0700	0,0500	0,0000

Tabela 6.19 - Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem de RM sintética com 9% de ruído.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,3250	0,1925	0,1125	0,1100	0,06500	0,0650	0,0300	0,0250	0,0200	0,0000
Te1	0,0000	0,3200	0,2600	0,1700	0,1500	0,1300	0,1200	0,0600	0,0400	0,0100	0,0000
Tr2	0,0000	0,3700	0,2025	0,1175	0,1150	0,0925	0,0850	0,0375	0,0325	0,0325	0,0000
Te2	0,0000	0,2800	0,1700	0,1200	0,1000	0,0800	0,0600	0,0100	0,0100	0,0000	0,0000
Tr3	0,0000	0,3375	0,1900	0,1300	0,1250	0,0775	0,0725	0,0325	0,0300	0,0300	0,0000
Te3	0,0000	0,3600	0,2400	0,1700	0,0900	0,0600	0,0600	0,0100	0,0000	0,0000	0,0000
Tr4	0,0000	0,3225	0,1775	0,1200	0,1175	0,0625	0,0575	0,0325	0,0325	0,0325	0,0000
Te4	0,0000	0,3100	0,2000	0,1200	0,1200	0,0800	0,0700	0,0700	0,0400	0,0400	0,0000
Tr5	0,0000	0,3500	0,1700	0,1025	0,1025	0,0625	0,0600	0,0375	0,0325	0,0325	0,0000
Te5	0,0000	0,3500	0,2400	0,1600	0,1200	0,1200	0,0800	0,0700	0,0300	0,0300	0,0000

possíveis de se ponderar a imprecisão aceitável, visto que o usuário deve adequá-la às necessidades e interesses de uma dada aplicação.

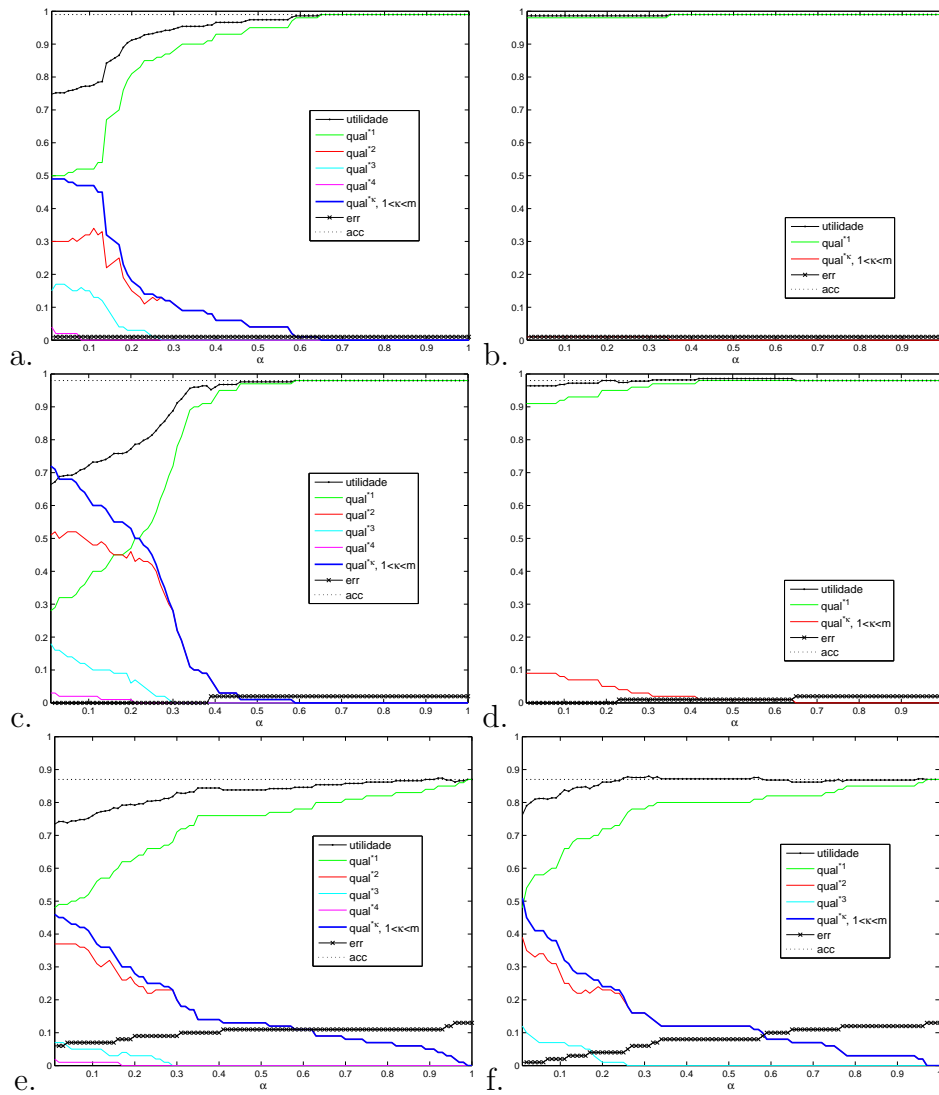


Figura 6.6 - Gráficos de classificação para as imagens de RM sintética: a) 0% ruído, AGRUP Te2; b) 0% ruído, FKNN Te2; c) 3% ruído, AGRUP Te1; d) 3% ruído, FKNN Te1; e) 9% ruído, AGRUP Te5; f) 9% ruído, FKNN Te5

6.2.1.3 Combinação de classificadores

A combinação é feita agregando-se dois classificadores parametrizados através do operador \cap , como descrito na Seção 5.5. Dado o número de experimentos realizados na avaliação de cada modelo, a escolha de quais classificadores devem ser combinados é auxiliada pela análise das tabelas, que sintetizam os índices de qualidade. Isto possibilita eleger um dos treinamentos, uma vez que, para cada experimento uma família de classificadores é definida.

Baseando-se no menor erro de teste para $\alpha = 1$, foram feitas as combinações dos classificadores obtidos nos experimento 4 e 5, para a imagem de RM sintética com 9% de ruído. Assim, tem-se o modelo AGRUP como $Cl_{\alpha_1}^1$ e o modelo FKNN como $Cl_{\alpha_2}^2$. A Figura 6.7a corresponde à combinação $Cl_{\alpha_1, \alpha_2}^{12}$ do experimento 4 e a Figura 6.7b para o experimento 5. Nos dois casos são apresentadas as superfícies de precisão, erro e imprecisão parcial.

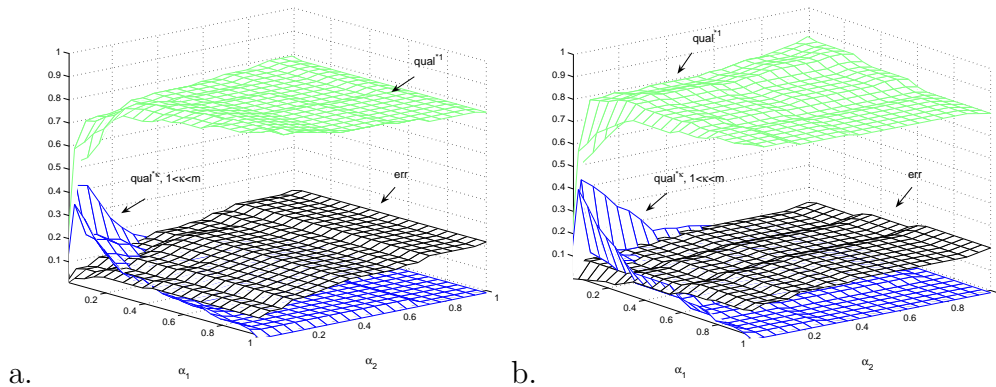


Figura 6.7 - Combinação dos modelos AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ e FKNN com $k = 5$: a) experimento 4 e b) experimento 5

As curvas de classificação que originaram $Cl_{\alpha_1, \alpha_2}^{12}$ para o experimento 4 são mostradas na Figura 6.8a e 6.8b). Nota-se, na Figura 6.8a, para a família de classificadores proveniente do modelo AGRUP, que a curva de utilidade de imprecisão supera a acurácia do classificador em uma faixa de valores de α ($0, 2 < \alpha < 0, 6$). Numa análise global da medida, é interessante a superação, que reflete ganho na classificação, justificado pelo erro crescente, quando se caminha no sentido da classificação precisa.

A escolha de um único classificador a partir da combinação é uma tarefa auxiliada pelos gráficos de classificação. A observação das curvas ajuda o usuário a ponderar os níveis de acurácia e imprecisão aceitáveis numa aplicação. Extraíndo o corte, $\alpha_1 = 0, 10$; tem-

se o classificador $Cl_{0,10;\alpha_2}$, mostrado na Figura 6.8c. Ou seja, é fixado um valor em α_1 e varia-se α_2 , obtendo a família de classificadores $Cl_{0,10;\alpha_2}$. É possível verificar que a combinação faz com que o classificador AGRUP seja mais preciso, e quando comparado com o próprio modelo individualmente, apresenta diminuição significativa no erro.

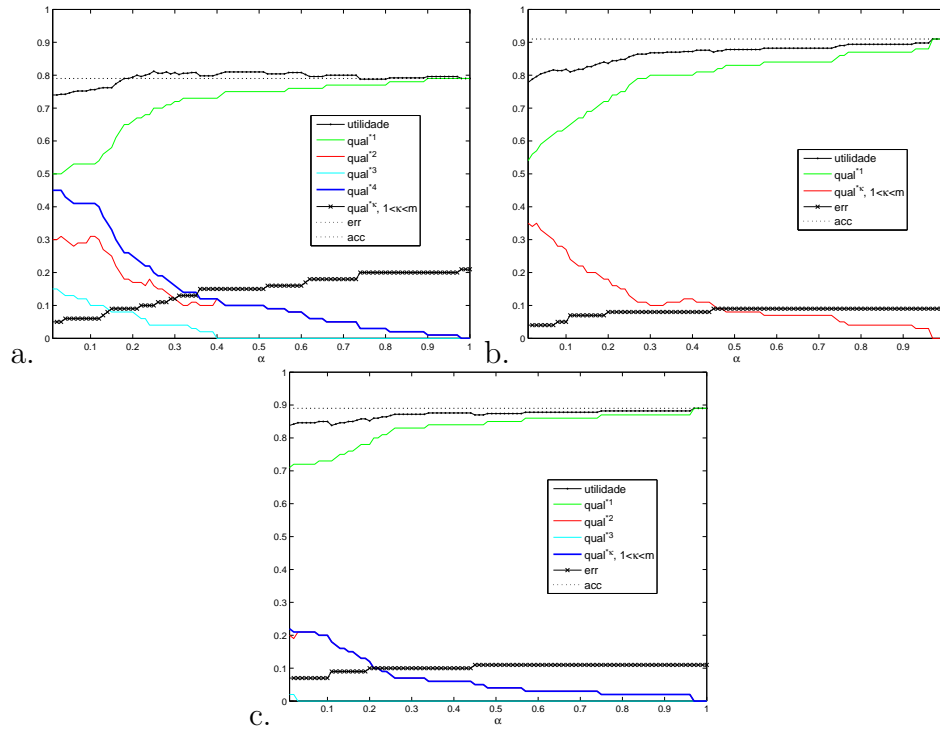


Figura 6.8 - Gráficos de classificação para as imagens de RM sintética: a) 9% ruído, AGRUP Te4; b) 9% ruído, FKNN Te4; c) classificador $Cl_{0,10;\alpha_2}$

6.2.2 Imagens de RM do cérebro com EM - Banco de dados real

O conjunto de dados de RM reais foi obtido a partir de um sistema de campo magnético de 1 Tesla. Dentre as seqüências obtidas foram selecionadas para os experimentos aquelas em que as áreas com desmielinização é mais visível, uma vez que as imagens são de um cérebro com lesões de EM. Diferentemente do caso sintético, em que a verdade existe, e os volumes são construídos a partir de um modelo anatômico, no caso real a definição do conjunto de amostras depende de um radiologista. Ou seja, um especialista é quem pode, dado seu conhecimento, apontar áreas que sejam pertencentes às classes definidas.

Na Figura 6.9 são mostrados os cortes 14 e 15, do tipo T2 e Flair, provenientes de um volume 512x512x20. Destacando uma das classes consideradas, lesão de EM, observam-se as lesões diagnosticadas a partir de um exame clínico: no corte 14, duas áreas de desmielinização do lado direito e no corte 15, duas outras no lado esquerdo. Estas áreas de desmielinização são as manchas esbranquiçadas que aparecem na porção mais interna do cérebro, a substância branca.

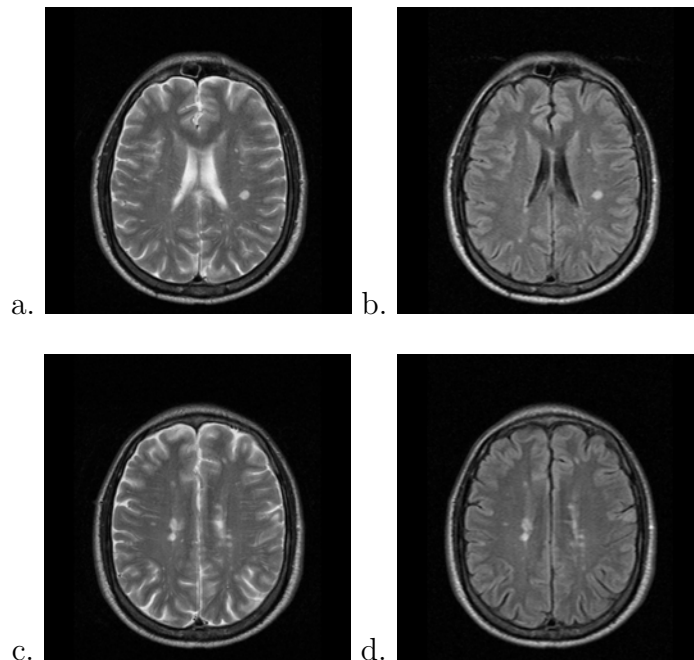


Figura 6.9 - Cortes provenientes de volumes de RM do banco de dados real: a) corte 14 T2, b) corte 14 Flair, c) corte 15 T2 e d) corte 15 Flair.

6.2.2.1 Espaço de atributos

Os experimentos que se seguem utilizam cortes axiais do tipo T2 e Flair provenientes de volumes 512x512x20. As amostras de treinamento são obtidas a partir do corte 14 e as amostras de teste, a partir do corte 15. As classes consideradas são as mesmas 5 classes consideradas na aplicação anterior usando o banco de imagens de RM simuladas.

O espaço de atributos, formado por 375 pontos de 5 classes distintas é apresentado na Figura 6.10.

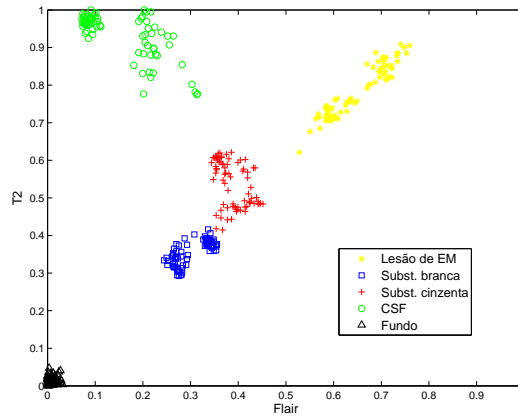


Figura 6.10 - Espaço de atributos formado pelas amostras das imagens T2 e Flair

6.2.2.2 Classificação imprecisa

Na classificação de um dos cortes do volume de RM real, foram avaliadas as mesmas estratégias, AGRUP e FKNN, variando os parâmetros. Para o AGRUP, foram testadas as relações de similaridade linear e não-linear e para o FKNN, diferentes valores de k . As estratégias, nas suas diferentes versões, apresentam resultados satisfatórios e muito similares. A observação das fronteiras entre as classes mostra um espaço de atributos simples (vide Figura 6.10), com confusão mínima entre amostras de classes distintas. Optou-se por mostrar as tabelas referentes à estratégia AGRUP frente a duas relações de similaridade distintas, $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ e $S_{\gamma_{g1}, \lambda_{tr1}}^G$. A configuração inicial para este método define os seguintes números de centros para cada experimento: $[T_{exp1} < 1; 2; 2; 1; 1 >$, $T_{exp2} < 1; 2; 2; 1; 1 >$, $T_{exp3} < 1; 2; 2; 1; 1 >$, $T_{exp4} < 2; 1; 2; 1; 1 >$, $T_{exp5} < 2; 1; 2; 1; 1 >$.

As Tabelas 6.20 e 6.21 apresentam as medidas de precisão. As medidas de erro podem ser observadas nas Tabelas 6.22 e 6.23.

