

**ANIBAL COTRINA - ATENCIO**

**AVALIAÇÃO DE MODELO DE MISTURAS APLICADO À  
CLASSIFICAÇÃO DE IMPRESSÕES DIGITAIS SEGUNDO A  
ARQUITETURA PCASYS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Evandro Ottoni Teatini Salles.

VITÓRIA  
2010

Dados Internacionais de Catalogação-na-publicação (CIP)  
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

---

A864a Atencio, Anibal Cotrina, 1978-  
Avaliação de modelo de misturas aplicado à classificação de impressões digitais segundo a arquitetura pcasys / Anibal Cotrina Atencio. – 2010.  
96 f. : il.

Orientador: Evandro Ottoni Teatini Salles.  
Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Impressão digital (Computação). 2. Classificação.  
3. Biometria. 4. Distribuição (Teoria da probabilidade).  
5. Algoritmos de expectativa de maximização. 6. Análise multivariada. 7. PCASYS (Sistema classificador de impressões digitais). I. Salles, Evandro Ottoni Teatini. II. Universidade Federal do Espírito Santo. Centro Tecnológico. III. Título.

CDU: 621.3

---

**ANIBAL COTRINA-ATENCIO**

**AVALIAÇÃO DE MODELO DE MISTURAS APLICADO À  
CLASSIFICAÇÃO DE IMPRESSÕES DIGITAIS SEGUNDO A  
ARQUITETURA PCASYS**

Dissertação submetida ao programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para a obtenção do Grau de Mestre em Engenharia Elétrica.

Aprovada em 29 de novembro de 2010.

**COMISSÃO EXAMINADORA**

---

Prof. Dr. Evandro Ottoni Teatini Salles  
Universidade Federal do Espírito Santo  
Orientador

---

Prof. Dr. João Marques Salomão  
Instituto Federal do Espírito Santo - Campus Vitória

---

Prof. Dr. Jugurta Rosa Montalvão Filho  
Universidade Federal de Sergipe

*à racionalidade do papai,*

*à fé da mamãe, e*

*à sensibilidade da Jane.*

## Agradecimentos

Eu fico muito grato a:

a jane,  
o luis, a lara,  
a sandra, o silvio,  
a ebel, o evandro, o ever,  
o wesley, a whitney, o willem  
a dalva, o david, o diego, o diogo,  
o heber, o henery, o henry, o hudson,  
o glauber, o gregorio, a gabriela, a geisa,  
o patric, o patrick, o paulo, o pedro, o peter,  
o felipe, o fabio, a fernanda, o flávio, o francis,  
o robson, o rodrigo, o romário, a rowena, o renato,  
a isabel, a izabel, a karin, o klauss, a talita, o thomas,  
o adilson, a alessandra, o alex, o alexandre, a anna, o andré,  
a carli, o carlos, o cássio, o César, o claudio, o cristiano, o cristofer,  
a marcia, o marcus, o mario, a mirian, a mayara, o moacir, a marcela,  
a janayna, o jardel, a jessica, o João, o jorge, a joseane, a joseline, a juliana, o julio,  
o brasil, o espirito santo, a cidade de vitória, a ufes, o ppgee, o laboratório cisne e ao cnpq.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Caracterização do problema . . . . .	2
1.2	Proposta . . . . .	3
1.3	Organização da dissertação . . . . .	4
<b>2</b>	<b>Definições e Estado da Arte</b>	<b>5</b>
2.1	Reconhecimento Biométrico . . . . .	5
2.2	Sistema biométrico . . . . .	5
2.3	Impressões digitais . . . . .	6
2.4	Classificação de Impressões Digitais . . . . .	8
2.5	Extração de características . . . . .	11
2.6	Técnicas de Classificação . . . . .	12
2.7	Arquitetura PCASYS . . . . .	15
2.8	Banco de dados DB4 . . . . .	16
2.9	Resultados atuais . . . . .	16
<b>3</b>	<b>Trabalho Prévio</b>	<b>18</b>
3.1	Processamento da Impressão Digital . . . . .	18
3.1.1	Determinação de Regiões na Imagem . . . . .	19
3.1.2	Melhora da Imagem . . . . .	19

3.1.3	Determinação da Imagem de Orientação . . . . .	20
3.1.4	Determinação do Ponto de Referência . . . . .	21
3.2	Classificação . . . . .	22
3.2.1	Geração de características . . . . .	22
3.2.2	Redução dimensional . . . . .	22
3.2.3	Classificação . . . . .	23
<b>4</b>	<b>Fundamentação Teórica</b>	<b>24</b>
4.1	Contexto . . . . .	24
4.2	Modelos de Misturas . . . . .	26
4.3	Verossimilhança máxima . . . . .	27
4.4	Algoritmo EM . . . . .	28
4.5	<i>Mean-Shift</i> . . . . .	30
4.6	Algoritmo SMEM . . . . .	31
4.7	GMBC . . . . .	35
4.8	Divergência Kullback-Leibler . . . . .	36
4.9	Região de Confiança . . . . .	37
<b>5</b>	<b>Avaliação dos parâmetros estimados</b>	<b>39</b>
5.1	Avaliação da média via teste de hipótese . . . . .	40
5.2	Avaliação da entropia mutua . . . . .	45
5.3	Resumo . . . . .	47
<b>6</b>	<b>Classificação Não-Supervisionada usando GMM e EM</b>	<b>48</b>
6.1	Metodologia usada da classificação Não-Supervisionada . . . . .	49
6.2	Resultados da classificação Não-Supervisionada inicializando o EM com parâmetros amostrais . . . . .	52

6.3	Resultados da classificação Não-Supervisionada inicializando o EM com centroides estimados via <i>K-means</i> . . . . .	62
6.4	Resultados da classificação Não-Supervisionada inicializando o EM com modas estimados via <i>Mean-Shift</i> . . . . .	65
6.5	Resumo . . . . .	69
<b>7</b>	<b>Classificação Supervisionada usando GMBC e SMEM</b>	<b>70</b>
7.1	Metodologia da classificação Supervisionada . . . . .	71
7.2	Resultados da classificação Supervisionada . . . . .	73
7.3	Resumo . . . . .	78
<b>8</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>81</b>
8.1	Conclusões . . . . .	83
8.2	Trabalhos futuros . . . . .	86
<b>A</b>	<b>Interface de usuário</b>	<b>88</b>
<b>B</b>	<b>Resultados Adicionais</b>	<b>91</b>
B.1	Valores de media e covariância . . . . .	91
B.2	Elipses de confiança . . . . .	92
B.3	Divergência de Kullback-Leibler . . . . .	93



# Lista de Tabelas

2.1	Trabalhos de classificação de impressões digitais, técnicas de extração de características (Orientação (O), Singularidades (S), fluxo de cristas (R) e Gabor (G)), técnicas de classificação utilizadas (regras (Rb), sintática (Sy), estruturais (Str) classificadores por redes neurais (Nn), múltiplos classificadores (Mc)) e ano de publicação. Informação extraída de (Maltoni et al., 2003) e (Samatelo, 2007) . . . . .	14
2.2	Taxas de erro de classificação sobre a base de dados DB4 considerando as distribuições natural (N) e balanceada (B). Dados baseados em informação fornecida em (Maltoni et al., 2003), (Maltoni et al., 2009), (Samatelo, 2007), (Liu et al., 2007). . . . .	17
3.1	Menores taxas de erro de classificação alcançados no trabalho prévio (Samatelo, 2007) com distintos classificadores para os problemas de 5 e 4 classes considerando as distribuições balanceada (B) e natural (N). . . . .	23
6.1	Taxas de erro (%) de classificação Não-Supervisionado para diferentes quantidades de coeficientes PCA no vetor de características quando foram considerados os parâmetros amostrais como parâmetros iniciais do EM para o problema de 4 classes. . . . .	56
6.2	Matriz de confusão da classificação com misturas gaussianas para o problema de quatro classes com: (a) 38 ; e (b) 60 coeficientes PCA. . . . .	57
6.3	Taxas de erro (%) de classificação Não-Supervisionada para diferentes quantidades de coeficientes PCA no vetor de características quando foram considerados os parâmetros amostrais como parâmetros iniciais do EM para o problema de 5 classes. . . . .	61
6.4	Matriz de confusão da classificação com misturas gaussianas para o problema de quatro classes com: (a) 38 ; e (b) 64 coeficientes PCA. . . . .	62

6.5	Taxas de erro (%) de classificação Não-Supervisionado para diferentes quantidades de coeficientes PCA no vetor de características quando foram usados os centroides como parâmetros iniciais do EM para os problemas de 4 e 5 classes. . . . .	64
6.6	Médias e desvios padrão das taxas de erro (%) de classificação Não-Supervisionado para diferentes quantidades de coeficientes PCA no vetor de características quando foram considerados as modas achadas via <i>Mean-Shift</i> como parâmetros iniciais do EM. . . . .	68
6.7	Comparação dos resultados obtidos com a classificação Não-Supervisionada baseada em GMM com os resultados obtidos no trabalho prévio. . . . .	69
7.1	Médias e desvios padrão das taxas de erro (%) de classificação Supervisionado para diferentes quantidades de coeficientes PCA no vetor de características . . . . .	76
7.2	Menores taxas de erro de classificação Supervisionada comparadas com os resultados alcançados no trabalho prévio para os problemas de 4 e 5 classes considerando as distribuições balanceada (B) e natural (N) e a se indicando a quantidade de coeficientes PCA. (Samatelo, 2007) . . . . .	79
8.1	Comparação com resultados atuais, das Taxas de erro de classificação alcançados neste trabalho. Informação baseada em (Maltoni et al., 2003), (Maltoni et al., 2009), (Samatelo, 2007), (Liu et al., 2007). . . . .	82
B.1	Parâmetros estimados da função de distribuição de probabilidades da mistura gaussiana para o problema de (a) quatro classes e (b) cinco classes. . . . .	91

# Lista de Figuras

2.1	Cristas e vales numa impressão digital. (Maltoni et al., 2003) . . . . .	6
2.2	Padrões de impressões digitais distintas no nível global. Os quadrados e triângulos denotam os pontos singulares <i>core</i> e <i>delta</i> respectivamente. (Maltoni et al., 2003) . . . . .	7
2.3	Minúcias (círculos pretos preenchidos) em uma porção da imagem de impressão digital; poros de suor (círculos vazios) em uma linha de crista. (Maltoni et al., 2003) . . . . .	7
2.4	Impressões digitais do tipo <i>Arch</i> , <i>Tented Arch</i> , <i>Left Loop</i> , <i>Right Loop</i> e <i>Whorl</i> . São apresentadas impressões de duas sub-classes da classe <i>Whorl</i> . (Maltoni et al., 2003) . . . . .	8
2.5	Três impressões de classes diferentes com pequena variação intra-classe (fila superior) e três impressões da mesma classes com grande variação entra-classe (fila inferior). . . . .	9
2.6	três exemplos de impressões digitais que contem ruído na àrea padrão (Maltoni et al., 2003). . . . .	10
2.7	Características mais frequentemente utilizados na classificação de impressões digitais: (a) imagem de orientação, (b) linhas de fluxo de crista, (c) pontos singulares (d) resposta de filtro de Gabor. (Samatelo, 2007) . . . . .	11
2.8	Esquema funcional do PCASYS (Candela et al., 1995) . . . . .	15
3.1	Resultados da segmentação. (Samatelo, 2007). . . . .	19
3.2	Imagens melhoradas; (a) Normalização. (b) Melhora no domínio da frequência (Samatelo, 2007). . . . .	20
3.3	(a) Impressão digital de boa qualidade. (b) Imagem de orientação com técnica padrão, (c-d) com técnicas melhoradas (Jiang, 2005), (Wang et al., 2006). . . . .	20

3.4	Sequência de passos no cálculo do ponto de referência. (Samatelo, 2007), (Samatelo e Salles, 2009) . . . . .	21
3.5	Sequência de obtenção do vetor de dados a partir da imagem de orientação. (Samatelo, 2007) . . . . .	22
4.1	Resumo esquematizado dos algoritmos usados das neste trabalho. . . . .	25
4.2	(a) Modelo de misturas gaussianas. (b) Curvas de nível. . . . .	26
4.3	Convergência do algoritmo <i>Expectation-Maximization</i> (Chuong e Batzoglou, 2008). . . . .	30
4.4	Busca de moda com <i>Mean-Shift</i> (Derpanis, 2005). . . . .	31
4.5	Inicialização para o caso de distribuição gaussiana bidimensional (a) e (b) Passo <i>Split</i> . (c) e (d) Passo <i>Merge</i> . . . . .	33
4.6	Estrutura do classificador GMBC com OCON. (Kim e Kang, 2007) . . . . .	35
4.7	Elipse de confiança para um nível de significância de 95% . . . . .	38
5.1	Distribuição dos vetores de características, médias amostrais, centroide estimado com <i>K-means</i> , média estimada com EM, moda estimadas com <i>Mean-Shift</i> e elipses de confiança $F$ e $\chi^2$ correspondentes a cada classe para o problema de 4 classes. . . . .	40
5.2	(a) Projeção bidimensional da distribuição de vetores de características da classe LL para o problema de 4 classes. (b) Elipses de confiança da média amostral ( $F$ e $\chi^2$ ), moda estimada com <i>Mean-Shift</i> e média estimada com EM para $d = \{3, 6, 10, 38, 61, 64\}$ . . . . .	41
5.3	Projeção no espaço bidimensional da distribuição de vetores de características, elipses de confiança, da média amostral ( $F$ e $\chi^2$ ), da moda estimada com <i>Mean-Shift</i> e da média estimada com EM para $d = \{3, 6, 10, 38, 61, 64\}$ para o problema de 4 classes. (a) e (b) Classe <b>W</b> . (c) e (d) Classe <b>RL</b> . (e) e (f) Classe <b>TA-A</b> . . . . .	43
5.4	Distribuição dos vetores de características, médias amostrais, centroide estimada com <i>K-means</i> , média estimada com EM, moda estimadas com <i>Mean-Shift</i> e elipses de confiança $F$ e $\chi^2$ correspondentes a cada classe para o problema de 5 classes. . . . .	44