Dados os parâmetros propostos para a relação de similaridade gaussiana empregada,

Tabela 6.20 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM real.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,4833	0,5067	0,6967	0,8833	0,9667	0,9833	0,9867	0,9967	0,9967	1,0000
Te1	0,0000	0,4000	0,4000	0,6267	0,9733	0,9867	0,9867	0,9867	0,9867	1,0000	1,0000
Tr2	0,0000	0,4900	0,5367	0,7467	0,9233	0,9700	0,9867	0,9900	0,9933	0,9933	0,9967
Te2	0,0000	0,4000	0,4400	0,6000	0,8933	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Tr3	0,0000	0,4200	0,5233	0,7100	0,9367	0,9667	0,9867	0,9900	0,9933	0,9967	1,0000
Te3	0,0000	0,4267	0,4933	0,5600	0,8133	0,9200	0,9733	1,0000	1,0000	1,0000	1,0000
Tr4	0,0000	0,4033	0,5500	0,6833	0,8933	0,9300	0,9633	0,9933	1,0000	1,0000	1,0000
Te4	0,0000	0,4000	0,5200	0,5867	0,6667	0,7600	0,8533	0,9067	0,9600	0,9733	0,9733
Tr5	0,0000	0,4200	0,5467	0,6967	0,9267	0,9600	0,9900	1,0000	1,0000	1,0000	1,0000
Te5	0,0000	0,4267	0,5733	0,6400	0,8267	0,9200	0,9467	0,9600	0,9733	0,9867	0,9867

Tabela 6.21 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ - imagem de RM real.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,9767	0,9767	0,9800	0,9833	0,9833	0,9833	0,9833	0,9833	0,9900	0,9933
Te1	0,0000	0,9867	0,9867	0,9867	0,9867	0,9867	1,0000	1,0000	1,0000	1,0000	1,0000
Tr2	0,0000	0,9700	0,9767	0,9800	0,9800	0,9800	0,9800	0,9800	0,9833	0,9900	0,9933
Te2	0,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Tr3	0,0000	0,9767	0,9800	0,9800	0,9833	0,9833	0,9833	0,9867	0,9900	0,9900	0,9933
Te3	0,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Tr4	0,0000	0,9067	0,9133	0,9167	0,9200	0,9267	0,9300	0,9333	0,9400	0,9400	0,9500
Te4	0,0000	0,7733	0,7733	0,7733	0,7733	0,7733	0,7733	0,7733	0,7733	0,8133	0,8533
Tr5	0,0000	0,9000	0,9100	0,9133	0,9167	0,9167	0,9300	0,9500	0,9767	0,9867	1,0000
Te5	0,0000	0,7200	0,7333	0,7333	0,7467	0,8133	0,8533	0,9067	0,9733	1,0000	1,0000

Tabela 6.22 - Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM real.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0033
Te2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0267
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0133

o comportamento dos classificadores imprecisos obtidos a partir do modelo AGRUP é bem diferente, quando se compara àquele submetido à relação de similaridade linear. Através das Tabelas 6.24 e 6.25 de imprecisão parcial, constata-se que o uso da relação gaussiana $S_{\gamma_{g1}, \lambda_{g1}}^G$ faz com que o classificador AGRUP tenha pouca imprecisão. Isso ocorre porque os parâmetros γ e λ definidos geram pequena sobreposição entre as classes. Este comportamento se deve aos parâmetros empregados e não exatamente à utilização de uma relação não linear.

Assim como apresentado para a aplicação anterior, os gráficos de classificação são

Tabela 6.23 - Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ - imagem de RM real.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0033	0,0033	0,0033	0,0067
Te1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0033	0,0033	0,0033	0,0067
Te2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0033	0,0067
Te3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0033	0,0100	0,0167	0,0500
Te4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0267	0,0400	0,0800	0,1467
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Tabela 6.24 - Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem de RM real.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,5167	0,4933	0,3033	0,1167	0,0333	0,0167	0,0133	0,0033	0,0033	0,0000
Te1	0,0000	0,6000	0,6000	0,3733	0,0267	0,0133	0,0133	0,0133	0,0133	0,0000	0,0000
Tr2	0,0000	0,5100	0,4633	0,2533	0,0767	0,0300	0,0133	0,0100	0,0067	0,0067	0,0000
Te2	0,0000	0,6000	0,5600	0,4000	0,1067	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr3	0,0000	0,5800	0,4767	0,2900	0,0633	0,0333	0,0133	0,0100	0,0067	0,0033	0,0000
Te3	0,0000	0,5733	0,5067	0,44	0,1867	0,0800	0,0267	0,0000	0,0000	0,0000	0,0000
Tr4	0,0000	0,5967	0,4500	0,3167	0,1067	0,0700	0,0367	0,0067	0,0000	0,0000	0,0000
Te4	0,0000	0,6000	0,4800	0,4133	0,3333	0,2400	0,1467	0,0933	0,0400	0,0267	0,0000
Tr5	0,0000	0,5800	0,4533	0,3033	0,0733	0,0400	0,0100	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,5733	0,4267	0,3600	0,1733	0,0800	0,0533	0,0400	0,0267	0,0133	0,0000

Tabela 6.25 - Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ - imagem de RM real.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0233	0,0233	0,0200	0,0167	0,0167	0,0133	0,0133	0,0133	0,0067	0,0000
Te1	0,0000	0,0133	0,0133	0,0133	0,0133	0,0133	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	0,0000	0,0300	0,0233	0,0200	0,0200	0,0200	0,0167	0,0167	0,0133	0,0067	0,0000
Te2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr3	0,0000	0,0233	0,0200	0,0200	0,0167	0,0167	0,0167	0,0133	0,0067	0,0067	0,0000
Te3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr4	0,0000	0,0933	0,0867	0,0833	0,0800	0,0733	0,0667	0,0567	0,0433	0,0433	0,0000
Te4	0,0000	0,2267	0,2267	0,2267	0,2267	0,2267	0,2000	0,1867	0,1467	0,0667	0,0000
Tr5	0,0000	0,1000	0,0900	0,0867	0,0833	0,0833	0,0700	0,0500	0,0233	0,0133	0,0000
Te5	0,0000	0,2800	0,2667	0,2667	0,2533	0,1867	0,1467	0,0933	0,0267	0,0000	0,0000

mostrados na Figura 6.11 e referem-se aos experimentos 2 e 5. Neste figura, a primeira coluna refere-se aos experimentos de número 2, para os modelos AGRUP com relações de similaridade linear e não linear e para o modelo FKNN. Já a segunda coluna, apresenta o mesmo, porém para o experimento 5. Para o modelo FKNN, analisado apenas graficamente, tem-se, para o experimento 2, medidas de erro e imprecisão nulas, e conseqüentemente, acurácia máxima. Este comportamento é verificado para toda a família de classificadores, mostrando generalização ótima.

Para o modelo AGRUP, a comparação para as famílias de classificadores já foi discutida nas tabelas. Observa-se o comportamento semelhante do modelo AGRUP com a relação

de similaridade gaussiana e o FKNN. Isto mostra a flexibilidade do método AGRUP quanto aos parâmetros que definem as relações de similaridade, e, conseqüentemente, a sobreposição que determina a imprecisão entre as classes.

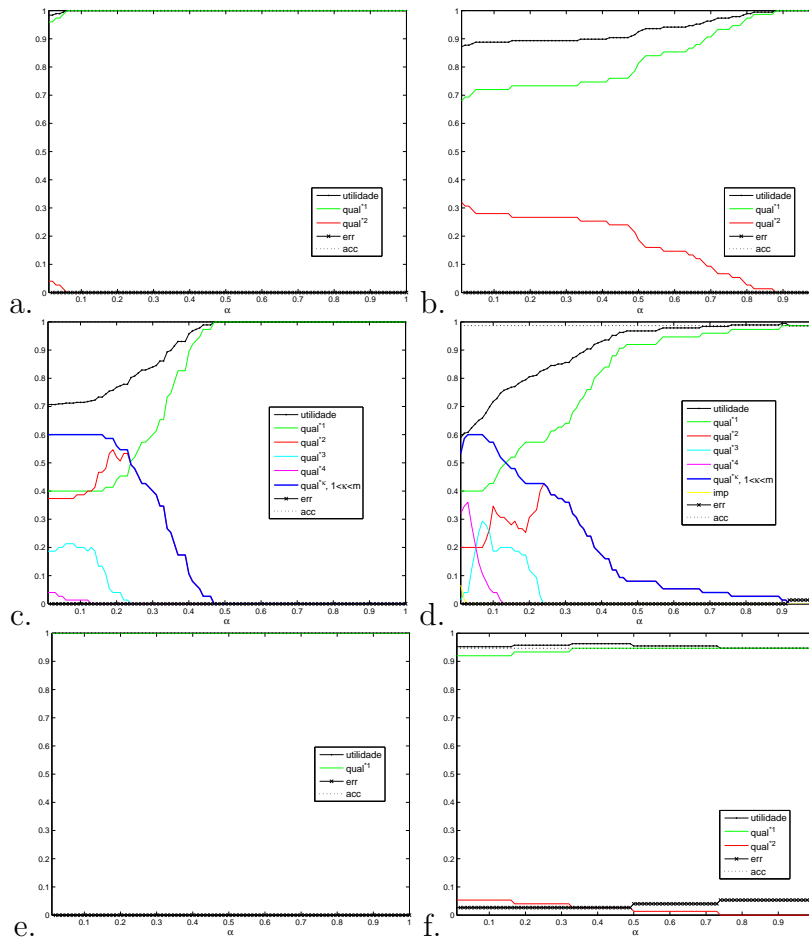


Figura 6.11 - Gráficos de classificação para a imagem de RM real: a) AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ Te2; b) AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ Te5; c) AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ Te2; d) AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ Te5; e) FKNN com $k = 5$ Te2 e f) FKNN com $k = 5$ Te5.

6.2.2.3 Combinação de classificadores

Da mesma forma como foi apresentado pra a aplicação anterior, os classificadores parametrizados são agregados através do operador \cap . Dos 5 experimentos realizados, a maioria apresenta comportamento semelhante quanto a medidas de erro verificadas no caso mais preciso de cada família de classificadores. Desta maneira, foram selecionados os experimentos 2 e 5, cujos gráficos foram estudados também para o modelo FKNN.

As combinações realizadas avaliam: (i) a agregação dos classificadores AGRUP com relação de similaridade linear, Cl^1 , e AGRUP com relação não linear, Cl^2 , treinados

com o mesmo conjunto, no caso, o experimento 5 (vide Figura 6.12a); (ii) a agregação do AGRUP Cl^3 com relação não linear treinado no experimento 2, com o FKNN, Cl^4 , treinado no experimento 5 (vide Figura 6.12b).

No caso (i), a queda na imprecisão é verificada no corte mostrado no gráfico da Figura 6.13. Ou seja, para o classificador $Cl_{0,4;\alpha_1}^{12}$, comparando com o classificador de base (vide Figura 6.11d), a imprecisão é menor e o erro se mantém. No caso (ii), as superfícies obtidas para o erro e imprecisão são praticamente coincidentes em 0. Os classificadores de base apresentam alta precisão, o erro verificado no modelo FKNN é atenuado pela combinação.

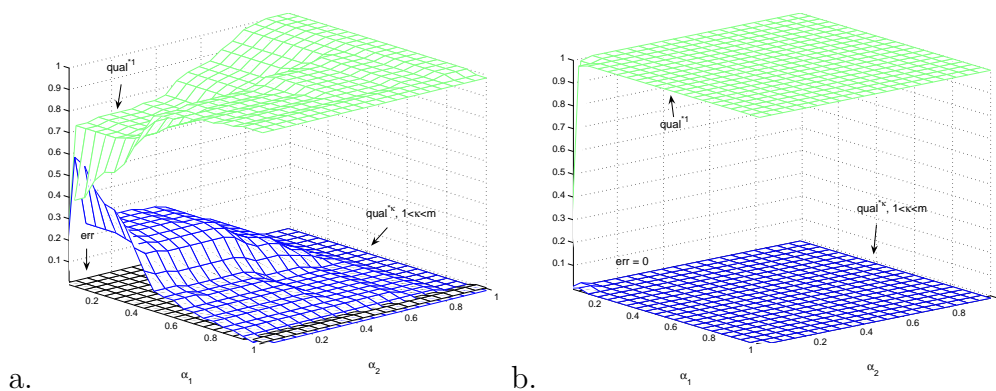


Figura 6.12 - Combinação dos classificadores: a) AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ + AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ experimento 5 e b) AGRUP com $S_{\gamma_{g1}, \lambda_{g1}}^G$ experimento 2 e FKNN experimento 5.

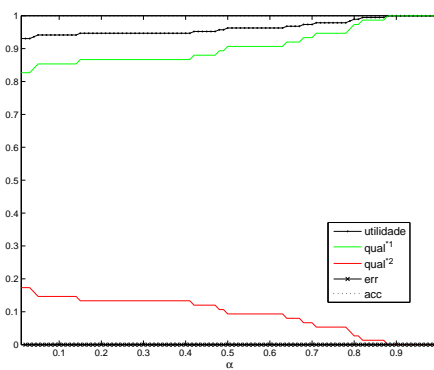


Figura 6.13 - Classificador $Cl_{0,4;\alpha_2}$ proveniente da combinação apresentada em 6.12a).

6.2.2.4 Classificação de imagens de RM restauradas

O método de classificação AGRUP foi também aplicado às imagens de RM que sofreram um processo de restauração. Ou seja, na fase de pré-processamento uma técnica de restauração é empregada, transformando o espaço de atributos.

A restauração de imagens, segundo Gonzalez e Woods (2000), tem como objetivo melhorar uma imagem em algum aspecto, tentando recuperar uma imagem que foi degradada, usando algum conhecimento a priori do fenômeno de degradação. A degradação pode aparecer na forma de ruído do sensor, desfocalização devido ao foco da câmera ou ao movimento relativo entre o objeto e a câmera, ruído atmosférico, entre outros. Desta forma, as técnicas de restauração são orientadas para a modelagem da degradação e aplicação do processo inverso no sentido de recuperar a imagem original.

Considerando que o processo de restauração transforma o espaço de atributos, foi estudado o desempenho do classificador frente a imagens degradadas e restauradas. Dentre as diversas técnicas de restauração definidas na literatura, umas das mais empregadas é o método de Wiener (GONZALEZ; WOODS, 2000). Castro *et al.* (2007) propõem um restaurador multiescala baseado em redes neurais artificiais (RNA), e uma aplicação em que o restaurador neural é empregado na fase de pré-processamento na classificação de imagens de RM.

As imagens de RM apresentadas são aquelas descritas na Seção 6.2.2. O conjunto de amostras é um subconjunto daquele já definido, e a classificação foi feita utilizando um conjunto de treinamento com 200 pontos e um conjunto de teste com 125 pontos.

Os cortes 14 e 15 classificados são apresentados na Figura 6.14 e a Tabela 6.26 mostra os índices *kappa* resultantes da avaliação das amostras de teste na classificação precisa das imagens. São avaliadas a imagem original e as imagens restauradas pela abordagem de Wiener e neural com diferentes parâmetros (RNA 80% e RNA 90%). Os índices apresentados demonstram bom desempenho do classificador, e para as imagens restauradas por redes neurais, o valor de *kappa* alcançado é maior. A proposta de se trabalhar com métodos de restauração, empregados no pré-processamento das imagens, é interessante do ponto de vista da transformação do espaço de atributos. Em geral, o efeito da restauração é comprimir as classes no espaço de atributos, o que não garante melhoria na classificação, mas pode ser favorável quando aplicada a abordagem baseada em agrupamento (classificador AGRUP). A etapa inicial do método cria centros ou protótipos para as classes, quanto menor o número de centros necessários, menos custosa é a classificação. Assim, se um restaurador obtém um espaço de atributos para

o qual o número de protótipos criados para cada classe é menor, isto reflete num ganho em termos de tempo computacional. Porém, trata-se de espaços distintos, o original e o restaurado, e a classificação em si pode não apresentar melhoria significativa.

Tabela 6.26 - Índices *kappa* para a imagem de RM real.

Degradada	Wiener	RNA 80%	RNA 90%
0,951389	0,950000	0,979167	0,970139

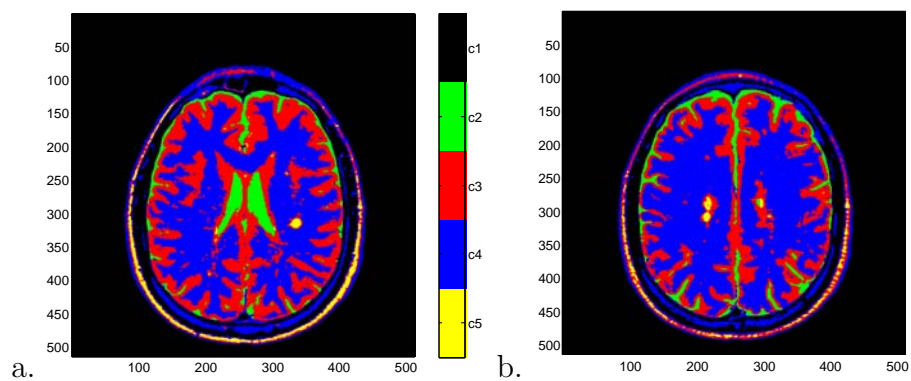


Figura 6.14 - Imagens de treinamento e teste classificadas: a) corte 14 e b) corte 15.

6.2.3 Imagem LANDSAT

As imagens utilizadas formam um conjunto das bandas 2, 3 e 4 do LANDSAT 5, da área de Brasília, Distrito Federal, e pode ser observada na Figura 6.15.

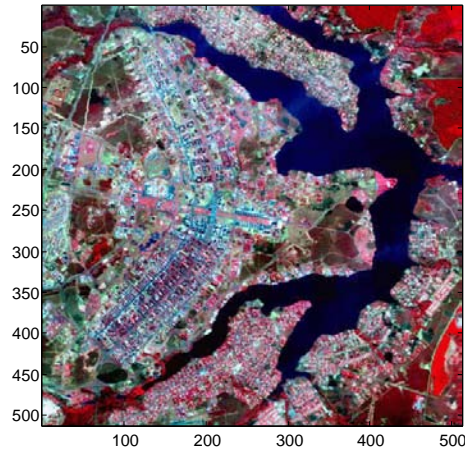


Figura 6.15 - Imagem LANDSAT.