5.5	Projeção no espaço bidimensional da distribuição de vetores de características, elipses de confiança, da média amostral ( $F$ e $\chi^2$ ), da moda estimada com <i>Mean-Shift</i> e da média estimada com EM para $d = \{3, 6, 10, 38, 61, 64\}$ e para o problema de 5 classes. (a) Classe <b>TA</b> . (b) Classe <b>A</b> . . . . .	45
5.6	Divergência de Jeffreys-Kullback-Leibler (a) Problema de 4 classes. (b) Problema de 5 classes. . . . .	46
6.1	Vetores de características com dois coeficientes PCA das imagens de treinamento rotuladas. (a) Problema de 4 classes: LL (vermelho), W (verde), RL (ciano) e TA-A (azul), (b) Problema de 5 classes: LL, W, RL, TA (amarelo) e A (azul). . . . .	51
6.2	Histograma dos vetores de características com dois coeficientes PCA das imagens de treinamento sem rotular (a) 15 intervalos (b) 31 intervalos por dimensão. . . . .	51
6.3	Função de densidade de probabilidade dos modelos de misturas gaussianas estimados para o problema de (a) 4 classes e (b) 5 classes. . . . .	51
6.4	Curvas das taxas de erro de classificação natural ((a), (c), (e), (g) e (i)) e balanceada ((b), (d), (f), (h) e (j)) para o problema de 4 classes quando os parâmetros amostrais foram consideradas como parâmetros iniciais do EM. .	53
6.5	Curvas das taxas de erro de classificação natural ((a), (c), (e), (g), (i) e (k)) e balanceada ((b), (d), (f), (h), (j) e (l)) para o problema de 5 classes quando foram consideradas os parâmetros amostrais como parâmetros iniciais do EM.	59
6.6	Curvas das taxas de erro de classificação natural (a) e balanceada (b) para o problema de 4 classes quando foram consideradas os centroides estimados com <i>K-means</i> como parâmetros iniciais do EM. . . . .	63
6.7	Curvas das taxas de erro de classificação natural (a) e balanceada (b) para o problema de 5 classes quando foram consideradas os centroides estimados com <i>K-means</i> como parâmetros iniciais do EM. . . . .	63
6.8	Taxas de erro modeladas como variáveis aleatórias para o problema de 4 classes quando foram consideradas as modas achadas via <i>Mean-Shift</i> como parâmetros iniciais do EM. (a) e (c) Distribuição natural. (b) e (d) Distribuição balanceada. . . . .	66

6.9	Taxas de erro modeladas como variáveis aleatórias para o problema de 5 classes quando foram consideradas as modas achadas via <i>Mean-Shift</i> como parâmetros iniciais do EM. (a) e (c) Distribuição natural. (b) e (d) Distribuição balanceada. . . . .	67
7.1	Modelos de misturas gaussianas estimadas correspondente a cada classe estimada com SMEM (a) Classe LL, (b) classe W, (c) classe RL (d) classe TA-A. . . . .	72
7.2	Taxas de erro de classificação Supervisionada modeladas como variáveis aleatórias para o problema de 4 classes considerando as distribuições: (a) natural e (b) balanceada. . . . .	73
7.3	Resultados da classificação Supervisionada para o problema de 4 classes considerando a distribuição balanceada avaliadas para $n = \{2, \dots, 64\}$ coeficientes PCA. Coluna da esquerda: médias e desvios padrão da taxas de erro. Coluna da direita: número de componentes gaussianas que modelam cada classe. . . . .	75
7.4	Taxas de erro de classificação Supervisionada modeladas como variáveis aleatórias para o problema de 5 classes considerando as distribuições: (a) natural e (b) balanceada. . . . .	77
7.5	Resultados da classificação Supervisionada para o problema de 5 classes considerando a distribuição balanceada avaliadas para $d = \{2, \dots, 64\}$ coeficientes PCA. Coluna da esquerda: Médias e desvios padrão da taxas de erro. Coluna da direita: Número de componentes gaussianas que modelam cada classe. . . . .	80
A.1	Interface gráfica de usuário desenvolvido em MATLAB. . . . .	89
A.2	Esquema da organização da interface de usuário. . . . .	90
B.1	Elipses de confiança da média amostral de Fisher e $\chi^2$ , moda achada com <i>Mean-Shift</i> e média achada com EM para $d = \{3, 6, 10, 38, 61, 64\}$ . (a) Classe <b>LL</b> . (b) Classe <b>W</b> . (c) Classe <b>RL</b> . . . . .	92
B.2	Divergência de Kullback-Leibler (a) Problema de 4 classes. (b) Problema de 5 classes. . . . .	93
B.3	Divergência de Jefferys-Kullback-Leibler entre distribuições geradas pelos parâmetros amostrais e os parâmetros estimados com EM da classe TA-A para o problema de quatro classes. . . . .	93

# Lista de Siglas

AFIS	<i>Automatic Fingerprint Identification System</i>
PCASYS	<i>Pattern-Level Classification Automation System</i>
PCA	<i>Principal Component Analysis</i>
GMM	Modelo de Misturas Gaussianas
EM	<i>Expectation-Maximization</i>
SMEM	<i>Split and Merge EM</i>
A	<i>Arch</i>
TA	<i>Tented Arch</i>
LL	<i>Left Loop</i>
RL	<i>Right Loop</i>
W	<i>Whorl</i>
NIST	<i>National Institute Standard Technology</i>
DB4	<i>NIST Special Database 4</i>
DB14	<i>NIST Special Database 14</i>
KLD	Divergência de Kullback-Leibler
JKLD	Divergência de Jeffreys Kullback-Leibler
ML	Verosimilhança máxima
GMBC	<i>Gaussian Mixture Based Classification</i>
OCON	<i>One Net One Class</i>
PDBNN	<i>Probabilistic Decision Based Neural Network</i>
PNN	Rede neural probabilística
$k$ -NN	$k$ -vizinhos mais próximos
MLP	Perceptron Multi-Camada

# Resumo

Atualmente, os sistemas de reconhecimento biométrico tem alcançado grande desenvolvimento e vem sendo usados em diversas áreas como a comercial, a financeira e de segurança. No caso específico do reconhecimento de impressões digitais os sistemas automáticos de identificação (AFIS) são capazes de identificar indivíduos localizados em grandes bancos de dados, lidando com dezenas de milhões de amostras, as que originam um custo computacional bastante elevado. Assim, a classificação de impressões digitais é um problema importante de ser tratado uma vez que diminui o tempo de busca por um indivíduo, reduzindo o espaço de busca a um determinado subgrupo.

No presente trabalho busca-se avaliar se, para a arquitetura de classificação de impressões digitais conhecida como PCASYS (*Pattern-Level Classification Automation System*), é possível assumir que as classes podem ser modeladas por funções de densidade de probabilidade gaussiana e que podem alcançar resultados de classificação comparáveis com os obtidos com técnicas mais amadurecidas, como as baseadas em redes neurais. Para isto, empregou-se os modelos de misturas gaussianas (GMM) com abordagens supervisionada e não-supervisionada na estimativa de parâmetros.

Os resultados foram avaliados comparando as taxas de erro de classificação obtidos neste trabalho frente aos resultados alcançados com outras técnicas no estado da arte; assim mesmo, analisou-se estatisticamente os parâmetros dos modelos de misturas gaussianas estimados usando teste de hipótese e entropia mútua. Os testes foram feitas usando o banco de imagens de impressões digitais número 4 do *National Institute of Standard and Technology* (NIST) considerando os problemas de classificação de 4 e 5 classes e as distribuições natural e balanceada.



# Abstract

Currently, biometric recognition systems have achieved great development and has been used in various fields including trade, finance and security. In the specific case of fingerprint recognition, automatic fingerprint identification systems (AFIS) are able to identify individuals located in large databases, dealing with tens of millions of samples, giving rise to a very high computational cost. Thus, the classification of fingerprints is an important problem to be treated as it reduces the search time by an individual, reducing the search space to a particular subgroup.

In this study we assessed if, for the architecture of fingerprint classification known as PCASYS (*Pattern-Level Classification Automation System*), it is possible to assume that classes can be modeled using Gaussian probability density function and can achieve classification results that can be compared with those obtained with more mature techniques such as those based on neural networks. For this, it is used the Gaussian mixture models (GMM), with supervised and unsupervised approaches in parameter estimation task.

It was evaluated the results comparing the rates of misclassification of the proposed technique with the results achieved with other techniques, anyway, we analyzed statistically the estimated parameters of Gaussian mixture models using hypothesis testing and mutual entropy. It was tested using the image bank fingerprints of number 4 of National Institute of Standards and Technology (NIST) considering the 4 class problem and 5 class problem considering the natural distribution and balanced distribution.

# Capítulo 1

## Introdução

A biometria é uma tecnologia que tem sido utilizada enormemente em identificações criminais e segurança em prisões (aplicação forense), assim mesmo, tem o potencial de ser adotada em larga faixa de aplicações civis (segurança bancária, controle de acesso físico, sistemas de identificação nacionais), em especial os sistemas conhecidos como AFIS (*Automatic Fingerprint Identification System*) estão sendo adotados cada vez mais para ambas as aplicações.

A principal vantagem da biometria baseada nas características anatômicas e fisiológicas é que estas não podem ser forjadas nem tampouco esquecidas (como acontece com as senhas), e obriga que a pessoa a ser autenticada esteja fisicamente presente no ponto de autenticação, eliminando o problema de roubo ou transferência de senhas para outras pessoas.

Sistemas biométricos desta monta possuem duas etapas a serem cumpridas para a identificação de indivíduos: Na primeira etapa se realiza o registro do usuário no sistema, onde a impressão digital é capturada, depois transformada em um modelo matemático e finalmente armazenada em um banco de dados. Na segunda etapa se realiza o reconhecimento, a partir do qual o sistema compara a impressão digital apresentada pelo usuário com o modelo previamente armazenado.

Um sistema de identificação pessoal é importante para uma série de transações do dia a dia, como por exemplo no controle de acesso a ambientes ou no uso de computadores. A identificação de uma pessoa requer a comparação das suas impressões digitais com todas as impressões digitais existentes em um banco de dados que, em muitos casos, pode ser muito grande (vários milhões de impressões digitais). Uma estratégia comum para reduzir o número de comparações de impressões digitais durante a recuperação e, conseqüentemente, melhorar o tempo de resposta do processo de identificação é dividir as impressões digitais em algumas categorias ou classes predefinidas.

Classificação de impressões digitais significa atribuir uma impressão digital a uma classe de forma consistente e confiável, tal que uma impressão digital desconhecida possa ser procurada apenas no sub-conjunto do banco de dados. O processo de classificação está baseada geralmente em macro-características, como a estrutura global das cristas. Enquanto ao processo de casamento da impressão digital, este é normalmente baseada em micro-estruturas tais como as terminações de cristas ou bifurcações.

A classificação de impressões digitais tem sido objeto de vários estudos na área de reconhecimento de padrões. No estado da arte há propostas diversas para resolver este problema, e já é possível projetar sistemas de classificação que sejam capazes de satisfazer o requisito de precisão do FBI (*Federal Bureau of Investigation*), 99% com uma rejeição máxima de 20% (Maltoni et al., 2003).

## 1.1 Caracterização do problema

Na dissertação “Um Sistema de Classificação de Impressões Digitais” (Samatelo, 2007) foi implementado um sistema de classificação de impressões digitais baseado na arquitetura PCASYS (Candela et al., 1995), na qual o autor indicou que nesse trabalho “o classificador que obteve os melhores resultados foi um classificador Bayesiano Normal Multivariável” (em que se obteve 6.75% de taxa de erro de classificação para o problema de 4 classes considerando uma distribuição balanceada) em comparação com outras técnicas de classificação como o classificador baseado em  $k$ -vizinhos mais próximos ( $k$ -NN), o perceptron multi-camada (MLP) e a rede neural probabilística (PNN) treinados e testados utilizando o banco de imagens de impressões digitais número 4 do NIST (*National Institute Standard Technology*) no qual os vetores de características estiveram constituídos pelos coeficientes PCA (*Principal Component Analysis*).

Por causa do resultado alcançado o autor concluiu que os vetores de características vinculados a cada classe apresentariam uma distribuição unimodal e que poderiam ser razoavelmente bem representadas por uma distribuição gaussiana multivariável. Não obstante, poder-se-ia considerar esta conclusão questionável, pois não foi discutido o porquê do fraco desempenho do MLP (em que se obteve uma taxa de erro de 7.27% para o problema de 4 classes considerando uma distribuição balanceada) cujo desempenho com poucos neurônios deveria pelo menos se aproximar ao desempenho do classificador normal multivariável.

Dessa forma, torna-se importante analisar se seria possível assumir que a distribuição no espaço multidimensional dos vetores de características -constituídos pelos coeficientes PCA- das impressões digitais de todas as classes poderiam ser descritas usando funções de densidade de probabilidade (fdp) Gaussianas, avaliando-se a factibilidade de modelar

as distribuições de cada classe usando apenas uma função de densidade Gaussiana (se a distribuição fosse simples) ou um conjunto de Gaussianas misturadas (se a distribuição fosse mais complexa).

## 1.2 Proposta

No presente trabalho de pesquisa propõe-se investigar, usando Modelo de Misturas Gaussianas (GMM) segundo uma abordagem Não-Supervisionada e um outra abordagem Supervisionada, se seria possível considerar a função de densidade de probabilidade Gaussiana como um modelo adequado para representar a distribuição dos vetores de características de imagens de impressões digitais e a partir do resultado enfrentar, usando a arquitetura PCASYS, o problema de classificação de impressões digitais.

Na **abordagem Não-Supervisionado** realiza-se a estimação de parâmetros de um GMM a partir da distribuição dos vetores de características não rotuladas de todas as impressões digitais de treinamento juntas com o objetivo de avaliar se seria possível modelar cada classe com apenas uma função de densidade de probabilidade Gaussiana. Por outro lado, na **abordagem Supervisionado** realiza-se a estimação de parâmetros de um GMM a partir da distribuição dos vetores de características das impressões digitais apenas de uma classe com o objetivo de avaliar se seria possível modelar cada classe pela combinação linear de funções de densidade de probabilidade Gaussiana (dada por o GMM).

No estado da arte da área de reconhecimento de padrões, assim como na área de mineração de dados, tem-se desenvolvida a teoria de agrupamento de dados usando modelos estatísticos dentro de um contexto probabilístico que trabalham sobre dados de treinamento não rotulados usando-se GMM (chamado também de *soft clustering*). No caso da abordagem Não-Supervisionado sugere-se usar o algoritmo EM (*Expectation-Maximization*) para realizar a estimação dos parâmetros do GMM, se esperando que a distribuição dos vetores de características de cada classe seja razoavelmente bem representada por uma das Gaussianas que compõem a mistura. Já no caso da abordagem Supervisionado, propõe-se utilizar o algoritmo SMEM (*Split and Merge EM*) para realizar a estimação dos parâmetros do GMM que modela a distribuição dos vetores de características das impressões digitais de cada classe.

Posto que em ambas as abordagens estimam-se os parâmetros das funções de densidade de probabilidade que modelam cada classe, a classificação, tanto de impressões digitais que foram parte do treinamento quanto das que não, pode ser feita usando o classificador Bayesiano no caso Não-Supervisionado. No caso Supervisionado a classificação pode ser feita usando-se a técnica GMBC (*Gaussian mixture based classification*).