6.2.3.1 Espaço de atributos

O conjunto de amostras empregado é definido em Silva (2004), que apresenta resultados usando classificação por redes neurais. São consideradas 4 classes distintas: água, vegetação, área residencial e solo exposto.

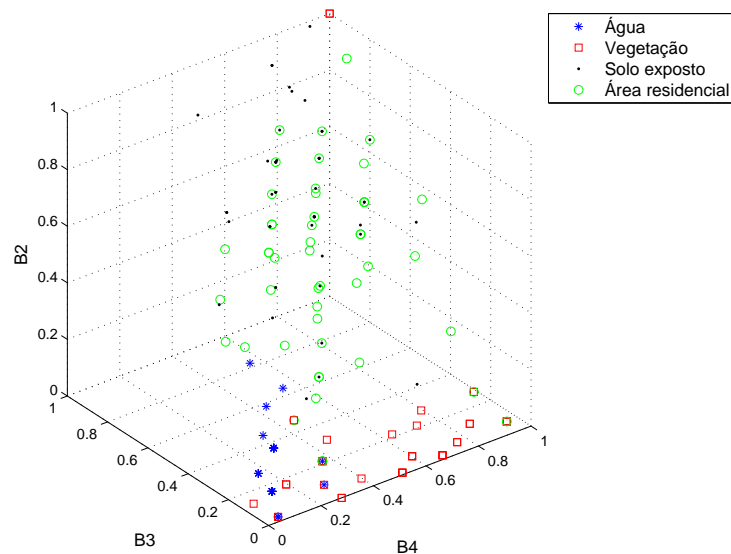


Figura 6.16 - Espaço de atributos normalizado formado pelas amostras da imagem LANDSAT nas bandas 2, 3 e 4.

O espaço de atributos é formado pelas respostas espectrais dos pixels nas bandas 2, 3 e 4. O espaço tridimensional normalizado, com 50 amostras para cada classe, ou seja, 200 pixels da imagem nas três faixas espectrais, é representado na Figura 6.16.

6.2.3.2 Classificação imprecisa

Os índices de imprecisão derivados das famílias parametrizadas de classificadores são sintetizados nas tabelas seguintes. As distribuições de possibilidade utilizadas são provenientes dos métodos AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ e FKNN com $k = 5$. A configuração inicial para esta aplicação não utiliza o meta-agrupamento. Neste caso, as próprias amostras são empregadas como protótipos, numa versão mais geral do modelo AGRUP (vide Seção 4.2.3).

As duas abordagens se mostram bem similares, sendo verificada, em alguns experimentos, a ocorrência da imprecisão completa para valores de $\alpha < 0,4$ tanto no treinamento, quanto no teste (Tabelas 6.27 e 6.28). As medidas de precisão, erro e imprecisão parcial podem ser verificadas nas Tabelas 6.29, 6.30, 6.31, 6.32, 6.33 e 6.34.

Tabela 6.27 - Imprecisão completa em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - LANDSAT.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	1,0000	0,0187	0,0187	0,0187	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	1,0000	0,025	0,0125	0,0125	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	1,0000	0,0500	0,0250	0,0250	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr3	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr4	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr5	1,0000	0,0250	0,0250	0,0125	0,0063	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	1,0000	0,0500	0,0500	0,0250	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Tabela 6.28 - Imprecisão completa em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem LANDSAT.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	1,0000	0,0063	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	1,0000	0,0250	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr3	1,0000	0,0125	0,0125	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr4	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr5	1,0000	0,0063	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	1,0000	0,0250	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Tabela 6.29 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem LANDSAT.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,4188	0,4437	0,4688	0,6438	0,7000	0,8187	0,8375	0,8625	0,8625	0,8625
Te1	0,0000	0,4500	0,4500	0,4500	0,5000	0,5500	0,6000	0,6000	0,6000	0,6000	0,6250
Tr2	0,0000	0,3937	0,5375	0,5563	0,5813	0,6312	0,8000	0,8125	0,8125	0,8313	0,8562
Te2	0,0000	0,1250	0,2750	0,2750	0,3250	0,4500	0,5000	0,5500	0,5500	0,5500	0,6000
Tr3	0,0000	0,3000	0,3438	0,3750	0,5000	0,5875	0,7500	0,7688	0,7875	0,8000	0,8125
Te3	0,0000	0,3000	0,3000	0,3000	0,5250	0,5500	0,5750	0,6000	0,6000	0,6000	0,6250
Tr4	0,0000	0,3375	0,3750	0,3812	0,5313	0,6500	0,7813	0,7813	0,8125	0,8187	0,8375
Te4	0,0000	0,3500	0,3750	0,3750	0,5250	0,6000	0,7000	0,7000	0,7250	0,7250	0,7750
Tr5	0,0000	0,3563	0,3875	0,4188	0,6188	0,6687	0,8125	0,8125	0,8313	0,8375	0,8875
Te5	0,0000	0,4000	0,4500	0,4750	0,5500	0,5750	0,6250	0,6250	0,6250	0,6250	0,6750

Tabela 6.30 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem LANDSAT.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,4063	0,4875	0,5687	0,6062	0,7188	0,7562	0,8187	0,8313	0,875	0,8875
Te1	0,0000	0,3500	0,4250	0,5000	0,5250	0,6000	0,6500	0,6750	0,6750	0,7000	0,7000
Tr2	0,0000	0,4063	0,5000	0,6375	0,6625	0,7438	0,7562	0,8313	0,8875	0,8875	0,9000
Te2	0,0000	0,2750	0,3250	0,3500	0,3750	0,3750	0,4500	0,4750	0,5250	0,5250	0,5500
Tr3	0,0000	0,3125	0,4562	0,5437	0,5813	0,7000	0,7312	0,8000	0,8438	0,8625	0,8812
Te3	0,0000	0,3750	0,4250	0,4500	0,4750	0,5250	0,5750	0,5750	0,6250	0,6500	0,7000
Tr4	0,0000	0,3063	0,4188	0,5313	0,5563	0,6875	0,7188	0,8187	0,8500	0,8500	0,8688
Te4	0,0000	0,3250	0,4000	0,4750	0,4750	0,6250	0,6750	0,6750	0,7000	0,7000	0,7500
Tr5	0,0000	0,3625	0,5000	0,6250	0,6625	0,7625	0,7875	0,8688	0,8812	0,8812	0,8938
Te5	0,0000	0,4750	0,5250	0,5250	0,5250	0,5500	0,5500	0,6750	0,6750	0,6750	0,6750

Tabela 6.31 - Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem LANDSAT.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0563	0,0625	0,0688	0,0938	0,0938	0,1000	0,1125	0,1375	0,1375	0,1375
Te1	0,0000	0,1750	0,1750	0,2000	0,2750	0,3000	0,3250	0,3250	0,3750	0,3750	0,3750
Tr2	0,0000	0,0313	0,0313	0,0313	0,0625	0,075	0,0875	0,1000	0,1000	0,1313	0,1437
Te2	0,0000	0,1000	0,1500	0,1500	0,1750	0,3250	0,3250	0,3250	0,3250	0,350	0,4000
Tr3	0,0000	0,0437	0,0500	0,0500	0,0813	0,0875	0,1000	0,1125	0,1375	0,1625	0,1688
Te3	0,0000	0,1250	0,1250	0,1500	0,1750	0,2250	0,3000	0,3250	0,3250	0,3500	0,3500
Tr4	0,0000	0,0813	0,0813	0,0813	0,0938	0,0938	0,0938	0,0938	0,1313	0,1375	0,1625
Te4	0,0000	0,1000	0,1000	0,1000	0,1000	0,1750	0,2250	0,2250	0,2250	0,2250	0,2250
Tr5	0,0000	0,0437	0,0437	0,0500	0,0625	0,0625	0,0625	0,0625	0,0875	0,0938	0,1125
Te5	0,0000	0,0750	0,0750	0,1000	0,1750	0,2000	0,2250	0,2250	0,2250	0,2500	0,3250

Na Figura 6.17 são apresentados os gráficos contendo as curvas de teste para o experimento 4 usando AGRUP e FKNN. A escolha deste experimento para estudo foi feita com base na acurácia precisa medida. Para as duas abordagens, dentre os 5 experimentos, o melhor caso é o experimento 4. Novamente, é observado no gráfico da Figura 6.17b, um caso em que a medida de utilidade de imprecisão supera a acurácia para alguns dos classificadores dentro da família parametrizada, o que mostra o ganho da imprecisão dentro do sistema.

Tabela 6.32 - Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem LANDSAT.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0125	0,0250	0,0437	0,0625	0,0688	0,1063	0,1125
Te1	0,0000	0,1000	0,1750	0,1750	0,2000	0,2250	0,2250	0,2500	0,2750	0,3000	0,3000
Tr2	0,0000	0,0063	0,0063	0,0187	0,0250	0,0250	0,0313	0,0625	0,0875	0,0875	0,1000
Te2	0,0000	0,1000	0,1500	0,1500	0,2000	0,2250	0,2750	0,3000	0,3750	0,3750	0,4500
Tr3	0,0000	0,0000	0,0125	0,0125	0,0375	0,0500	0,0625	0,0750	0,0938	0,1063	0,1187
Te3	0,0000	0,0000	0,0500	0,0750	0,0750	0,1000	0,1250	0,2250	0,2750	0,2750	0,3000
Tr4	0,0000	0,0000	0,0063	0,0187	0,0375	0,0563	0,075	0,1187	0,1250	0,1250	0,1313
Te4	0,0000	0,0000	0,0000	0,0500	0,0750	0,0750	0,0750	0,1750	0,1750	0,2250	0,2500
Tr5	0,0000	0,0063	0,0063	0,0187	0,0500	0,0563	0,0688	0,0938	0,1000	0,1000	0,1063
Te5	0,0000	0,1000	0,1500	0,1750	0,1750	0,2500	0,2500	0,2500	0,3250	0,3250	0,3250

Tabela 6.33 - Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - LANDSAT.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,5062	0,4750	0,4437	0,2625	0,2062	0,0813	0,0500	0,0000	0,0000	0,0000
Te1	0,0000	0,3750	0,3750	0,3500	0,2250	0,1500	0,0750	0,0750	0,0250	0,0250	0,0000
Tr2	0,0000	0,5500	0,4188	0,4000	0,3563	0,2938	0,1125	0,0875	0,0875	0,0375	0,0000
Te2	0,0000	0,7250	0,5500	0,5500	0,5000	0,2250	0,1750	0,1250	0,1250	0,1000	0,0000
Tr3	0,0000	0,6563	0,6062	0,575	0,4188	0,3250	0,1500	0,1187	0,075	0,0375	0,0187
Te3	0,0000	0,5750	0,5750	0,5500	0,3000	0,2250	0,1250	0,0750	0,0750	0,0500	0,0250
Tr4	0,0000	0,5813	0,5437	0,5375	0,3750	0,2562	0,1250	0,1250	0,0563	0,0437	0,0000
Te4	0,0000	0,5500	0,5250	0,5250	0,3750	0,2250	0,0750	0,0750	0,0500	0,0500	0,0000
Tr5	0,0000	0,5750	0,5437	0,5188	0,3125	0,2687	0,1250	0,1250	0,0813	0,0688	0,0000
Te5	0,0000	0,4750	0,4250	0,4000	0,2750	0,2250	0,1500	0,1500	0,1500	0,1250	0,0000

Tabela 6.34 - Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem LANDSAT.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,5938	0,5125	0,4312	0,3812	0,2562	0,2000	0,1187	0,1000	0,0187	0,0000
Te1	0,0000	0,5500	0,4000	0,3250	0,2750	0,1750	0,1250	0,0750	0,0500	0,0000	0,0000
Tr2	0,0000	0,5812	0,4938	0,3438	0,3125	0,2312	0,2125	0,1063	0,0250	0,0250	0,0000
Te2	0,0000	0,6000	0,5250	0,5000	0,4250	0,4000	0,2750	0,2250	0,1000	0,1000	0,0000
Tr3	0,0000	0,6750	0,5188	0,4437	0,3813	0,2500	0,2062	0,1250	0,0625	0,0313	0,0000
Te3	0,0000	0,6250	0,525	0,4750	0,4500	0,3750	0,3000	0,2000	0,1000	0,0750	0,0000
Tr4	0,0000	0,6937	0,575	0,4500	0,4063	0,2562	0,2062	0,0625	0,0250	0,0250	0,0000
Te4	0,0000	0,6750	0,6000	0,4750	0,4500	0,3000	0,2500	0,1500	0,1250	0,0750	0,0000
Tr5	0,0000	0,6250	0,4938	0,3563	0,2875	0,1812	0,1437	0,0375	0,0187	0,0187	0,0000
Te5	0,0000	0,4000	0,3250	0,3000	0,3000	0,2000	0,2000	0,0750	0,0000	0,0000	0,0000

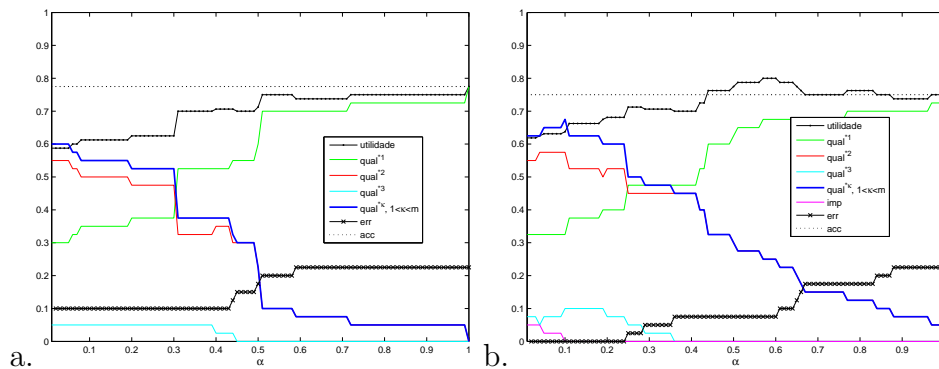


Figura 6.17 - Gráficos de classificação - imagem satélite LANDSAT - a) AGRUP Te4 e b) FKNN Te4.

6.2.3.3 Classificação precisa

O refinamento pela vizinhança, descrito na Seção 5.6.2, é empregado para obtenção de uma classificação em termos de classes puras, dado um classificador impreciso. O refinamento é aplicado a uma família parametrizada de classificadores imprecisos, obtida através das distribuições de possibilidade derivadas do classificador AGRUP, treinado no experimento 4. A escolha do experimento foi feita com base nos valores de precisão observados, em $\alpha = 1$, tanto para o AGRUP quanto para o FKNN, sendo este o experimento com maior precisão de teste (vide Tabelas 6.29 e 6.30).

A aplicação da vizinhança gera uma nova classificação (imagem classificada), sendo calculado o índice *kappa* e a taxa de acerto, sobre as amostras de teste, para cada uma das novas imagens obtidas. Os valores de *kappa* e as taxas de acerto (*acc*) são apresentados na Tabela 6.35. A Figura 6.18 mostra a imagem classificada após o refinamento da classificação obtida pelo classificador impreciso $Cl_{\alpha=0,5}$.

Tabela 6.35 - Índices *kappa* e *acc* para a imagem LANDSAT após refinamento.

α	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
<i>acc</i>	0,825	0,825	0,825	0,750	0,800	0,775	0,775	0,775	0,775
<i>kappa</i>	0,766	0,766	0,766	0,666	0,733	0,700	0,700	0,700	0,700

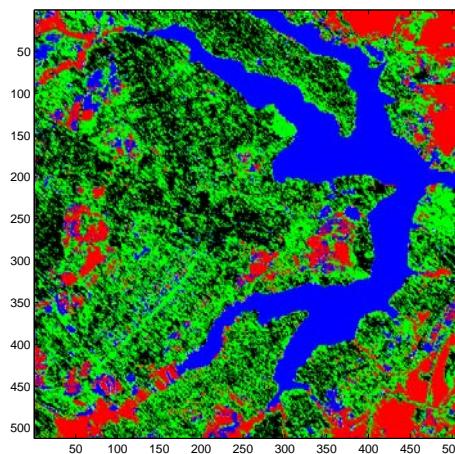


Figura 6.18 - Imagem LANDSAT classificada.

O trabalho de Silva (2004) apresenta três experimentos, dos quais, dadas as matrizes de confusão, pode-se extrair as medidas de desempenho descritas na Tabela 6.36. Estas classificações, sobre a mesma imagem, empregando o mesmo conjunto de amostras,

foram obtidas por classificadores neurais. Uma análise dos resultados, mostra um desempenho satisfatório do classificador. Avaliando os classificadores imprecisos após o refinamento, constata-se a vantagem em se combinar classificação no espaço de atributos com refinamento no espaço do problema, como forma de agregar informação presente da vizinhança, em favor da classificação correta do pixel. E ainda a comparação dos resultados aqui obtidos com aqueles apresentados por Silva (2004), demonstra que as abordagens são comparáveis.

Tabela 6.36 - Índices *kappa* e *acc* para a imagem LANDSAT classificada por redes neurais em Silva (2004).

Experimento	1	2	3
<i>acc</i>	0,7400	0,7250	0,8300
<i>kappa</i>	0,6533	0,6284	0,7800

6.2.4 Imagem CBERS

A imagem CBERS, utilizada em Drummond e Sandri (2004), é multiespectral, com 3 bandas, dimensão 839×650 (vide Figura 6.19), escala $\Omega = \{0, \dots, 255\}$ e representa a região nordeste do Rio Grande do Sul, no Brasil.