A avaliação da eficiência da classificação segundo as duas abordagens é realizado usando

a taxa de erro de classificação, já que também permite comparar o resultados da proposta deste trabalho com os resultados alcançados em outros trabalhos que constituem o estado da arte na classificação de impressões digitais. Adicionalmente propõe-se realizar a análise dos resultados da estimação dos parâmetros do GMM obtidos na abordagem Não-Supervisionado usando o teste de hipótese e a medição da entropia mutua.

### 1.3 Organização da dissertação

O conteúdo desta dissertação está organizado da seguinte forma:

No capítulo 2 apresentam-se os conceitos básicos do reconhecimento biométrico e a classificação de impressões digitais, revisa-se as técnicas de classificação propostas para a solução deste problema, descreve-se a arquitetura PCASYS e faz-se uma revisão do estado da arte da classificação de impressões digitais e revisa-se as abordagens feitas até a atualidade juntamente com os resultados alcançados.

Já que o presente trabalho foi desenvolvido a partir das conclusões extraídas no trabalho “Um Sistema de Classificação de Impressões Digitais” (Samatelo, 2007), No capítulo 3 é feito uma revisão daquele trabalho; revê-se as técnicas baseadas na arquitetura PCASYS que foram propostas e usadas pelo autor na etapa de processamento da imagem da impressão digital e também revê-se a etapa de classificação e os resultados alcançados naquele trabalho.

No capítulo 4 contextualiza-se a proposta deste trabalho e apresenta-se o marco teórico das técnicas de estimação de parâmetros das misturas Gaussianas, também são apresentadas as técnicas de inicialização dos algoritmos de estimação propostos. Assim mesmo, são revisadas de forma resumida as ferramentas de avaliação estatística que foram adaptadas, integradas e implementadas neste trabalho.

Nos três capítulos seguintes são apresentados os resultados alcançados neste trabalho, dessa forma no capítulo 5 apresenta-se os resultados da avaliação de parâmetros estimados via o teste de hipótese e a entropia mútua, no capítulo 6 apresenta-se a metodologia empregada e os resultados da classificação alcançados sob a abordagem Não-Supervisionado; da mesma forma no capítulo 7 são apresentados a metodologia e os resultados da classificação obtidos sob a abordagem Supervisionado.

Finalmente, no capítulo 8 são apresentadas as conclusões derivadas do trabalho e levantam-se possibilidades futuras relativas à continuidade da pesquisa.

# Capítulo 2

## Definições e Estado da Arte

### 2.1 Reconhecimento Biométrico

Reconhecimento biométrico refere-se ao uso de características fisiológicas distintivas (por exemplo, impressões digitais, retina, rosto, íris) e comportamentais (por exemplo, assinatura, maneira de andar) chamados de identificadores biométricos, ou simplesmente de biometrias, que permitem a autenticação ou reconhecimento de indivíduos de forma automática.

### 2.2 Sistema biométrico

Um sistema biométrico é essencialmente um sistema de reconhecimento de padrões que identifica um indivíduo pela determinação de la autenticidade de uma determinada característica fisiológica e/ou comportamental. Por isso, uma questão de suma importância no desenvolvimento de um sistema biométrico é determinar *como* o indivíduo é reconhecido. Dependendo do contexto de aplicação, um sistema biométrico pode ser chamado de sistema de verificação ou sistema de identificação.

Um **sistema de verificação** autentica a identidade de uma pessoa fazendo a comparação das características biométricas captadas contra seus próprios modelos biométricos pré-armazenados no sistema. Isto conduz a comparações *um-para-um* que determinam se a identidade declarada pelo indivíduo é verdadeira. Um sistema de verificação rejeita ou aceita a alegação de identificação respondendo a pergunta: “Sou eu quem afirmo que eu sou?”.

Um **sistema de identificação** reconhece um indivíduo pesquisando todo o banco de dados de modelos para um casamento. Isto conduz a uma comparação *um-a-muitos* que deter-

mina a identidade do indivíduo. Um sistema de identificação estabelece a identidade de um sujeito (ou não, se o sujeito não foi inscrito no banco de dados do sistema) respondendo a pergunta: “Quem sou eu?” sem que o mesmo tenha que reivindicar uma identidade.

### 2.3 Impressões digitais

Uma impressão digital é a reprodução do relevo das elevações da epiderme na ponta dos dedos, produzida quando o dedo é pressionado contra uma superfície lisa. A característica estrutural mais evidente de uma impressão digital é um padrão de cristas e vales intercalados. Em uma imagem de impressão digital as cristas (também chamadas de linhas de crista) são escuras enquanto os vales são claros (veja a Figura 2.1). As cristas variam em largura de  $100\mu\text{m}$ , no caso de cristas muito finas, até  $300\mu\text{m}$  no caso das cristas mais grossas. Os padrões das impressões digitais, quando são analisadas em diferentes escalas, apresentam diferentes tipos de características: no nível global, no nível local, e no nível fino.

No **nível global** o fluxo das cristas traçam padrões geométricos reconhecíveis por simples observação, semelhantes ao mostrado na Figura 2.2; geralmente a determinação do padrão a que pertence uma impressão digital é obtido mediante o conhecimento das linhas tipo e dos pontos singulares chamados *core* e *delta*. As linhas tipo são definidas como duas cristas que iniciam-se paralelamente e divergem sobre área principal da impressão digital ou área padrão. Os pontos *core* são os pontos onde as cristas apresentam maior curvatura e pode ser definido como o ponto superior da linha de crista mais interna. Os pontos *delta* são os pontos de divergência das linhas tipo mais internas e são originadas pela bifurcação de uma linha simples ou pela brusca divergência de duas linhas paralelas. Os pontos singulares e as formas das linhas tipo são importantes na hora de fazer a indexação e a classificação de impressões digitais, mas, a sua distinção não é suficiente para fazer o casamento com precisão.

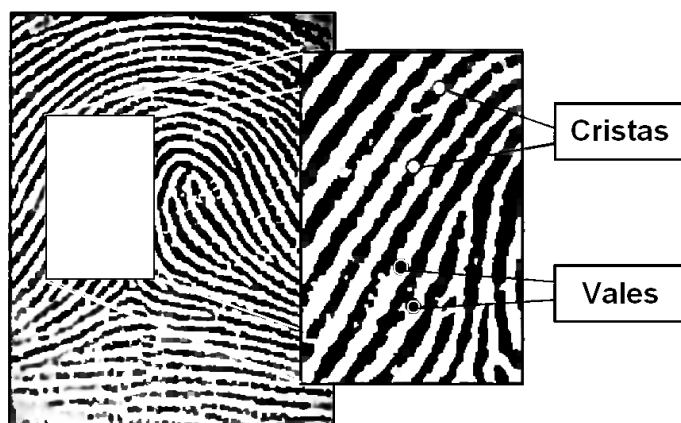


Figura 2.1: Cristas e vales numa impressão digital. (Maltoni et al., 2003)