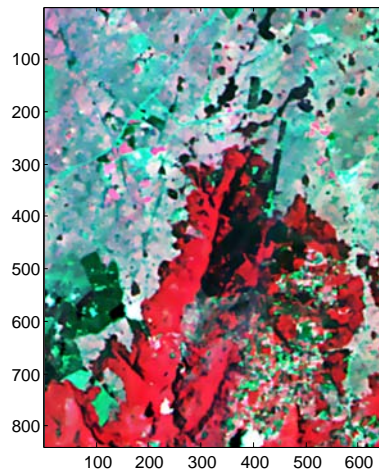


Figura 6.19 - Imagem CBERS

6.2.4.1 Espaço de atributos

O espaço de atributos, apresentado na Figura 6.20, é formado pelas amostras de 8 classes, que correspondem a áreas com: Araucária, Mata Atlântica (Nativa), reflorestamento (Pinus), campo, solo exposto, estrada, nuvens e sombra; em três bandas. Cada classe do conjunto contém 150 pontos de amostra.

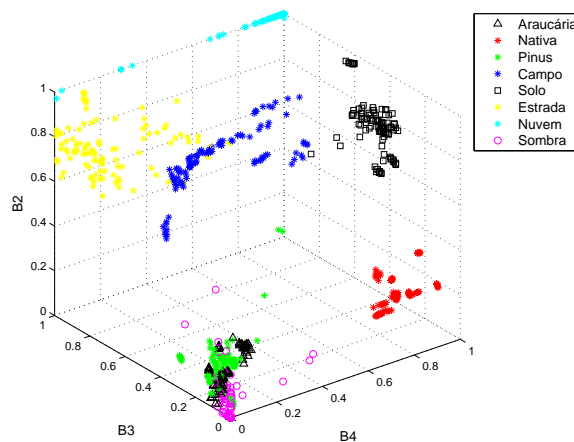


Figura 6.20 - Espaço de atributos normalizado formado pelas amostras da imagem CBERS nas bandas 2, 3 e 4.

6.2.4.2 Classificação imprecisa

Os modelos AGRUP e FKNN foram utilizados na obtenção das distribuições de possibilidade para a imagem CBERS. Em cada um dos 5 experimentos realizados, o classificador é treinado com 120 pontos e testado com os 30 restantes, para cada classe, o que totaliza 960 pontos de treinamento e 240 de teste. Como configuração inicial, o número de centros gerados em cada experimento pelo meta-agrupamento é dado por: $[T_{exp_1} < 33; 4; 81; 37; 15; 74; ; 8; 71 >, T_{exp_2} < 24; 3; 100; 13; 75; 91; 9; 63 >, T_{exp_3} < 40; 6; 77; 21; 21; 74; 4; 63 >, T_{exp_4} < 55; 1; 75; 57; 14; 97; 8; 49 >, T_{exp_5} < 34; 5; 99; 69; 2; 97; 11; 58 >]$.

As tabelas que se seguem mostram precisão, erro e imprecisão parcial, variando o parâmetro α . Observando os índices de acerto preciso para a família parametrizada obtida a partir dos dois métodos, nas Tabelas 6.37 e 6.38, verifica-se que, para a maioria dos experimentos de teste, 4 dos 5 realizados, o índice de acerto preciso é melhor para o modelo AGRUP. O mesmo ocorre para os conjuntos de treinamento.

A comparação do erro, nas Tabelas 6.39 e 6.40, tanto de treinamento quanto de teste, mostra que o FKNN, para valores mais baixos de α , tem erro acima daquele observado para o modelo AGRUP. Esta, por sua vez, tem maior imprecisão parcial (Tabela 6.41). A Tabela 6.42 mostra a imprecisão parcial para o modelo FKNN. A imprecisão completa para os dois classificadores é nula para $\alpha > 0$. Como já foi discutido, na análise feita em outras aplicações, a imprecisão parcial é característica do modelo, que cria, a partir das relações de similaridade, hipersferas que se sobrepõem.

Tabela 6.37 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem CBERS.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,4010	0,5354	0,7010	0,8792	0,9292	0,9344	0,9833	0,9854	0,9854	0,9885
Te1	0,0000	0,3667	0,4958	0,5875	0,6625	0,7083	0,7542	0,8417	0,8625	0,8708	0,8833
Tr2	0,0000	0,3500	0,5073	0,7125	0,8948	0,9323	0,9385	0,9396	0,9885	0,9885	0,9896
Te2	0,0000	0,3333	0,45	0,5917	0,6292	0,6583	0,6958	0,7083	0,8	0,825	0,8458
Tr3	0,0000	0,4583	0,5594	0,7010	0,8760	0,9208	0,9271	0,9281	0,9313	0,9323	0,9885
Te3	0,0000	0,4375	0,5083	0,6000	0,7042	0,7417	0,7625	0,8000	0,8083	0,8167	0,8667
Tr4	0,0000	0,3958	0,5698	0,7490	0,8896	0,9187	0,9219	0,9875	0,9885	0,9885	0,9896
Te4	0,0000	0,3792	0,5208	0,6208	0,7375	0,8083	0,8208	0,8500	0,8583	0,8667	0,8792
Tr5	0,0000	0,4437	0,5833	0,7406	0,9437	0,9885	0,9906	0,9927	0,9948	0,9990	1,0000
Te5	0,0000	0,3833	0,4542	0,5292	0,6167	0,6583	0,7083	0,7583	0,7750	0,7917	0,8000

Os gráficos mostrados na Figura 6.21, referentes ao experimento 1, são apresentados de forma mais suscinta, contendo apenas as curvas de erro, precisão, imprecisão parcial, utilidade de imprecisão e acurácia. O experimento 1 é aquele que tem menor erro em

Tabela 6.38 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem CBERS.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,7885	0,8438	0,8792	0,8792	0,9052	0,9052	0,9281	0,9313	0,9313	0,9521
Te1	0,0000	0,7333	0,7625	0,7792	0,7875	0,8000	0,8000	0,8125	0,8167	0,8167	0,8292
Tr2	0,0000	0,8115	0,8583	0,8885	0,8885	0,9125	0,9125	0,9302	0,9333	0,9333	0,9510
Te2	0,0000	0,6917	0,7208	0,7333	0,7542	0,7667	0,7833	0,7875	0,8167	0,8292	0,8292
Tr3	0,0000	0,7990	0,8531	0,8792	0,8802	0,9052	0,9063	0,9187	0,9198	0,9198	0,9437
Te3	0,0000	0,7000	0,7750	0,7833	0,8125	0,8208	0,8333	0,8333	0,8375	0,8417	0,8458
Tr4	0,0000	0,7937	0,8552	0,8771	0,8771	0,8958	0,8958	0,9135	0,9187	0,9187	0,9354
Te4	0,0000	0,7083	0,7792	0,8208	0,8333	0,8333	0,8417	0,8583	0,8583	0,8625	0,8708
Tr5	0,0000	0,8260	0,8938	0,9302	0,9302	0,9583	0,9583	0,9792	0,9792	0,9792	1,0000
Te5	0,0000	0,7333	0,7750	0,7792	0,7833	0,7875	0,8042	0,8083	0,8083	0,8083	0,8167

Tabela 6.39 - Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - imagem CBERS.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0094	0,0094	0,0094	0,0115
Te1	0,0000	0,0750	0,0792	0,0792	0,0917	0,0958	0,1000	0,1042	0,1083	0,1125	0,1167
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0094	0,0094	0,0104
Te2	0,0000	0,0292	0,0417	0,0583	0,0708	0,0917	0,1125	0,1208	0,1292	0,1417	0,1542
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0010	0,0010	0,0115
Te3	0,0000	0,0583	0,0583	0,0583	0,0625	0,0792	0,0875	0,1125	0,1292	0,1292	0,1333
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0094	0,0094	0,0104	0,0104
Te4	0,0000	0,0458	0,0500	0,0542	0,0583	0,0583	0,0792	0,0833	0,1000	0,1083	0,1208
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,1042	0,1250	0,1292	0,1333	0,1417	0,1500	0,1583	0,1792	0,1875	0,2000

Tabela 6.40 - Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem CBERS.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0469	0,0469	0,0469	0,0469	0,0469	0,0469	0,0469	0,0469	0,0469	0,0479
Te1	0,0000	0,1292	0,1417	0,15	0,1583	0,1667	0,1667	0,1667	0,1708	0,1708	0,1708
Tr2	0,0000	0,0479	0,0479	0,0479	0,0479	0,0479	0,0479	0,0479	0,0479	0,0479	0,049
Te2	0,0000	0,1083	0,1333	0,1458	0,1500	0,1500	0,1500	0,1542	0,1583	0,1583	0,1708
Tr3	0,0000	0,0552	0,0552	0,0552	0,0552	0,0552	0,0552	0,0563	0,0563	0,0563	0,0563
Te3	0,0000	0,1083	0,1375	0,1458	0,1458	0,1500	0,1500	0,1500	0,1500	0,1542	0,1542
Tr4	0,0000	0,0646	0,0646	0,0646	0,0646	0,0646	0,0646	0,0646	0,0646	0,0646	0,0646
Te4	0,0000	0,0708	0,0917	0,1000	0,1083	0,1083	0,1167	0,1250	0,1250	0,1250	0,1292
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,0792	0,0875	0,0875	0,1292	0,1333	0,1417	0,1542	0,1750	0,1792	0,1833

$\alpha = 1$ para o modelo AGRUP. Comparando-se com a família de classificadores do modelo FKNN, a característica de menor imprecisão parcial é verificada. Ainda para o modelo FKNN, o erro supera a imprecisão parcial em valores de α muito baixos, fazendo com que, mesmo para valores de α maiores, a precisão não seja favorecida.

6.2.4.3 Classificação precisa

O classificador AGRUP foi originalmente proposto para a classificação da imagem CBERS em questão, com os resultados apresentados em Drummond e Sandri (2004). O conjunto de amostras empregado é composto por 54 pontos de treinamento e 108 de

Tabela 6.41 - Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\lambda_{tr-1}, \lambda_{tr-1}}^L$ - imagem CBERS.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,5990	0,4646	0,299	0,1208	0,0708	0,0656	0,0073	0,0052	0,0052	0,0000
Te1	0,0000	0,5583	0,4250	0,3333	0,2458	0,1958	0,1458	0,0542	0,0292	0,0167	0,0000
Tr2	0,0000	0,6500	0,4927	0,2875	0,1052	0,0677	0,0615	0,0604	0,0021	0,0021	0,0000
Te2	0,0000	0,6375	0,5083	0,3500	0,3000	0,2500	0,1917	0,1708	0,0708	0,0333	0,0000
Tr3	0,0000	0,5417	0,4406	0,2990	0,1240	0,0792	0,0729	0,0719	0,0677	0,0667	0,0000
Te3	0,0000	0,5042	0,4333	0,3417	0,2333	0,1792	0,1500	0,0875	0,0625	0,0542	0,0000
Tr4	0,0000	0,6042	0,4302	0,2510	0,1104	0,0813	0,0781	0,0031	0,0021	0,0010	0,0000
Te4	0,0000	0,5750	0,4292	0,3250	0,2042	0,1333	0,1000	0,0667	0,0417	0,0250	0,0000
Tr5	0,0000	0,5563	0,4167	0,2594	0,0563	0,0115	0,0094	0,0073	0,0052	0,0010	0,0000
Te5	0,0000	0,5125	0,4208	0,3417	0,2500	0,2000	0,1417	0,0833	0,0458	0,0208	0,0000

Tabela 6.42 - Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 5$ - imagem CBERS.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,1646	0,1094	0,074	0,074	0,0479	0,0479	0,0250	0,0219	0,0219	0,0000
Te1	0,0000	0,1375	0,0958	0,0708	0,0542	0,0333	0,0333	0,0208	0,0125	0,0125	0,0000
Tr2	0,0000	0,1406	0,0938	0,0635	0,0635	0,0396	0,0396	0,0219	0,0187	0,0187	0,0000
Te2	0,0000	0,2000	0,1458	0,1208	0,0958	0,0833	0,0667	0,0583	0,0250	0,0125	0,0000
Tr3	0,0000	0,1458	0,0917	0,0656	0,0646	0,0396	0,0385	0,0250	0,0240	0,0240	0,0000
Te3	0,0000	0,1917	0,0875	0,0708	0,0417	0,0292	0,0167	0,0167	0,0125	0,0042	0,0000
Tr4	0,0000	0,1417	0,0802	0,0583	0,0583	0,0396	0,0396	0,0219	0,0167	0,0167	0,0000
Te4	0,0000	0,2208	0,1292	0,0792	0,0583	0,0583	0,0417	0,0167	0,0167	0,0125	0,0000
Tr5	0,0000	0,1740	0,1063	0,0698	0,0698	0,0417	0,0417	0,0208	0,0208	0,0208	0,0000
Te5	0,0000	0,1875	0,1375	0,1333	0,0875	0,0792	0,0542	0,0375	0,0167	0,0125	0,0000

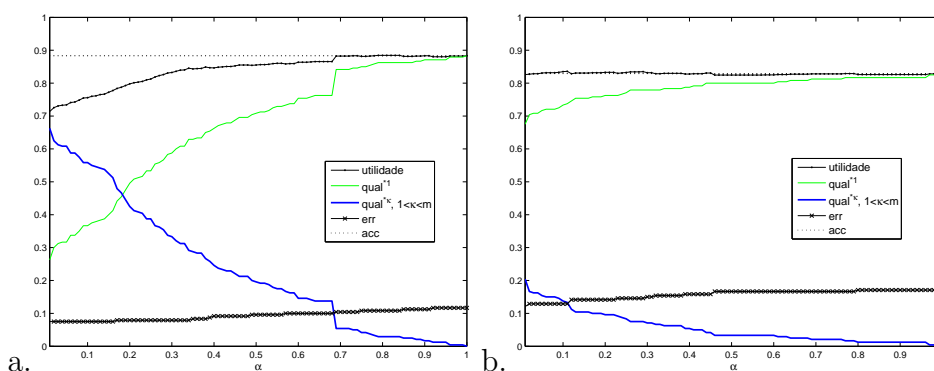


Figura 6.21 - Gráficos de classificação para imagem CBERS: a) classificador AGRUP Te1 e b) classificador FKNN Te1.

teste, para cada uma das classes. A configuração inicial utiliza um número de centros fixo, 15 centros por classe. O processo de classificação envolve a heurística 2-MAX (vide seção 5.6.1), que devolve uma imagem obtida em termos de classes puras e compostas. Na obtenção da classificação final é empregado o refinamento pela vizinhança.

Os experimentos tomam a imagem não normalizada, ou seja, nos limites $[0, 255]$. As relações de similaridade empregadas são lineares cujos parâmetros são como se segue:

- $\lambda_1 = \min(200, \tau * d_{max}), \tau = 20$
- $\gamma_1 = \gamma_{tr2}$
- $\lambda_2 = f(\min(d_{ext}, d_{max} * \tau_1), \max(d_{in}, d_{min} * \tau_2)); \tau_1 = \tau_2 = 2; f(a, b) = (a + b)/2$

Na Tabela 6.43 são mostrados os valores de $kappa'$, antes do refinamento pela vizinhança e $kappa''$, que indica o desempenho do classificador após o refinamento. Os valores correspondem a 4 experimentos realizados variando-se as relações de similaridade empregadas. Todos os experimentos empregam o mesmo conjunto de teste e treinamento em apenas uma execução, sendo os valores de $kappa$ calculado sobre as 108 amostras de teste.

Tabela 6.43 - Valores de $kappa$ antes e após o refinamento pela vizinhança.

	$kappa'$	$kappa''$
Experimento 1	0,757	0,847
Experimento 2	0,77	0,834
Experimento 3	0,757	0,82
Experimento 4	0,667	0,774

Os resultados obtidos são satisfatórios. A abordagem de refinamento pela vizinhança aumenta de forma significativa o desempenho da classificação. Uma classificação imprecisa combinada a uma heurística adequada reflete ganho na solução de um problema de classificação.

6.2.5 Dados de registro de queda na distribuição de energia elétrica

Llanos *et al.* (2003b) apresenta a classificação de registro de queda de distribuição de energia como parte dos sistemas automáticos de monitoramento de falhas. O objetivo é aperfeiçoar a detecção de falhas, buscando sua origem e localização. O conjunto de dados utilizados nesta tese tem a mesma origem, daqueles empregados em Llanos *et al.* (2003b), Llanos *et al.* (2003a).

O conjunto de dados consiste em registros de perturbações que afetam a qualidade da onda elétrica. Os registros são feitos por instrumentos situados em pontos das redes elétricas e caracterizam o fenômeno de perturbação (evento) por atributos que indicam: forma, amplitude máxima, duração e o nível de tensão (médio/alto). O atributo nível de tensão é utilizado para a classificação supervisionada, ou seja, o problema tem duas classes: alta tensão (AT) e média tensão (MT).

6.2.5.1 Espaço de atributos

Para os experimentos realizados são empregados dois atributos: amplitude máxima e duração do evento, que definem um espaço bidimensional. O conjunto de amostras, apresentado na Figura 6.22, contém um total de 80 pontos, 40 por classe. A validação subdivide o conjunto em 5 partes, definindo conjuntos de treinamento com 32 pontos e de teste com 8 pontos.

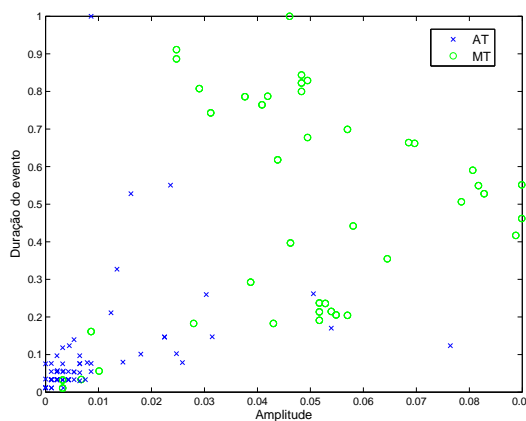


Figura 6.22 - Espaço de atributos formado pelos atributos amplitude e duração do evento.

6.2.5.2 Classificação imprecisa

Neste caso, como o problema tem apenas duas classes, a imprecisão completa coincide com a imprecisão parcial, ou seja, ou o sistema impreciso define pontos como precisos corretamente classificados ou a saída é a imprecisão completa. A relação de similaridade que se mostrou mais adequada para o método AGRUP é a linear $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ e o método FKNN teve seu comportamento avaliado para $k = 1$ e $k = 3$. A configuração inicial dada pelo agrupamento define os seguintes números de centros para cada classe, em cada experimento: $[T_{exp1} < 11; 20 >, T_{exp2} < 12; 18 >, T_{exp3} < 10; 13 >, T_{exp4} < 7; 11 >, T_{exp5} < 10; 19 >]$.

Avaliando o acerto preciso (Tabelas 6.44, 6.45 e 6.46), em $\alpha = 1$, os modelos apresentam, na média, resultados similares. A imprecisão completa mostra a flexibilidade maior do método AGRUP quando inserido no ambiente da classificação imprecisa destacada neste trabalho. As Tabelas de imprecisão completa são referenciadas como 6.47, 6.48 e 6.49 e as Tabelas de erro são 6.50, 6.51 e 6.52.

Tabela 6.44 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - Dados registro de rede elétrica.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,4375	0,5000	0,5938	0,9219	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Te1	0,0000	0,3125	0,3125	0,3125	0,6875	0,7500	0,7500	0,7500	0,7500	0,8750	1,0000
Tr2	0,0000	0,4844	0,5000	0,5625	0,8750	0,9844	0,9844	0,9844	0,9844	0,9844	1,0000
Te2	0,0000	0,5000	0,5000	0,5000	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375
Tr3	0,0000	0,4844	0,4844	0,5313	0,9375	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Te3	0,0000	0,3750	0,5000	0,5000	0,6250	0,6875	0,6875	0,6875	0,6875	0,6875	0,6875
Tr4	0,0000	0,4688	0,7031	0,7656	0,8438	0,9063	0,9219	0,9219	0,9531	0,9688	0,9844
Te4	0,0000	0,5000	0,5625	0,5625	0,5625	0,5625	0,5625	0,5625	0,6875	0,6875	0,6875
Tr5	0,0000	0,3281	0,5000	0,5313	0,9219	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Te5	0,0000	0,4375	0,5000	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500

Tabela 6.45 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 1$ - Dados registro de rede elétrica.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,7031	0,8594	0,8594	0,8594	0,9375	0,9375	0,9375	0,9375	0,9375	1,0000
Te1	0,0000	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	1,0000
Tr2	0,0000	0,7969	0,8594	0,8594	0,8594	0,9219	0,9219	0,9219	0,9219	0,9219	1,0000
Te2	0,0000	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375
Tr3	0,0000	0,7813	0,9219	0,9219	0,9219	0,9531	0,9531	0,9531	0,9531	0,9531	1,0000
Te3	0,0000	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500	0,8750
Tr4	0,0000	0,7969	0,9375	0,9375	0,9375	0,9531	0,9531	0,9531	0,9531	0,9531	1,0000
Te4	0,0000	0,6875	0,6875	0,6875	0,6875	0,6875	0,6875	0,6875	0,6875	0,6875	0,6875
Tr5	0,0000	0,7969	0,8281	0,8281	0,8281	0,9375	0,9375	0,9375	0,9375	0,9375	1,0000
Te5	0,0000	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500

Tabela 6.46 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados registro de rede elétrica.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,7031	0,8594	0,8594	0,8594	0,9375	0,9375	0,9375	0,9375	0,9375	1,0000
Te1	0,0000	0,5625	0,6875	0,7500	0,8750	0,9375	0,9375	0,9375	0,9375	0,9375	1,0000
Tr2	0,0000	0,7969	0,8594	0,8594	0,8594	0,9219	0,9219	0,9219	0,9219	0,9219	1,0000
Te2	0,0000	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375	0,9375
Tr3	0,0000	0,7813	0,9219	0,9219	0,9219	0,9531	0,9531	0,9531	0,9531	0,9531	1,0000
Te3	0,0000	0,6250	0,7500	0,8125	0,8125	0,8125	0,8125	0,8125	0,8125	0,8125	0,8125
Tr4	0,0000	0,7969	0,9375	0,9375	0,9375	0,9531	0,9531	0,9531	0,9531	0,9531	1,0000
Te4	0,0000	0,5625	0,6875	0,6875	0,6875	0,6875	0,6875	0,6875	0,6875	0,6875	0,6875
Tr5	0,0000	0,7969	0,8281	0,8281	0,8281	0,9375	0,9375	0,9375	0,9375	0,9375	1,0000
Te5	0,0000	0,6250	0,6250	0,6250	0,6250	0,7500	0,7500	0,7500	0,7500	0,7500	0,7500

Tabela 6.47 - Imprecisão completa em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^L$ - Dados registro de rede elétrica.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	1,0000	0,5625	0,5000	0,4063	0,0781	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	1,0000	0,6875	0,6875	0,6875	0,3125	0,2500	0,2500	0,2500	0,2500	0,1250	0,0000
Tr2	1,0000	0,5156	0,5000	0,4375	0,1250	0,0156	0,0156	0,0156	0,0156	0,0156	0,0000
Te2	1,0000	0,5000	0,5000	0,5000	0,0625	0,0625	0,0625	0,0625	0,0000	0,0000	0,0000
Tr3	1,0000	0,5156	0,5156	0,4688	0,0625	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	1,0000	0,5625	0,3750	0,3750	0,2500	0,1250	0,1250	0,1250	0,0000	0,0000	0,0000
Tr4	1,0000	0,5313	0,2969	0,2344	0,1563	0,0938	0,0781	0,0781	0,0469	0,0156	0,0000
Te4	1,0000	0,3125	0,2500	0,1875	0,1875	0,1875	0,1875	0,1875	0,0625	0,0000	0,0000
Tr5	1,0000	0,6719	0,5000	0,4688	0,0781	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	1,0000	0,3125	0,2500	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Tabela 6.48 - Imprecisão completa em 5 experimentos usando o classificador FKNN com $k = 1$ - Dados registro de rede elétrica.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	1,0000	0,2969	0,1406	0,1406	0,1406	0,0625	0,0625	0,0625	0,0625	0,0625	0,0000
Te1	1,0000	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0000
Tr2	1,0000	0,2031	0,1406	0,1406	0,1406	0,0781	0,0781	0,0781	0,0781	0,0781	0,0000
Te2	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr3	1,0000	0,2188	0,0781	0,0781	0,0781	0,0469	0,0469	0,0469	0,0469	0,0469	0,0000
Te3	1,0000	0,1875	0,1875	0,1875	0,1875	0,1875	0,1875	0,1875	0,1875	0,1875	0,0000
Tr4	1,0000	0,2031	0,0625	0,0625	0,0625	0,0469	0,0469	0,0469	0,0469	0,0469	0,0000
Te4	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr5	1,0000	0,2031	0,1719	0,1719	0,1719	0,0625	0,0625	0,0625	0,0625	0,0625	0,0000
Te5	1,0000	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0000

Tabela 6.49 - Imprecisão completa em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados registro de rede elétrica.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	1,0000	0,2969	0,1406	0,1406	0,1406	0,0625	0,0625	0,0625	0,0625	0,0625	0,0000
Te1	1,0000	0,4375	0,3125	0,25	0,125	0,0625	0,0625	0,0625	0,0625	0,0625	0,0000
Tr2	1,0000	0,2031	0,1406	0,1406	0,1406	0,0781	0,0781	0,0781	0,0781	0,0781	0,0000
Te2	1,0000	0,0625	0,0625	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr3	1,0000	0,2188	0,0781	0,0781	0,0781	0,0469	0,0469	0,0469	0,0469	0,0469	0,0000
Te3	1,0000	0,3125	0,1875	0,125	0,0625	0,0625	0,0000	0,0000	0,0000	0,0000	0,0000
Tr4	1,0000	0,2031	0,0625	0,0625	0,0625	0,0469	0,0469	0,0469	0,0469	0,0469	0,0000
Te4	1,0000	0,1875	0,0625	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr5	1,0000	0,2031	0,1719	0,1719	0,1719	0,0625	0,0625	0,0625	0,0625	0,0625	0,0000
Te5	1,0000	0,1875	0,1875	0,1875	0,1875	0,0625	0,0625	0,0625	0,0000	0,0000	0,0000

Tabela 6.50 - Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr1}, \lambda_{tr1}}^{GL}$ - Dados registro de rede elétrica.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0625	0,0625	0,0625
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	0,0000	0,0625	0,1250	0,1250	0,1250	0,1875	0,1875	0,1875	0,3125	0,3125	0,3125
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0156	0,0156
Te4	0,0000	0,1875	0,1875	0,2500	0,2500	0,2500	0,2500	0,2500	0,2500	0,3125	0,3125
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,2500	0,2500	0,2500	0,2500	0,2500	0,2500	0,2500	0,2500	0,2500	0,2500

Tabela 6.51 - Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 1$ - Dados registro de rede elétrica.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	0,0000	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	0,0000	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,1250
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	0,0000	0,3125	0,3125	0,3125	0,3125	0,3125	0,3125	0,3125	0,3125	0,3125	0,3125
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,1875	0,1875	0,1875	0,1875	0,1875	0,1875	0,1875	0,1875	0,1875	0,2500

Tabela 6.52 - Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados registro de rede elétrica.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	0,0000	0,0000	0,0000	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625	0,0625
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	0,0000	0,0625	0,0625	0,0625	0,1250	0,1250	0,1875	0,1875	0,1875	0,1875	0,1875
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	0,0000	0,2500	0,2500	0,3125	0,3125	0,3125	0,3125	0,3125	0,3125	0,3125	0,3125
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,1875	0,1875	0,1875	0,1875	0,1875	0,1875	0,1875	0,2500	0,2500	0,2500

Os gráficos de classificação apresentados na Figura 6.23 referem-se ao melhor e pior casos alcançados para a classificação precisa, em $\alpha = 1$, em cada configuração estudada nas tabelas acima. O melhor caso se refere ao experimento 2 (Te2) e o pior, ao experimento 4 (Te4). Para o melhor caso empregando o modelo FKNN, usando 3 vizinhos (Figuras 6.23e), constata-se em $\alpha < 0,25$ a imprecisão, que mais tarde se transforma em erro; o mesmo erro verificado para o modelo FKNN empregando apenas 1 vizinho. Para o modelo AGRUP, o melhor caso (Figura 6.23a) é completamente impreciso até valores de α próximos de 0,8. Em problemas com apenas duas classes, a imprecisão completa pode parecer não muito útil, porém, para um sistema de detecção de falhas numa rede elétrica, como é o caso, o custo da imprecisão pode compensar o erro.

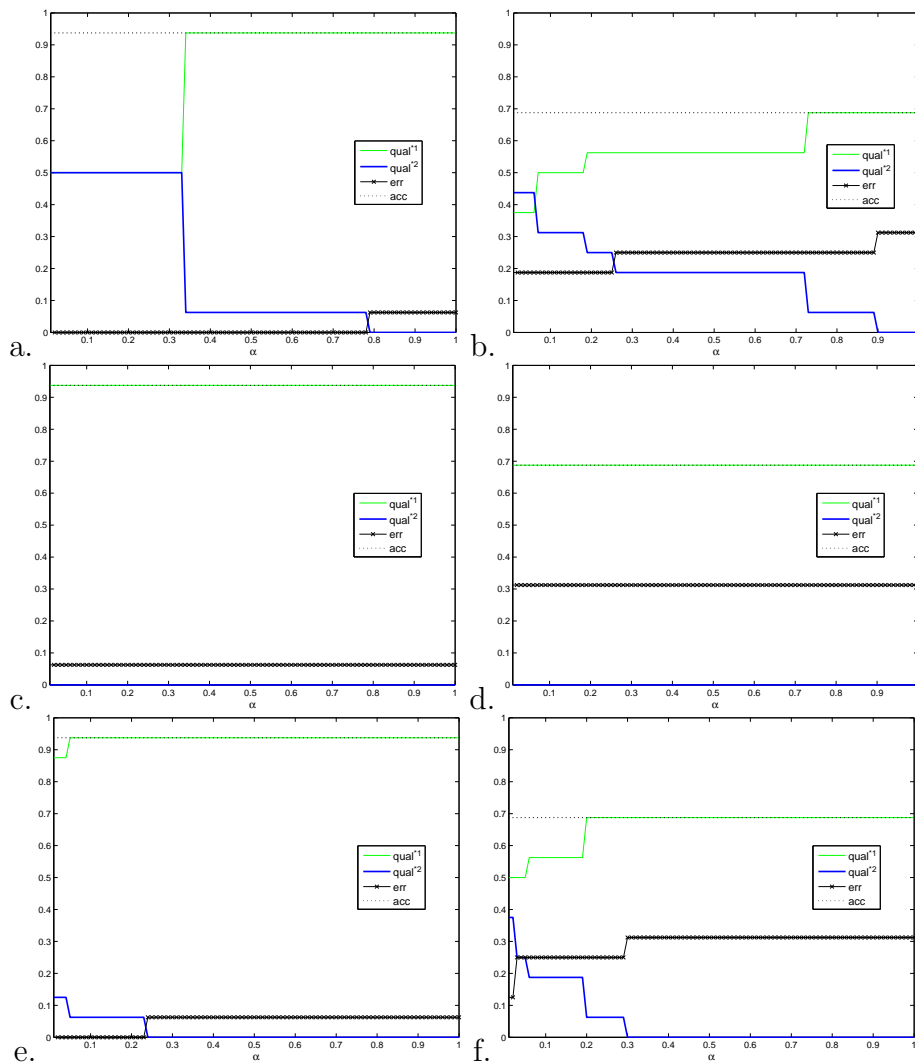


Figura 6.23 - Gráficos de classificação para os dados de registro de rede elétrica do classificador AGRUP com $S^L_{\gamma_{tr1}, \lambda_{tr1}}$ e FKNN com $k = 1$ e $k = 3$: a) AGRUP Te2; b) AGRUP Te4; c) FKNN com $k = 1$ Te2; d) FKNN com $k = 1$ Te4; e) FKNN com $k = 3$ Te2 ; f) FKNN com $k = 3$ Te4.

6.2.5.3 Combinação de classificadores

Em Drummond *et al.* (2006) é apresentada uma aplicação, usando um subconjunto destes dados, onde é empregado o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$. Os dois classificadores agregados se diferem apenas no conjunto de centros empregados na fase inicial. Foram calculados 5 centros utilizando o mesmo método, o FCM, com inicialização aleatória. Não é empregado, neste caso, o meta-agrupamento proposto, como na maioria das aplicações descritas.

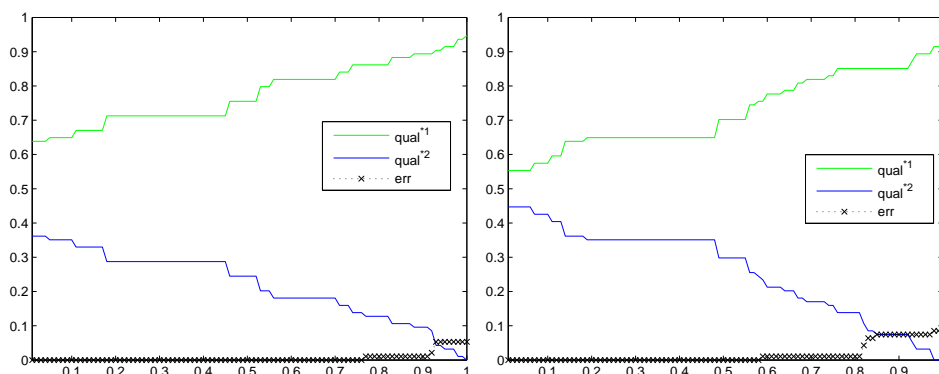


Figura 6.24 - Gráficos de classificação, Cl^1 e Cl^2 , para o conjunto de dados de registros de rede elétrica. Ambos empregando o método AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ e número de centros fixos em 5 para cada classe.

Na Figura 6.25 são apresentados os classificadores $Cl_{\alpha1;0,58}^{12}$ e $Cl_{0,60;\alpha2}^{12}$, obtidos pela agregação dos classificadores Cl^1 e Cl^2 . Comparando a família de classificadores obtidos para a agregação, o resultado proveniente da combinação apresenta queda na imprecisão sem aumento do erro, o que evidencia a vantagem de se trabalhar com as combinações em lugar dos classificadores individuais.

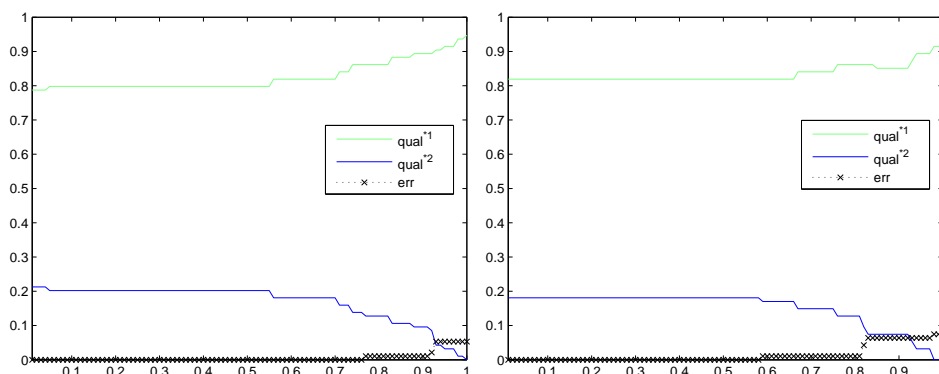


Figura 6.25 - Gráficos de classificação: a) $Cl_{\alpha1;0,58}^{12}$ e b) $Cl_{0,60;\alpha2}^{12}$.