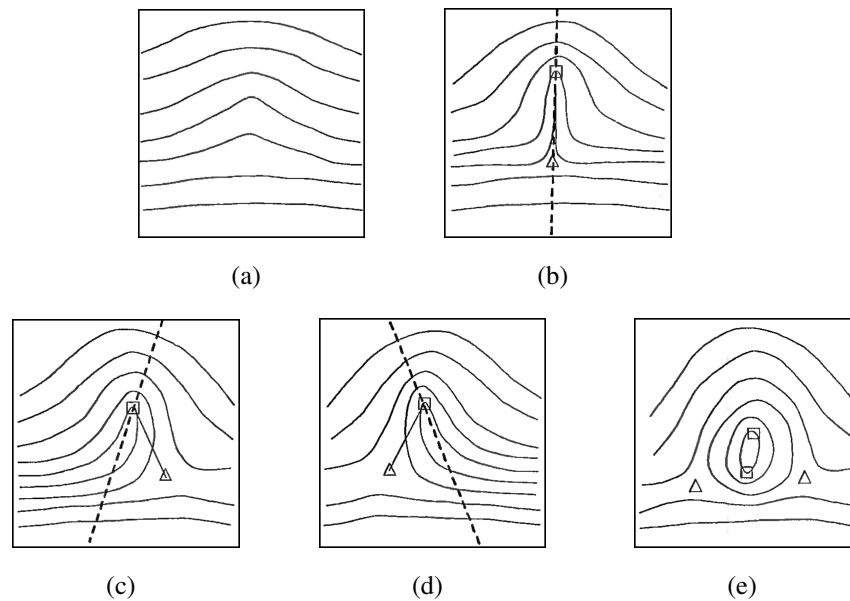


Figura 2.2: Padrões de impressões digitais distintas no nível global. Os quadrados e triângulos denotam os pontos singulares *core* e *delta* respectivamente. (Maltoni et al., 2003)

No **nível local** as linhas das cristas não são contínuas nem retas devido a que elas mudam de direção ou se bifurcam; estas características locais são chamados de detalhes minuciosos ou minúcias e devido a que não estão uniformemente distribuídas proveem informação suficiente para determinar a individualidade da impressão digital. As minúcias de finalização (onde a crista termina abruptamente) e de bifurcação (onde a crista se bifurca em duas ou mais cristas) são consideradas como os tipos mais importantes, posto que poderiam gerar outros tipos de minúcias.

No **nível fino** podem ser detectados detalhes intra-cristas das impressões digitais e estão dados essencialmente pelos poros de suor dos dedos (veja Figura 2.3) cujas posições e formas são consideradas altamente distintivas. No entanto, a extração de poros é factível apenas em imagens de boa qualidade e de alta resolução (por exemplo, 1000 pontos por polegada).

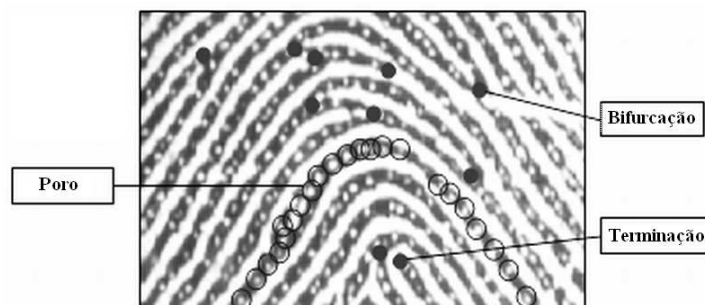


Figura 2.3: Minúcias (círculos pretos preenchidos) em uma porção da imagem de impressão digital; poros de suor (círculos vazios) em uma linha de crista. (Maltoni et al., 2003)



## 2.4 Classificação de Impressões Digitais

A classificação de impressões digitais refere-se ao problema da atribuição da imagem de uma impressão digital a uma classe de maneira consistente e confiável. Embora que a correspondência da impressão digital normalmente seja realizada conforme as características locais (por exemplo minúcias), a classificação de impressões digitais está geralmente baseado nas características globais, tais como a estrutura global das cristas e as singularidades.

O primeiro estudo científico em profundidade sobre classificação de impressões digitais foi realizada por Francis Galton, no qual dividiu-se as impressões digitais em três classes principais: *Arch*, *Loop* e *Whorl*. Ao mesmo tempo, Juan Vucetich foi o primeiro em fazer a identificação de um suspeito em 1892. Alguns anos mais tarde Edward Henry redefiniu a classificação de Galton incrementando o número de classes. O sistema de classificação de Galton-Henry dividiu as impressões digitais em cinco categorias: *Arch* (A), *Tented Arch* (TA), *Left Loop* (LL), *Right Loop* (RL) e *Whorl* (W) cujas respectivas traduções são Arco, Arco Angular, Presilha Esquerda, Presilha Direita e Verticilo; as que são mostradas na Figura 2.4. A maioria dos esquemas de classificação utilizados atualmente pelas agências de aplicação da lei em todo o mundo são variantes do esquema de classificação de Galton-Henry.

As impressões digitais tipo *Arch* tem sulcos que entram de um lado, subindo e depois descendo, e que saem do lado oposto com que entraram. Este tipo de impressão digital não

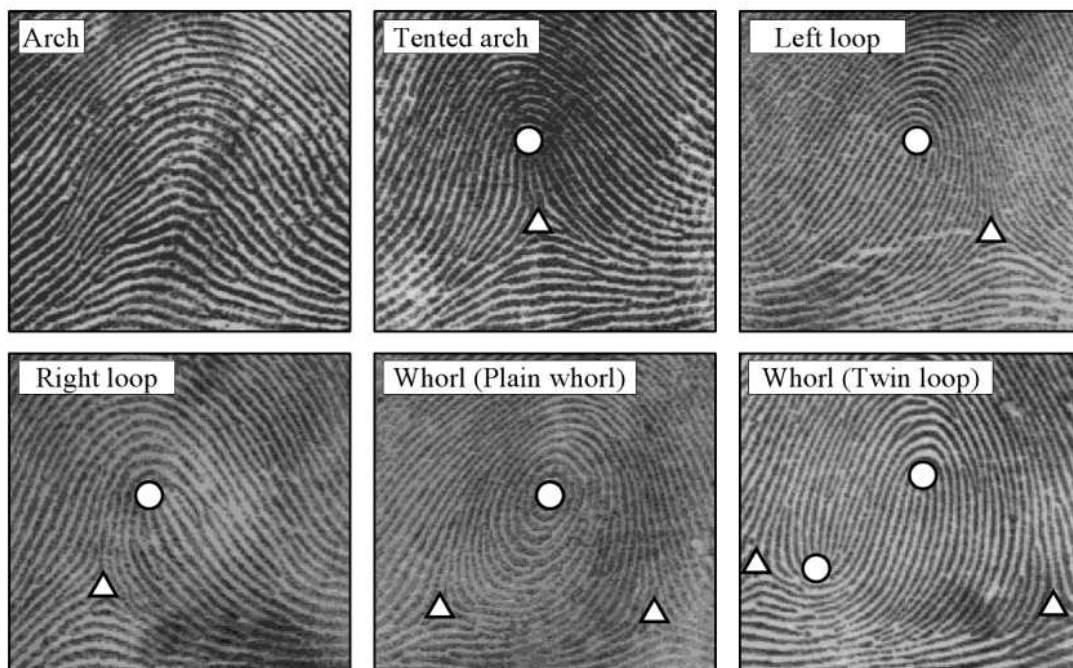


Figura 2.4: Impressões digitais do tipo *Arch*, *Tented Arch*, *Left Loop*, *Right Loop* e *Whorl*. São apresentadas impressões de duas sub-classes da classe *Whorl*. (Maltoni et al., 2003)

tem pontos *cores* nem *deltas*. As impressões digitais do tipo *Tented Arch* são semelhantes aos do tipo *Arch*, exceto que ao menos uma *crista* exibe uma grande curvatura; este tipo de impressão digital apresenta um ponto *core* e um ponto *delta*. As impressões digitais correspondentes aos tipos *Left Loop* e *Right Loop* têm pelo menos um sulco entrando de um lado, fazendo uma curva de volta, e saindo do mesmo lado do que entrou. Estas impressões apresentam um ponto *core* e um ponto *delta*; o *delta* é assumido como se estivesse no sul do *core*. Quando as impressões digitais têm cristas que entram e saem do lado esquerdo são chamados de *Left Loop* e quando têm cristas que entram e saem do lado direito são chamados *Right Loop*. As impressões digitais tipo *Whorl* contém pelo menos uma crista que faz um trajeto de 360° em torno do centro da impressão digital. Nestas impressões digitais podem ser encontrados dois pontos *core* (ou um *Whorl*) e dois pontos *delta*. Em alguns sistemas de classificação, estas impressões são divididos em sub-classes: *Twin Loop* ou duplo *Loop* e *Plain Whorl*.

A variação entre impressões digitais de classes distintas é conhecida como **variação inter-classe** e a variação entre impressões digitais de uma mesma classe é conhecida como **variação intra-classe**. Usualmente o problema de classificação de impressões digitais torna-se difícil de resolver devido a que há impressões pertencentes a classes distintas que têm pequena variação inter-classe (veja as três impressões digitais da fila superior da Figura 2.5) e porque há impressões digitais de uma mesma classe que têm grande variação intra-classe (veja as três impressões pertencentes à classe *Whorl* da fila inferior da Figura 2.5).

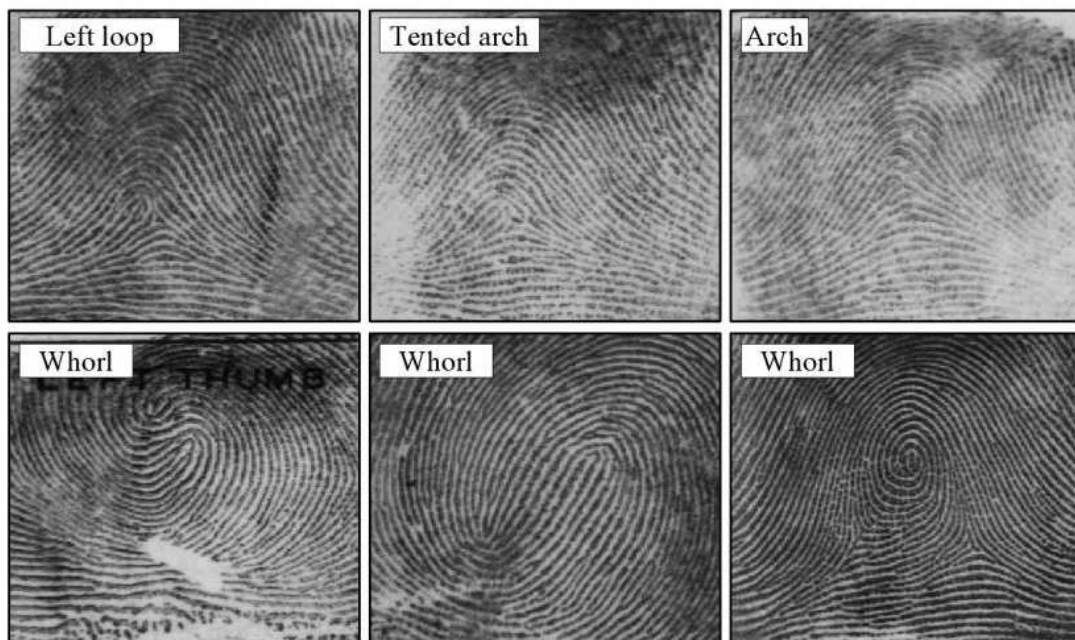


Figura 2.5: Três impressões de classes diferentes com pequena variação intra-classe (fila superior) e três impressões da mesma classes com grande variação intra-classe (fila inferior).

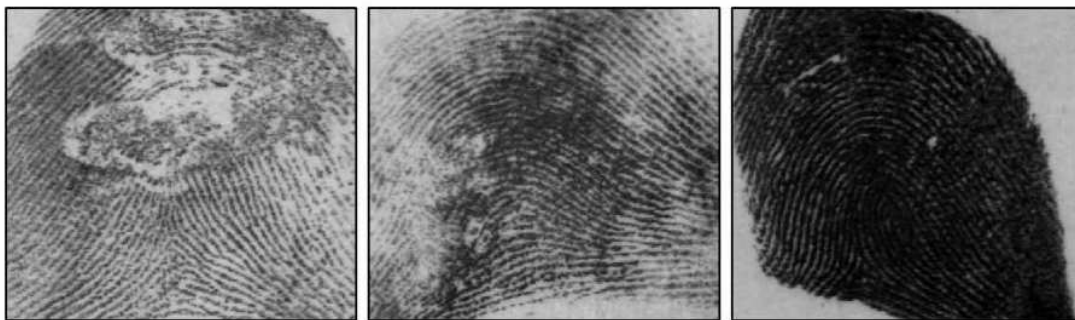


Figura 2.6: três exemplos de impressões digitais que contêm ruído na área padrão (Maltoni et al., 2003).

Um outro problema que complica ainda mais a tarefa de classificação surge quando as imagens das impressões digitais contêm ruído na área padrão originado pela técnica de aquisição ou pelas características do dispositivo de leitura da imagem. Exemplos de impressões digitais com ruído são mostradas na Figura 2.6.

Em muitos dos sistemas de identificação automatizados as impressões digitais são particionadas em classes predefinidas baseadas em macro-estruturas tais como a quantidade de pontos singulares ou as formas das linhas tipo. Sendo assim que o número de classes usado comumente é pequeno e que a quantidade de impressões digitais não estão uniformemente distribuídas; por exemplo em sistemas que consideram as classes A, TA, LL, RL e W a proporção de ocorrência na natureza das impressões digitais destas classes são de 3,7%, 2,9%, 33,8%, 31,7% e 27,9% respectivamente (Wilson et al., 1994).

Num experimento, se como resultado da partição num banco de imagens de impressões digitais a quantidade de impressões em cada classe tem-se a mesma proporção do que na ocorrência na natureza diz-se que o banco de dados experimental tem **distribuição natural**, mas, há alguns bancos de imagens de impressões digitais que consideram a mesma quantidade de impressões digitais de todas as classes, nesse caso a diz-se que o banco tem **distribuição balanceada**. Por outro lado, devido à alta porcentagem de falsa classificação de impressões da classe A como se fossem TA e vice-versa, assim mesmo, tendo em consideração que essas duas classes não são muito comuns na natureza, alguns pesquisadores propuseram juntar as duas classes em uma única classe chamada TA-A, originando dois tipos de problemas de classificação: o **problema de 5 classes** que considera as classes A, TA, LL, RL e W; e o **problema de 4 classes** que considera as classes TA-A, LL, RL e W.

## 2.5 Extração de características

O problema de classificação de impressões digitais tem atraído grande interesse na comunidade científica devido a sua importância e dificuldade intrínseca, e muitos trabalhos têm sido publicados sobre este tema durante os últimos trinta e cinco anos. Embora uma grande variedade de algoritmos de classificação já tenham sido desenvolvidos para este problema, apenas um número relativamente pequeno de técnicas de extração de características de impressões digitais são usados pela maioria dos autores. Em particular, quase todos as técnicas são baseados em imagem de orientação, fluxo de linhas das cristas, pontos singulares ou respostas do filtro de Gabor.