6.2.6 Dados do projeto GeoSchisto

O projeto GeoSchisto (GEOSCHISTO, 2007) propõe o desenvolvimento de modelos de análise, representação e construção de cenários espaciais que permitam a caracterização de fenômenos relacionados à distribuição da esquistossomose em Minas Gerais. O objetivo principal do projeto é determinar as relações entre a distribuição dos hospedeiros intermediários e a prevalência da esquistossomose no Estado de Minas Gerais, utilizando os recursos de geoprocessamento e de imagens ambientais produzidas por satélites.

A esquistossomose mansoni é uma doença endêmica, conhecida pelos brasileiros como barriga d'água, xistosa ou doença do caramujo. O diagnóstico e o tratamento são simples, mas a erradicação da doença só é possível com medidas que interrompam o ciclo evolutivo do parasita, como a realização de saneamento básico e a mudança do comportamento das pessoas que vivem em áreas de risco.

O projeto dispõe de uma base de dados da área de estudos contendo variáveis climáticas, sociais e de sensoriamento remoto que caracterizam a região. Além disto, dados históricos de prevalência da doença são conhecidos e um dos pontos de estudo do projeto é, justamente, a estimativa da prevalência da esquistossomose através de modelos de classificação de dados que podem, de alguma maneira, caracterizar as regiões que têm maior ou menor chance de ocorrência da doença.

Com base nos trabalhos publicados por Guimarães *et al.* (2006), Freitas *et al.* (2006), Martins *et al.* (2007), Fonseca *et al.* (2007), foi estudado o comportamento de um classificador impreciso nesta aplicação.

6.2.6.1 Espaços de atributos

A área de estudo é o Estado de Minas Gerais que tem área de aproximadamente 590.000km^2 , dividido politicamente em 853 municípios, com 18 milhões de habitantes, em média, segundo dados fornecidos pelo IBGE (2007). A porcentagem de casos positivos da doença em relação a um mínimo de 80% da população é definido como prevalência da doença num determinado município. A prevalência dos dados é disponibilizada pela Secretaria de Vigilância em Saúde e Secretaria de Estado de Saúde de Minas Gerais e consiste em dados históricos de 197 municípios. Existem associadas a cada município, em média, 90 variáveis distintas, que são informações socioeconômicas, demográficas, de saneamento básico, presença de água e vigor da vegetação no verão e no inverno, clima e relevo. É evidente a necessidade de um estudo dos atributos que

caracterizam cada município, no intuito de extrair no banco de dados as informações que sejam realmente relevantes, e com isso, buscar a melhor classificação dos dados. O presente trabalho não se dispõe a tal estudo, se atendo apenas à etapa de classificação. Com isto, foi empregado os atributos considerados interessantes por Martins *et al.* (2007), que utiliza árvores de decisão e por Fonseca *et al.* (2007), que usa modelos de regressão. E ainda, outros testes, com um número menor de variáveis foi realizado como simples experimentos avaliando os classificadores aqui propostos.

Martins *et al.* (2007) determina 4 classes para o problema, definindo as categorias de prevalência: baixa (0 a 5%); média (>5 a 15%); alta (>15 a 25%) e muito alta (>25%). Trabalhos mais recentes, em vias de conclusão dentro do Projeto, simplificam o problema em apenas 3 classes: baixa, média e alta.

Com os dados disponibilizados, os experimentos foram realizados considerando 3 classes, variando o conjunto de atributos. Num primeiro momento, foram empregadas 5 variáveis, 3 delas de sensoriamento remoto: VEG_I (Modelo linear de mistura espectral - vegetação no inverno), AC_mediana (Mediana da acumulação hídrica) e DEC (Declividade do terreno); 1 variável climática, Tmin_V (Média de temperatura mínima no verão) e 1 variável social, SAN1 (% de domicílios com esgoto ligado a rio ou lago). As variáveis selecionadas são provenientes de estudos realizados pelo Projeto. Dentro do conjunto total de amostras, foram selecionadas 45 amostras por classe, totalizando um conjunto com 135 amostras. Apenas para ilustrar, o espaço de atributos formado por 3, das 5 variáveis empregadas, é apresentado na Figura 6.26.

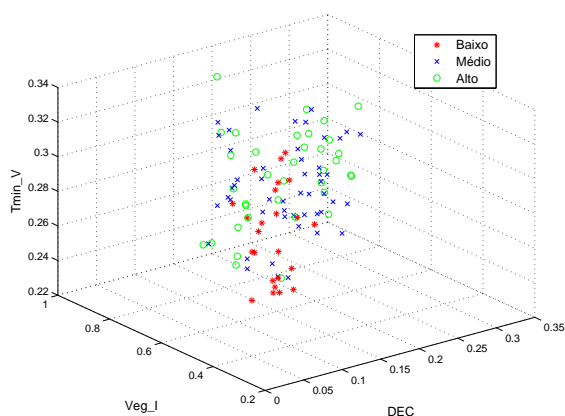


Figura 6.26 - Espaço de atributos formado pelos dados do Projeto GeoSchisto, caracterizados pelas variáveis DEC, Veg_I e Tmin_V.

6.2.6.2 Classificação imprecisa

As tabelas de resultados apresentadas a seguir, são provenientes dos métodos AGRUP com a similaridade $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ e do FKNN com $k = 3$. Pode-se observar pelas Tabelas de precisão 6.53 e 6.54 que, para valores de α próximos de 1, ou seja, no caso mais preciso dentro da família parametrizada, as duas estratégias têm alta precisão para o treinamento, ao contrário do que dizem os baixos valores verificados para o conjunto de teste. São classificadores que respondem perfeitamente ao treinamento, para todos os experimentos. Porém, não apresentam boa generalização, uma vez que a taxa de acerto preciso para as amostras de teste são baixas. Este comportamento pode caracterizar um super treinamento, ou ainda evidenciar que os atributos que estabelecem as classes não são bem definidos.

Observa-se alta imprecisão parcial (Tabelas 6.57 e 6.58), para os dois modelos de classificação. O gráfico da Figura 6.27 mostra um experimento em que a imprecisão do FKNN, indicado como método 2 (met2), é superior ao do AGRUP (met1) para valores de $\alpha < 0,6$. Este comportamento está relacionado aos dados e não ao método em si. O espaço de atributos representado na Figura 6.26, mesmo com apenas 3 variáveis, mostra que as classes têm grande confusão, o que faz com que os graus de probabilidade atribuídos aos elementos da amostra no método FKNN sejam valores próximos para classes distintas.

As medidas de erro se encontram nas Tabelas 6.55 e 6.56, e a imprecisão total nas Tabelas 6.59 e 6.60.

Tabela 6.53 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - Dados Projeto GeoSchisto.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,2407	0,2963	0,4167	0,8981	0,9537	0,9722	1,0000	1,0000	1,0000	1,0000
Te1	0,0000	0,1852	0,1852	0,1852	0,2593	0,2593	0,2593	0,2593	0,3333	0,3704	0,4444
Tr2	0,0000	0,2778	0,3241	0,4259	0,9074	0,9630	0,9722	0,9907	1,0000	1,0000	1,0000
Te2	0,0000	0,3704	0,4074	0,4074	0,4074	0,4444	0,4815	0,4815	0,5926	0,6296	0,6296
Tr3	0,0000	0,2593	0,3056	0,4259	0,9537	0,9722	1,0000	1,0000	1,0000	1,0000	1,0000
Te3	0,0000	0,2222	0,2593	0,2593	0,2963	0,4074	0,4444	0,4444	0,5556	0,5556	0,5556
Tr4	0,0000	0,2963	0,3889	0,4815	0,9074	0,9352	0,963	1,0000	1,000	1,0000	1,0000
Te4	0,0000	0,1481	0,1481	0,1481	0,1481	0,1481	0,2222	0,2593	0,2963	0,2963	0,3333
Tr5	0,0000	0,2778	0,3148	0,3981	0,8704	0,9444	0,9722	1,0000	1,0000	1,0000	1,0000
Te5	0,0000	0,2222	0,2593	0,3333	0,3704	0,3704	0,4074	0,4074	0,4074	0,4444	0,5185

Tabela 6.54 - Medidas de precisão (elementos corretamente classificados com cardinalidade 1) em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados Projeto GeoSchisto.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,1667	0,4167	0,5648	0,5648	0,7963	0,7963	0,8981	0,8981	0,8981	1,0000
Te1	0,0000	0,037	0,037	0,0741	0,2222	0,2593	0,2963	0,4074	0,4074	0,4074	0,4444
Tr2	0,0000	0,1296	0,4074	0,5185	0,5185	0,7407	0,7407	0,9444	0,9444	0,9444	1,0000
Te2	0,0000	0,1481	0,2222	0,2222	0,2963	0,3704	0,3704	0,4074	0,4815	0,5556	0,5556
Tr3	0,0000	0,1574	0,3981	0,4907	0,4907	0,7315	0,7315	0,9259	0,9259	0,9259	1,0000
Te3	0,0000	0,1481	0,1481	0,2222	0,3333	0,3704	0,4815	0,6296	0,6296	0,6296	0,6667
Tr4	0,0000	0,1944	0,4537	0,6389	0,6389	0,8148	0,8148	0,9444	0,9444	0,9444	1,0000
Te4	0,0000	0,1111	0,1111	0,1111	0,1111	0,1852	0,2593	0,3333	0,3333	0,3704	0,3704
Tr5	0,0000	0,1204	0,4444	0,5648	0,5648	0,8056	0,8056	0,9167	0,9167	0,9167	1,0000
Te5	0,0000	0,0741	0,1481	0,2222	0,2593	0,2593	0,3333	0,4815	0,4815	0,5926	0,5926

Tabela 6.55 - Medidas de erro em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - Dados Projeto GeoSchisto.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	0,0000	0,2222	0,2222	0,3333	0,3704	0,4074	0,4074	0,4074	0,4444	0,4444	0,4815
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	0,0000	0,1852	0,2222	0,2222	0,2222	0,2963	0,2963	0,2963	0,3333	0,3333	0,3704
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	0,0000	0,2222	0,2222	0,2963	0,3333	0,4074	0,4074	0,4074	0,4074	0,4074	0,4074
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	0,0000	0,2593	0,2963	0,3333	0,3704	0,3704	0,4444	0,5185	0,5926	0,6296	0,6296
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,1852	0,2222	0,2963	0,2963	0,2963	0,3333	0,3333	0,3333	0,4074	0,4815

Tabela 6.56 - Medidas de erro em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados Projeto GeoSchisto.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	0,0000	0,1481	0,2222	0,2593	0,3333	0,3333	0,3704	0,4444	0,4815	0,5185	0,5556
Tr2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	0,0000	0,0370	0,1481	0,1481	0,1481	0,1852	0,2222	0,2593	0,4074	0,4074	0,4444
Tr3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	0,0000	0,1111	0,1111	0,1481	0,1852	0,1852	0,2593	0,2963	0,3333	0,3333	0,3333
Tr4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	0,0000	0,1481	0,2222	0,2593	0,2963	0,3704	0,4074	0,5556	0,6296	0,6296	0,6296
Tr5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,0741	0,1481	0,1852	0,2222	0,2593	0,2963	0,3704	0,3704	0,4074	0,4074

Tabela 6.57 - Imprecisão parcial em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - Dados Projeto GeoSchisto.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,5463	0,6019	0,5370	0,1019	0,0463	0,0278	0,0000	0,0000	0,0000	0,0000
Te1	0,0000	0,2963	0,3333	0,2963	0,2963	0,2593	0,2593	0,2593	0,1481	0,1111	0,0000
Tr2	0,0000	0,4815	0,5093	0,4907	0,0926	0,0370	0,0278	0,0093	0,0000	0,0000	0,0000
Te2	0,0000	0,3333	0,3333	0,3333	0,3333	0,2593	0,2222	0,2222	0,0741	0,0370	0,0000
Tr3	0,0000	0,5185	0,5185	0,5000	0,0463	0,0278	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	0,0000	0,4074	0,4074	0,3333	0,2963	0,1111	0,0741	0,0741	0,0000	0,0000	0,0000
Tr4	0,0000	0,4630	0,4259	0,3981	0,0833	0,0556	0,0370	0,0000	0,0000	0,0000	0,0000
Te4	0,0000	0,3333	0,3704	0,4815	0,4444	0,4444	0,2963	0,1852	0,0741	0,0370	0,0000
Tr5	0,0000	0,4259	0,4907	0,5093	0,1296	0,0556	0,0278	0,0000	0,0000	0,0000	0,0000
Te5	0,0000	0,4815	0,4444	0,3704	0,3333	0,3333	0,2593	0,2593	0,2593	0,1481	0,0000

Tabela 6.58 - Imprecisão parcial em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados Projeto GeoSchisto.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	0,0000	0,5833	0,3333	0,3333	0,4352	0,2037	0,2037	0,1019	0,1019	0,1019	0,0000
Te1	0,0000	0,3704	0,4074	0,4444	0,3333	0,2963	0,2222	0,1111	0,0741	0,0741	0,0000
Tr2	0,0000	0,5556	0,2778	0,2778	0,4815	0,2593	0,2593	0,0556	0,0556	0,0556	0,0000
Te2	0,0000	0,3704	0,2963	0,4074	0,3333	0,3704	0,3704	0,3333	0,1111	0,0370	0,0000
Tr3	0,0000	0,5556	0,3148	0,3148	0,5093	0,2685	0,2685	0,0741	0,0741	0,0741	0,0000
Te3	0,0000	0,5185	0,5926	0,5185	0,4074	0,3704	0,1852	0,037	0,0370	0,0370	0,0000
Tr4	0,0000	0,4907	0,2315	0,2315	0,3611	0,1852	0,1852	0,0556	0,0556	0,0556	0,0000
Te4	0,0000	0,1852	0,2593	0,3333	0,2963	0,2593	0,2593	0,0741	0,037	0,0000	0,0000
Tr5	0,0000	0,6481	0,3241	0,3241	0,4352	0,1944	0,1944	0,0833	0,0833	0,0833	0,0000
Te5	0,0000	0,4074	0,3704	0,2963	0,4074	0,4444	0,3333	0,1111	0,1481	0,0000	0,0000

Tabela 6.59 - Imprecisão completa em 5 experimentos usando o classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$ - Dados Projeto GeoSchisto.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	1,0000	0,213	0,1019	0,0463	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	1,0000	0,2222	0,1852	0,1111	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr2	1,0000	0,2407	0,1667	0,0833	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	1,0000	0,1111	0,0370	0,0370	0,0370	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr3	1,0000	0,2222	0,1759	0,0741	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	1,0000	0,1111	0,0741	0,0741	0,0370	0,0370	0,0370	0,0370	0,0000	0,0000	0,0000
Tr4	1,0000	0,2407	0,1852	0,1204	0,0093	0,0093	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	1,0000	0,2222	0,1481	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Tr5	1,0000	0,2963	0,1944	0,0926	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	1,0000	0,1111	0,0741	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Tabela 6.60 - Imprecisão completa em 5 experimentos usando o classificador FKNN com $k = 3$ - Dados Projeto GeoSchisto.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Tr1	1,0000	0,2500	0,2500	0,1019	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te1	1,0000	0,4444	0,3333	0,2222	0,1111	0,1111	0,1111	0,0370	0,0370	0,0000	0,0000
Tr2	1,0000	0,3148	0,3148	0,2037	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te2	1,0000	0,4444	0,3333	0,2222	0,2222	0,0741	0,0370	0,0000	0,0000	0,0000	0,0000
Tr3	1,0000	0,2870	0,2870	0,1944	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te3	1,0000	0,2222	0,1481	0,1111	0,0741	0,0741	0,0741	0,0370	0,0000	0,0000	0,0000
Tr4	1,0000	0,3148	0,3148	0,1296	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te4	1,0000	0,5556	0,4074	0,2963	0,2963	0,1852	0,0741	0,0370	0,0000	0,0000	0,0000
Tr5	1,0000	0,2315	0,2315	0,1111	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Te5	1,0000	0,4444	0,3333	0,2963	0,1111	0,0370	0,0370	0,0370	0,0000	0,0000	0,0000

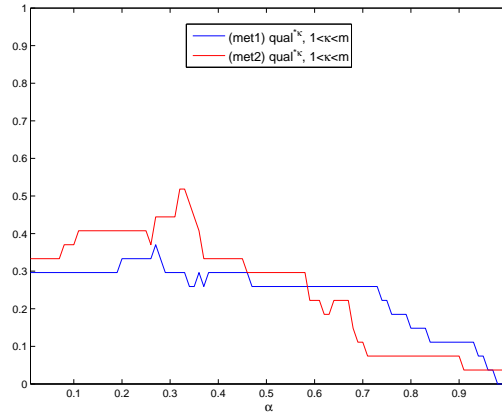


Figura 6.27 - Imprecisão parcial referente ao experimento 1 (Te1) para os métodos AGRUP (met1) e FKNN (met2).

Os gráficos de classificação dos experimentos de teste 2 e 3, para os modelos AGRUP e FKNN são apresentados na Figura 6.28. Pode-se observar a imprecisão completa bem acentuada nos gráficos das Figuras 6.28b e 6.28d. E ainda, nota-se que a curva de utilidade de imprecisão para o modelo AGRUP, nos dois experimentos, se apresenta ligeiramente acima da acurácia para valores $0,8 < \alpha < 0,99$ (Figuras 6.28a e 6.28b).

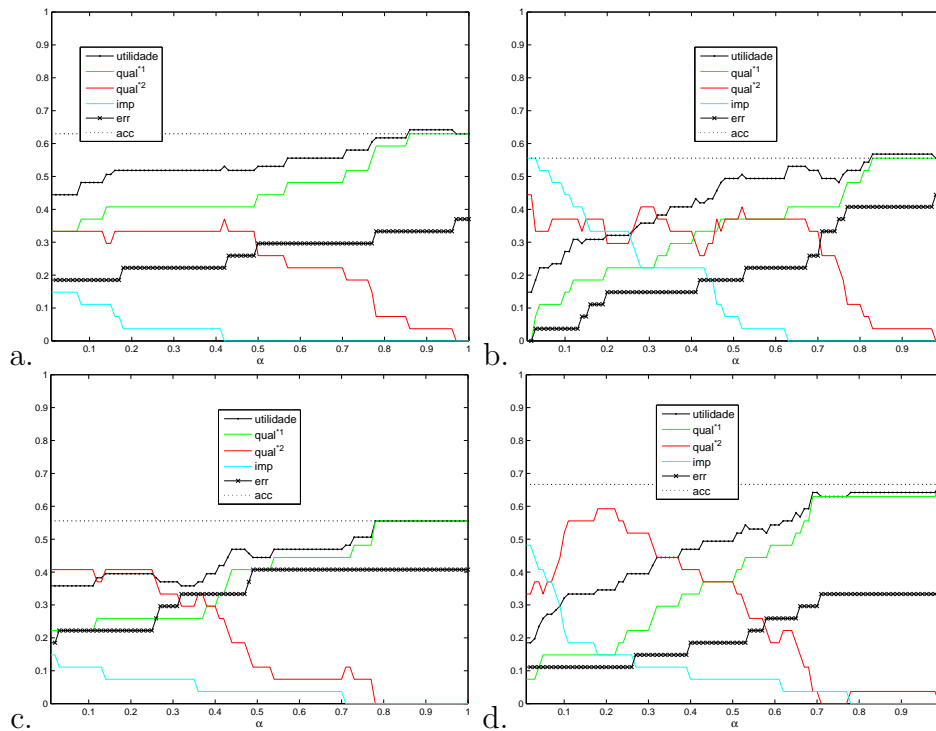


Figura 6.28 - Gráficos de classificação para os dados do Projeto Schisto: a) AGRUP Te2; b) FKNN Te2; c) AGRUP Te3 e d) FKNN Te3.

6.2.6.3 Classificação precisa

Outro experimento realizado utiliza 142 amostras de treinamento e 55 de teste, em uma única execução do classificador AGRUP com $S_{\gamma_{tr2}, \lambda_{tr1}}^L$. São empregadas 44 variáveis, definidas em Martins *et al.* (2007), das quais 22 são derivadas de dados de SR, 6 são climáticas e 16 são socioeconômicas. Das variáveis de SR, 18 delas foram derivadas do sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*) e 4 do SRTM (*Shuttle Radar Topography Mission*). As variáveis climáticas foram obtidas através da plataforma de coleta de dados do CPTEC/INPE. E, por fim, as variáveis socioeconômicas foram obtidas do IBGE.

A classificação precisa é obtida tomando como classe aquela cujo grau de pertinência é máximo. E a avaliação do desempenho do classificador é feito através das matrizes de confusão, apresentadas para as amostras de treinamento e teste, como pode ser observado na Tabela 6.61.

Tabela 6.61 - Matrizes de confusão: a) amostras de treinamento e b) amostras de teste

a)	Baixa	Média	Alta	b)	Baixa	Média	Alta
Baixa (34)	34	1	0	Baixa (12)	5	5	2
Média (48)	0	46	0	Média (25)	1	9	4
Alta (60)	0	1	60	Alta (18)	6	11	12

Pode-se observar que o índice de acerto no treinamento é muito alto, o que não acontece no teste. O classificador é praticamente perfeito quando classifica as amostras de treinamento com uma taxa de acerto $acc = 0,98$; e falha na classificação das amostras de teste cuja taxa de acerto é $acc = 0,47$, caracterizando o mesmo comportamento observado para os testes realizados com 5 variáveis. Os resultados alcançados são comparáveis àqueles apresentados na literatura, Martins *et al.* (2007) e Fonseca *et al.* (2007), evidenciando a necessidade de estudo dos atributos que caracterizam os dados.

6.3 Considerações Finais

Os experimentos realizados buscaram avaliar a abordagem imprecisa proposta. Para tanto foram usados conjuntos de dados sintéticos e reais, de natureza bem distinta, com e sem ruído, definidos em espaços multidimensionais.

Os experimentos apontam para a viabilidade da classificação imprecisa, especificamente da metodologia proposta. Dentro da abordagem imprecisa, com criação de uma família

parametrizada de classificadores imprecisos, o modelo AGRUP se mostrou flexível com relação à sua adequação a aplicações que necessitem de mais ou menos precisão. No que tange à imprecisão, o AGRUP também se mostrou robusto à presença ou ausência de ruído.

A combinação de classificadores se mostrou viável com base no comportamento de cada classificador em separado, considerando as medidas de precisão, erro, imprecisão parcial e completa. O mecanismo impreciso dá suporte à tomada de decisão, quando, para uma dada aplicação, o custo da imprecisão, ainda que completa, compensa o erro. Pôde-se verificar que a combinação realizada para alguns dos experimentos favoreceu o classificador AGRUP concernente à precisão, apresentando ainda uma diminuição significativa no erro.

Um outro processo investigado neste trabalho, o refinamento, mostrou-se importante para se obter um classificador preciso. Os resultados obtidos para as imagens de satélite, submetidas a tal vizinhança, tiveram um ganho significativo, deixando evidente as vantagens em se combinar classificação no espaço de atributos com refinamento no espaço do problema, como forma de agregar informação presente na vizinhança.

CAPÍTULO 7

CONCLUSÃO

A tarefa de classificação de dados é, na maior parte das vezes, sujeita a imperfeições, o que ocorre, por exemplo, quando os dados são ruidosos, quando o conjunto de atributos necessário para classificar um dado está incompleto, ou quando o próprio conceito que define uma classe é vago. Para lidar com estas imperfeições, os métodos de classificação definidos na literatura empregam teorias que tratam de incerteza, imprecisão ou do caráter vago da informação. A teoria de possibilidades é particularmente interessante neste contexto, já que pode modelar todos estes aspectos.

Nos classificadores possibilistas, o resultado da classificação de um dado é uma distribuição de possibilidades no domínio formado pelo conjunto das classes, onde o grau atribuído a uma classe reflete o quanto o classificador “acredita” que aquela é a classe correta do dado.

Normalmente, ter somente uma distribuição de possibilidade sobre a classe de um ponto pode não ser suficiente em uma dada aplicação; na maior parte das vezes, é necessário tomar-se uma decisão sobre a classificação dos dados. A primeira solução é, naturalmente, escolher a classe com o maior grau de possibilidade e definir critérios para tratar casos de empate.

Esta solução pode não ser conveniente no caso de existência de imperfeições. Uma outra solução, aceitável em algumas aplicações, é a de retornar não (necessariamente) somente a classe com o maior grau de possibilidade, mas um subconjunto de classes, escolhido levando-se em conta informação global sobre todos os dados que estão sendo classificados. O uso de classificadores imprecisos diminui o erro à custa da precisão. A escolha de uma classe para um elemento como sendo somente aquela com maior grau é um tipo de tratamento local, enquanto que a derivação de classes compostas é uma abordagem mais global.

Este trabalho apresenta a obtenção de uma família parametrizada de classificadores imprecisos a partir de distribuições de possibilidade provenientes de um método de classificação. Uma coleção de índices de classificação é derivada com base na partição do conjunto de dados segundo precisão e cardinalidade. O elemento pode ser classificado de forma correta ou incorreta, e ter associado a ele qualquer subconjunto de classes definidas no problema. Desta forma, são definidas as medidas de precisão, imprecisão parcial e imprecisão completa para os elementos classificados correta ou incorretamente.

É proposto ainda um novo método de classificação possibilista baseado em agrupamento, o AGRUP, através do qual se associa, a cada elemento do conjunto de dados, uma distribuição de possibilidade. Na etapa inicial, é realizado o agrupamento dos dados por classe, gerando protótipos para cada uma delas. Como forma de otimizar a obtenção de centros, é proposto um meta-agrupamento, cujo índice de agrupamento tem por base parâmetros de distância entre pontos de uma classe e os protótipos criados. O índice de agrupamento pretende encontrar o melhor conjunto de centros para cada classe, de forma que o novo espaço de atributos seja favorável à aplicação das relações de similaridade.

A aplicação da relação de similaridade sobre um elemento cria uma hiperesfera em torno dele, dando forma ao núcleo e ao suporte de um conjunto difuso. Foram testadas diferentes implementações para o suporte e o núcleo, e as que apresentaram melhores resultados estão detalhadas no trabalho.

Fez-se a opção por um modelo mais simples, baseado na métrica de distância euclidiana, que gera as hiperesferas difusas. O emprego de modelos mais complexos, como a geração de hiperelipses, que ocorre quando se faz uso de outra métrica de distância, por exemplo, a de Mahalanobis, é possível, porém mais custosa.

Para o método AGRUP, na fase de classificação propriamente dita, são derivadas as distribuições de possibilidade, a partir das quais se cria uma família parametrizada de classificadores imprecisos. Estratégias difusas, possibilistas ou probabilistas, que de alguma forma, são capazes de devolver graus de compatibilidade dos elementos para com as classes, podem ser empregadas na abordagem proposta.

A definição de uma classe precisa para um elemento é dada por uma heurística ou critério de classificação. O emprego do refinamento pela vizinhança é interessante quando se define a classificação imprecisa no espaço de atributos e faz-se a vizinhança no espaço do problema.

O classificador proposto trabalha no espaço de atributos multidimensional, diferentemente de trabalhos anteriores, como o de Malinski (2000). A derivação das medidas de possibilidade num espaço tratado conjuntamente resulta em redução na perda de informação, quando se compara às abordagens que tratam o espaço dimensão a dimensão.

Uma das contribuições mais significativas neste trabalho é a investigação da imprecisão proveniente de uma classificação. É proposto um conjunto de métodos próprio para lidar

com a classificação imprecisa, consistindo-se ainda num ferramental flexível no sentido de que outras abordagens encontradas na literatura, sejam elas difusas, possibilistas ou probabilistas, podem ser empregadas, desde que as distribuições fornecidas por elas possam ser transformadas em possibilidades. Os mecanismos de avaliação da classificação imprecisa se mostram aplicáveis a diferentes conjuntos de dados, além de apresentar resultados promissores. No que se refere à combinação de classificadores, é clara a vantagem trazida pela abordagem, quando se agregam distribuições vindas de métodos distintos. A noção de ordem imposta pela distribuição de possibilidade, que caracteriza o classificador impreciso, facilita a agregação das distribuições com significados diferentes.

Este trabalho tem, pois, também como contribuição a proposta de um modelo que emprega relações de similaridade. O uso destas relações difusas disponibiliza um método mais flexível quanto à imprecisão, uma vez que seus parâmetros ditam a sobreposição entre as classes, definindo os graus de pertinência associados a cada elemento. O ajuste dos parâmetros que determinam núcleo e suporte das funções depende do espaço de atributos e uma relação de similaridade é específica para cada protótipo.

O trabalho pode ser continuado em várias frentes. Do mesmo modo que se obtém classificadores imprecisos usando um corte (o parâmetro α) para definir as classes compostas, pode-se parametrizar a classificação imprecisa usando-se o número de graus maximais distintos a serem considerados para criar uma classe composta para um dado. Fixar um número de graus maximais não é o mesmo que fixar um número de classes; um único valor maximal pode levar a uma classificação composta. Para um problema com $|C| = c$ classes, o número máximo de graus maximais é, obviamente c também, já que é possível ter uma distribuição onde todos os graus são distintos. Neste caso, analogamente à tabela de confusão imprecisa definida para cada valor de α da família parametrizada de classificadores imprecisos, a tabela de confusão seria definida para cada número de graus maximais. A partir destas tabelas de confusão seriam gerados os gráficos com as curvas de erro, precisão, imprecisão parcial e completa. A diferença em relação à abordagem com cortes α é que o gráfico de classificação seria mais simples: ao invés de variar α num dos eixos, a variação seria de até $|C|$ valores. As curvas essenciais indicariam o erro, acerto preciso, imprecisão completa e imprecisão parcial, mas pode-se definir curvas por cardinalidade como foi feito no presente trabalho.

Um outro aspecto interessante é o emprego de variáveis como custo, que também poderiam ser levadas em conta na escolha do classificador impreciso a partir de um classificador possibilista. Por exemplo, em uma dada aplicação, o custo de uma escolha

precisa errada (e tomada como certa) pode ser eventualmente maior que o custo de investigação extra para decidir no caso de uma classificação imprecisa. Por outro lado, o ganho no acerto pode ser muito maior que o custo da investigação extra, por exemplo, em função do tempo despendido. Dados como custos e probabilidade, que possibilitem estimar a esperança de ganho, podem ser incorporados na metodologia apresentada, o que interessa em aplicações como a de registros de falhas em subestações elétricas, apresentada na Seção 6.2.5.

E, por fim, dada a aplicação a imagens de RM, que foi restrita à classificação bidimensional de um corte do volume do cérebro, destaca-se a potencialidade de uso do método AGRUP na classificação tridimensional, pela aplicação do método corte a corte. O mecanismo impreciso pode também ser empregado e a vizinhança, tida como heurística de refinamento, também pode ser estendida ao volume tridimensional.

REFERÊNCIAS BIBLIOGRÁFICAS

BALL, G. H.; HALL, D. J. Isodata, an iterative method of multivariate analysis and pattern classification. **The Journal of Applied Behavioral Science**, NTL Institute, v. 12, n. 2, p. 153–155, March 1967. 48

BELACEL, N.; HANSEN, P.; MLADENOVIC, N. Fuzzy j-means: a new heuristic for fuzzy clustering. **Pattern Recognition**, Elsevier Science Ltd., v. 35, n. 10, p. 2193–2200, October 2002. 52

BELLMAN, R.; KALABA, R.; ZADEH, L. Abstraction and pattern classification. **Journal of Mathematical Analysis and Applications**, Elsevier, Santa Monica, California, U.S.A., v. 13, n. 1, p. 1–7, January 1966. 28, 44

BEZDEK, J.; EHRLICH, R.; FULL, W. Fcm: The fuzzy c-means algorithm. **Computers & Geosciences**, Elsevier Science Ltd., v. 10, n. 2-3, p. 191–203, May 1984. 50

BEZDEK, J. C. **Pattern recognition with fuzzy objective function algorithms**. New York: Plenum Press, 1981. 256 p. ISBN 9780306406713. 51

BEZDEK, J. C.; CORAY, C.; GUNDERSON, R.; WATSON, J. Detection and characterization of cluster substructure. ii. fuzzy c-varieties and convex combinations thereof. **SIAM Journal on Applied Mathematics**, JSTOR, v. 40, n. 2, p. 358–372, April 1981. 51

BEZDEK, J. C.; PAL, S. K. **Fuzzy models for pattern recognition**. New York: IEEE Press, 1992. 539 p. ISBN 9780780304222. 27

BISHOP, C. M. **Neural networks for pattern recognition**. New York: OXFORD University Press, 1995. 482 p. ISBN 019853864. 61

BISHOP, Y. M.; FEINBERG, S. E.; HOLLAND, P. W. **Discrete multivariate analysis: theory and practice**. Cambridge, Mass.: MIT Press, 1975. 557 p. ISBN 9780262021135. 66

BRAGA, A. P.; LUDERMIR, T. B.; CARVALHO, A. C. P. de L. F. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: Livros Técnicos e Científicos Editora S.A., 2000. 262 p. 60, 61

BRAINWEB. **Simulated brain database (BrainWeb)**. Montréal: McConnell Brain Imaging Centre - Montréal Neurological Institute, McGill University, 2006.

Disponível em: <<http://www.bic.mni.mcgill.ca/brainweb/>>. Acesso em: Junho 2006. 30, 100

CARPENTER, G. A.; GROSSBERG, S. Art 2: Self-organization of stable category recognition codes for analog input patterns. **Applied Optics: Special Issue on Neural Network.**, OSA, v. 26, n. 23, p. 4919–4930, December 1987. 62

_____. A massively parallel architecture for a self-organizing neural pattern recognition machine. **Computer Vision, Graphics, and Image Processing.**, Academic Press Professional, Inc., San Diego, CA, USA, v. 37, n. 1, p. 54–115, January 1987. ISSN 0734-189X. 62

_____. Art 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. **Neural Networks.**, Elsevier Science Ltd., v. 3, n. 2, p. 129–152, August 1990. 62

CASTRO, A. P. A. de; DRUMMOND, I. N.; SILVA, J. D. S. da. A multiscale neural network method for image restoration. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 30, 2007, Florianópolis, Brasil. **Anais...** Florianópolis: SBMAC, 2007. 118

CBERS. **Satélite sino-brasileiro de recursos terrestres - CBERS.** São José dos Campos: INPE - Instituto Nacional de Pesquisas Espaciais, 2004. Disponível em: <<http://www.cbears.inpe.br>>. Acesso em: Junho 2007. 30, 99

CLARKE, R. T.; BITTENCOURT, H. R. Uso de árvores de decisão na classificação de imagens digitais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 11, 2003, Belo Horizonte, Brasil. **Anais...** Belo Horizonte: Digital, 2003. v. 1, p. 2043–2045. 60

COLE, R. M. **Clustering with genetic algorithms.** Dissertação (Mestrado) — University of Western Australia, Nedlands 6907, Australia, January 1998. Disponível em: <<http://citeseer.ist.psu.edu/435115.html>>. 47

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, IEEE, v. 13, n. 1, p. 21–27, January 1967. 57

DASARATHY, B. V. **Nearest neighbor (NN) norms:** Nn pattern classification techniques. Washington: IEEE Computer Society Press Tutorial, 1991. 447 p. ISBN 9780818689307. 57

DIDAY, E. The dynamic cluster method in non-hierarchical clustering. **Journal of Computing and Information Science**, Springer, v. 2, n. 1, p. 61–88, March 1973. 48

DRUMMOND, I.; MELENDEZ, J.; SANDRI, S. Assessing the aggregation of parameterized imprecise classification. **Frontiers in Artificial Intelligence and Applications**, IOS Press, v. 146, n. 1, p. 227–235, October 2006. 136

DRUMMOND, I.; SANDRI, S. A clustering-based possibilistic method for image classification. **Lecture Notes in Computer Science**, Springer, v. 3171, n. 1, p. 454–463, October 2004. 70, 126, 128

_____. A clustering-based fuzzy classifier. **Frontiers in Artificial Intelligence and Applications**, IOS Press, v. 131, n. 1, p. 247–254, October 2005. 70, 72

_____. Parameterized imprecise classification: elicitation and assessment. **Lecture Notes in Computer Science**, Springer, v. 4140, n. 1, p. 319–328, October 2006. 29, 82

DUBOIS, D.; ESTEVA, F.; GARCIA, P.; GODO, L.; MANTARAS, R. L.; PRADE, H. Fuzzy set modelling in case-based reasoning. **International Journal of Intelligent Systems**, John Wiley & Sons, Ltda., v. 13, n. 4, p. 345–373, April 1998. 38, 39

DUBOIS, D.; PRADE, H. **Possibility theory**: an approach to computerized processing of uncertainty. New York: Plenum Publishing Corporation, 1988. 263 p. ISBN 0306425203. 37, 41

DUBOIS, D.; PRADE, H.; SANDRI, S. On possibility/probability transformations. In: **FUZZY LOGIC: STATE OF ART**. Boston: Kluwer Academic Publishers, 1993. cap. 12, p. 103–112. ISBN 9780792323242. 41, 42

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. New York: John Wiley & Sons, Inc., 2000. 654 p. ISBN 9780471056690. 27, 28, 43, 55, 60, 61

EFRON, B.; TIBSHIRANI, R. **An introduction to the bootstrap**. New York: Chapman & Hall, 1993. 436 p. ISBN 9780412042317. 68

FONSECA, F. R.; FREITAS, C.; DUTRA, L. V.; MARTINS, F. T.; GUIMARÃES, R. J. P. S.; SCHOLTE, R. G. C.; AMARAL, R. S.; DRUMMOND, S. C.; M., M. A. C.; ROCHA, L.; CARVALHO, O. S. Desenvolvimento de um modelo de regressão linear para a predição da prevalência de esquistossomose no estado de minas gerais.

- In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13, 2007, Florianópolis, Brasil. **Anais...** Florianópolis: URLibService, 2007. p. 2573–2580. ISBN 978-85-17-00031-7. 137, 138, 143
- FREITAS, C. C.; GUIMARÃES, R. J. P. S.; DUTRA, L. V.; SANTOS, R. A. T.; MARTINS, F. T.; GOUVÊA, E. J. C.; MOURA, A. C.; DRUMMOND, S. C.; AMARAL, R. S.; CARVALHO, O. S. Remote sensing and geographic information systems for the study of schistosomiasis in the state of minas gerais, brazil. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 26, 2006, Denver, USA. **Proceedings...** Denver: IEEE, 2006. p. 2841–2848. 137
- GAO, Q.-B.; WANG, Z.-Z. Center-based nearest neighbor classifier. **Pattern Recognition**, Elsevier Science Ltd., v. 40, n. 1, p. 346–349, january 2007. 57
- GEOSCHISTO. **Projeto GeoSchisto**. [online]. 2007. Disponível em: <<<http://www.dpi.inpe.br/geoschisto/>>>. Acesso em: Junho 2007. 30, 137
- GONZALEZ, R. C.; WOODS, R. E. **Digital image processing**. Upper Saddle River, New Jersey: Prentice-Hall, Inc., 2000. 505 p. ISBN 0-201-18075-8. 94, 95, 118
- GUIMARÃES, R. J. P. S.; FREITAS, C. C.; DUTRA, L. V.; MOURA, A. C. M.; AMARAL, R. S.; DRUMMOND, S. C.; GUERRA, M.; SCHOLTE, R. G. C.; FREITAS, C. R.; CARVALHO, O. S. Analysis and estimative of schistosomiasis prevalence for minas gerais state, brazil, using multiple regression with social and environmental spatial data. **Memórias do Instituto Oswaldo Cruz**, Scielo, v. 101, p. 91–96, October 2006. 137
- GUSTAFSON, D. E.; KESSEL, W. C. Fuzzy clustering with a fuzzy covariance matrix. In: IEEE CONFERENCE ON DECISION AND CONTROL, SYMPOSIUM ON ADAPTIVE PROCESSES, 17, 1978, San Diego, Canadá. **Proceedings...** San Diego: IEEE Press, 1978. v. 17, p. 761–766. 50
- HAMERLY, G.; ELKAN, C. Alternative to the k-means algorithm that find better clusterings. In: INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 11, 2002, McLean, Virginia (USA). **Proceedings...** McLean: ACM Press, 2002. v. 3, p. 600–607. ISBN 1-58113-492-4. 48
- HAYKIN, S. **Redes neurais: princípios e prática**. Porto Alegre: Bookman, 2001. 898 p. ISBN 8573077182. 60, 62
- IBGE. **Instituto Brasileiro de Geografia e Estatística (IBGE)**. Minas Gerais, Brasil, 2007. Disponível em:

<<http://www.ibge.gov.br/estadosat/perfil.php?sigla=mg>>. Acesso em: Junho 2007. 137

JAIN, A. K. **Fundamentals of digital image processing**. New Jersey: Prentice hall, Inc, 1989. 569 p. ISBN 9780133361650. 96

JAIN, A. K.; DUIN, R. P. H.; MAO, J. Statistical pattern recognition: A review. **IEEE Transactions on Pattern Analysis and Machine Learning**, v. 22, n. 1, p. 4–37, 2000. 28

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: A review. **ACM Computing Surveys**, ACM Press, v. 31, p. 264–323, September 1999. 45, 46, 47, 48, 62

KAYMAK, U.; SETNES, M. **Extended fuzzy clustering algorithms**. Rochester, NY, USA, November 2000. Erasmus University Rotterdam Technical Report. 51, 52

KELLER, J.; HUNT, D. Incorporating fuzzy membership functions into the perceptron algorithm. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, n. 6, p. 693–699, July 1985. 59

KELLER, J. M.; GRAY, M. R.; GIVENS, J. J. A. A fuzzy k-nearest neighbor algorithm. **IEEE Transactions on Systems, Man and Cybernetics**, IEEE Press, v. 15, n. 4, p. 580–585, July/August 1985. 29, 57, 59, 69, 78

KITTLER, J.; HATEF, M.; DUIN, R.; MATAS, J. On combining classifiers. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 20, n. 3, p. 226–239, March 1998. 62

KLIR, G. J.; FOLGER, T. A. **Fuzzy sets, uncertainty and information**. New Jersey: Prentice Hall, 1988. 368 p. ISBN 0133459845. 36

KOHONEN, T. **Self-organizing maps**. New York: Springer, 2001. 501 p. ISBN 9783540679219. 62

KRISHNAPURAM, R.; KELLER, J. A possibilistic approach to clustering. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 1, n. 2, p. 98–110, May 1993. 53, 54, 78

KUNCHEVA, L.; BEZDEK, J. Nearest prototype classifier designs: An experimental study. **International Journal of Intelligent Systems**, Wiley, v. 16, n. 12, p. 1445–1473, December 2001. 59

KUNCHEVA, L.; WHITAKER, C. Pattern recognition. **Encyclopedia of Statistics in Behavioral Science**, John Wiley & Sons, Ltd., v. 3, p. 1532–1535, October 2005. 43

KUNCHEVA, L. I. **Fuzzy classifier design**. New York: A Springer-Verlag Company, 2000. 314 p. ISBN 9783790812985. 17, 28, 29, 44, 62, 64, 70, 84

_____. **Combining pattern classifiers: methods and algorithms**. Hoboken, NJ: John Wiley & Sons, Ltd., 2004. 350 p. ISBN 9780471210788. 43, 62, 63, 64, 66, 67, 82, 88

LAM, L.; SUEN, C. Y. Optimal combination of pattern classifiers. **Pattern Recognition Letters**, Elsevier Science Ltd., v. 16, n. 9, p. 945–954, September 1995. 62

LANDSAT. **Land Remote Sensing Satellite - LANDSAT**. Washington, DC: NASA - National Aeronautics and Space Administration, 2004. Disponível em: <<http://landsat.gsfc.nasa.gov/images/lambertian.html>>. Acesso em: Junho 2007. 30, 99

LEONDES, T. C. **Image processing and pattern recognition**. California: Academic Press, 1998. 386 p. ISBN 9780124438651. 28, 43, 44

LESKI, J. Towards a robust fuzzy clustering. **Fuzzy Sets and Systems**, Elsevier Science Ltd., v. 137, n. 2, p. 215–233, July 2003. 52

LIKAS, A.; VLASSIS, N.; VERBEEK, J. The global k-means clustering algorithm. **Pattern Recognition**, Elsevier Science Ltd., v. 76, n. 2, p. 151–461, February 2003. 48

LLANOS, D.; MELÉNDEZ, J.; COLOMER, J.; CORBELLA, X.; SÁNCHEZ, J. Abstraction of significant temporal features of voltage sags recorded in a 25kv substation. In: INTERNATIONAL CONFERENCE ON RENEWABLE ENERGIES AND POWER QUALITY, 2003, Vigo, ES. **Proceedings...** Vigo: Digital, 2003. Online. Disponível em: <<http://www.icrepq.com/index2.htm>>. 30, 131

LLANOS, D.; MORA, J.; MELÉNDEZ, J.; RUIZ, M.; COLOMER, J.; SÁNCHEZ, J.; CORBELLA, X. Classification of short duration faults (voltage sags) in transmission and distribution power systems. In: EUROPEAN CONTROL CONFERENCE, 7, 2003, Cambridge, UK. **Proceedings...** Cambridge: University of Cambridge, 2003. Online. Disponível em: <<http://exit.udg.es/articles/90.pdf>>. 30, 131

MALINSKI, P. R. **Uso da teoria de possibilidades e de algoritmos genéticos para construção de um classificador supervisionado com aplicação em imagens de radar de abertura sintética (SAR)**. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, Setembro 2000. 29, 54, 69, 89, 91, 146

MALINSKI, P. R.; SANDRI, S.; FREITAS, C. An imprecion-based image classifier. In: INTERNATIONAL IEEE CONFERENCE ON FUZZY SYSTEMS, 10, 2001, Melbourne, Australia. **Proceedings...** Melbourne: IEEE Press, 2001. v. 1, p. 825–828. 55

MARQUES, P. F. L. Os componentes básicos de um sistema mri. In: WORKSHOP ON BIOMEDICAL ENGINEERING, 3, 1998, Lisboa, Portugal. **Proceedings...** Lisboa: DEEC, 1998. p. 10–14. 96

MARTINS, F. T.; FREITAS, C.; DUTRA, L. V.; FONSECA, F. R.; GUIMARÃES, R. J. P. S.; SCHOLTE, R. G. C.; AMARAL, R. S.; DRUMMOND, S. C.; FREITAS, C. R.; CARVALHO, O. S. Uso de árvore de decisão para predição da prevalência de esquistossomose no estado de minas gerais, brasil. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13, 2007, Florianópolis, Brasil. **Anais...** Florianópolis: URLibService, 2007. p. 2841–2848. ISBN 978-85-17-00031-7. 137, 138, 143

MCQUENN, J. Some methods for classification and analysis of multivariate observations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 2, 1967, Berkeley. **Proceedings...** Berkeley: University of California Press, 1967. v. 1, p. 281–257. 48

MICHIE, D.; SPIEGELHALTER, D.; TAYLOR, C. **Machine learning, neural and statistical classification**. New Jersey: Englewood Cliffs - Prentice Hall, 1994. 298 p. ISBN 9780131063600. 67

NADLER, M.; SMITH, E. P. **Pattern recognition engeneering**. New York: John Wiley & Sons Inc, 1993. 588 p. ISBN 0471622931. 27

PAWLAK, Z. Rough sets. **International Journal of Parallel Programming**, Springer, v. 11, n. 5, p. 341–356, October 1982. 81

PEDRYCZ, W. Collaborative fuzzy clustering. **Pattern Recognition Letters**, Elsevier Science Ltd., v. 23, n. 14, p. 1675–1686, December 2002. 52

- PLEET, B.; HICKS, R. **Computer vision laboratory**. Nov 2004.
[Http://vis-www.cs.umass.edu/projects/mristroke/mri.html](http://vis-www.cs.umass.edu/projects/mristroke/mri.html). 98
- RICHARDS, J. A. **Remote sensing digital image analysis: an introduction**.
 Berlin: Springer-Verlag, 1986. 281 p. ISBN 9780387160078. 56
- ROGOVA, G. Combining the results of several neural networks classifiers. **Neural Networks**, Elsevier Science Ltd., Oxford, UK, UK, v. 7, n. 5, p. 777–781, May 1994. ISSN 0893-6080. 62
- RUSPINI, E. A new approach to clustering. **Information and Control**, v. 15, n. 1, p. 22–32, 1969. 48, 49
- RUSPINI, E.; BONISSONE, P.; PEDRYCZ, W. **Handbook of fuzzy computation**. London: IOP Publishing Ltd, 1998. 504 p. ISBN 9780750304276. 38
- SHAPIRO, L. G.; STOCKMAN, G. C. **Computer vision**. Upper Saddle River, NJ: Prentice Hall, 2001. 580 p. ISBN 9780130307965. 56, 96
- SILVA, J. D. S. d. **Uso de redes neurais em visão computacional e processamento de imagens**. São José dos Campos, 2004. 70 p. Relatório de atividades em estágio SPE/RHAE (CNPq-Processo no. 260.048/95–0), realizado no Nuclear Engineering Department da University of Tennessee, Knoxville, USA, sob supervisão do Dr. Robert E. Uhrig, no período de 01/03/1996 a 28/02/1997.
 Disponível em: <<http://mtc-m16.sid.inpe.br/rep-/sid.inpe.br/marciana/2004/09.17.13.47>>. 21, 120, 124, 125
- SILVA, L. R. S. **Aprendizagem participativa em agrupamento nebuloso de dados**. Dissertação (Mestrado) — Universidade Estadual de Campinas, Campinas, Março 2003. 45, 52
- SMETS, P. What's dempster-shafer model? In: YAGER, R. R.; FEDRIZZI, M.; KACPRZYK, J. (Ed.). **Advances in the Dempster-Shafer theory of evidence**. New York: Wiley, 1994. cap. 1, p. 5–34. 81
- SYMON, M. J. Clustering criterion and multivariate normal mixture. **Biometrics**, JSTOR, v. 37, n. 1, p. 35–43, March 1981. 48
- WU, K.; YANG, M. Alternative c-means clustering algorithms. **Pattern Recognition**, Elsevier Science Ltd., v. 35, n. 2, p. 2267–2278, February 2002. 52

- YAGER, R. R. Families of owa operators. **Fuzzy Sets and Systems**, Elsevier Science Ltd., v. 59, n. 2, p. 125–148, October 1993. 37
- YAGER, R. R.; OVCHINIKOV, S.; TONG, R.; T., N. H. **Fuzzy sets and applications**: selected papers by I. A. Zadeh. New York: John Wiley & Sons, Ltd., 1987. 684 p. ISBN 9780471857105. 39
- YANG, T.; WANG, S. Competitive algorithms for the clustering of noisy data. **Fuzzy Sets and Systems**, Elsevier Science Ltd., v. 141, n. 2, p. 281–299, January 2004. 53
- ZADEH, L. Similarity relations and fuzzy orderings. **Information Sciences**, Elsevier Science Ltd., v. 3, n. 2, p. 177–200, April 1971. 38
- ZADEH, L. A. Fuzzy sets. **Fuzzy Sets, Information and Control**, v. 8, n. 3, p. 338–353, June 1965. 28, 33
- _____. Fuzzy sets as a basis for theory of possibility. **Fuzzy Sets and Systems**, Elsevier Science Ltd., v. 1, n. 1, p. 3–28, January 1978. 42
- ZHANG, B.; SRIHARI, S. N. Fast k-nearest neighbor classification using cluster-based trees. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 26, n. 4, p. 525–528, 2004. 57
- ZHANG, G. P. Neural networks for classification: A survey. **IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews**, IEEE, v. 30, n. 4, p. 451–462, November 2000. 61

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programa de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. São aceitos tanto programas fonte quanto executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.