A **imagem de orientação** é uma matriz discreta cujos elementos representam a orientação média local das cristas da impressão digital como se mostra na Figura 2.7(a). Normalmente a imagem de orientação é registrada com relação ao ponto central das impressões digitais antes de ser processada. Na medida em que tal recurso caso seja calculado com suficiente precisão e detalhe pode conter todas as informações necessárias para a classificação (Capelli et al., 1999) e (Candela et al., 1995).

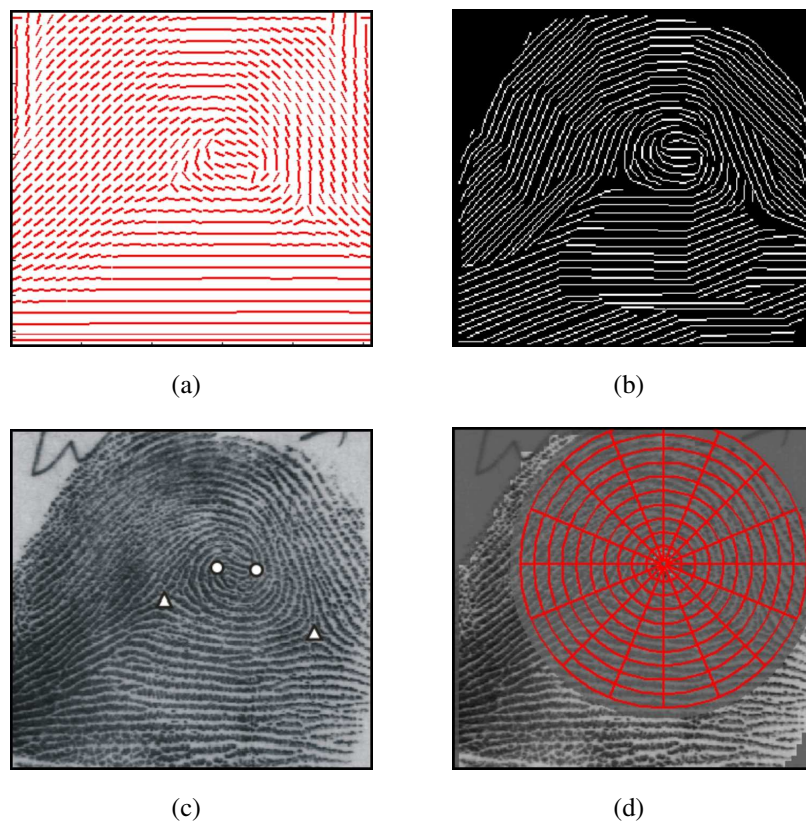


Figura 2.7: Características mais frequentemente utilizados na classificação de impressões digitais: (a) imagem de orientação, (b) linhas de fluxo de crista, (c) pontos singulares (d) resposta de filtro de Gabor. (Samatelo, 2007)

O **fluxo de linhas de cristas** é geralmente representado como um conjunto de curvas paralelas às linhas de cristas e vales. Essas curvas não coincidem necessariamente com as cristas de impressões digitais, mas, exibem a mesma orientação local como se mostra na Figura 2.7(b). O fluxo de linhas das cristas pode ser traçado fazendo o desenho das curvas localmente orientados de acordo com a imagem de orientação (Candela et al., 1995).

As posições dos **pontos singulares** (*core* e *delta*) como se mostra na Figura 2.7(c) são muito úteis para o alinhamento de impressões digitais com respeito a um ponto fixo. Muitas técnicas de detecção de singularidades estão baseadas no índice de Poincaré, o qual é calculado sobre uma curva fechada ao redor de um ponto particular (Kawagoe e Tojo, 1984).

Na classificação de impressões digitais a resposta dos **filtros de Gabor** é usada para representar as impressões digitais de forma análoga à representação do *Iris Code* - introduzido em (Daugman, 1993) -. Tal representação que se mostra na figura 2.7(d) é chamada de *Finger Code* (Jain et al., 1999) e pode ser usado tanto em tarefas de classificação quanto de casamento. Além disso, os filtros de Gabor também podem ser usados no melhoramento das imagens devido à propriedade seletiva de frequência e orientação (Hong et al., 1998).

## 2.6 Técnicas de Classificação

A maioria das técnicas de classificação existentes no estado da arte da classificação de impressões digitais pode ser atribuída às seguintes categorias: técnicas baseadas em regras, técnicas sintáticas, técnicas estruturais, técnicas estatísticas, técnicas baseadas em redes neurais e técnicas de múltiplos classificadores.

Nas técnicas de **classificação baseadas em regras**, uma impressão digital pode ser simplesmente classificada de acordo com o número e a posição das singularidades. Este método é comumente usado pelos especialistas forenses no processo de classificação manual. Por conseguinte, vários autores adotaram a mesma técnica para a classificação automática.

As técnicas de **classificação sintáticas** descrevem os padrões das impressões digitais por meio de símbolos terminais e regras de produção. Assim, uma gramática é definida para cada classe e um processo de análise é responsável por classificar cada novo padrão (Fu e Booth, 1986).

As técnicas de **classificação estruturais** são baseadas na organização relacional de recursos de baixo nível em estruturas de alto nível. Esta organização relacional é representada por meio de estruturas simbólicas de dados como grafos ou árvores que permitem estabelecer uma organização hierárquica da informação. A imagem de orientação é a adequada para a representação estrutural, já que a imagem pode ser particionado em regiões ligadas que são

caracterizadas por orientações “homogêneas”. Estas regiões e as relações entre elas contêm informações úteis para a classificação.

Nas técnicas de **classificação estatísticas**, um vetor numérico de tamanho fixo é derivado de cada impressão digital e um classificador estatístico de propósito geral é utilizado para a classificação. Um dos classificadores mais usados é a técnica de  $k$ -NN. Muitas técnicas usam a imagem de orientação como vetor de características, ao representar cada elemento da imagem de orientação através de seus componentes vectoriais e aninhar as filas das matrizes conformantes de tais componentes. Por exemplo, uma típica imagem de orientação de  $30 \times 30$  elementos resulta num vetor de 1800 ( $= 30 \times 30 \times 2$ ) elementos. O treinamento dos classificadores com vetores de tão alta dimensão requer grande quantidade de dados de treinamento, memória e tempo de cálculo. Por isso, há propostas técnicas estatísticas de redução dimensional do vetor de características. Normalmente é adotada a técnica de análise das componentes principais (PCA), já que ela garante uma boa preservação da distancia euclidiana.

A maioria das **técnicas de classificação por redes neurais** são baseadas no perceptron de múltiplas camadas e usam os elementos da imagem de orientação como características de entrada. Alguns autores apresentam uma arquitetura piramidal, constituída de várias redes tipo perceptron, cada uma das quais é treinada para reconhecer as impressões digitais pertencentes a uma classe diferente.

Em alguns estudos de classificação de padrões foram observados que diferentes classificadores rotulam de forma distinta os mesmos padrões. Isso motiva o interesse recente na combinação de diferentes abordagens para as tarefas de classificação de impressões digitais. As **técnicas baseadas em múltiplos classificadores** potencialmente oferecem informações complementares a respeito dos padrões a serem classificados, os quais podem ser utilizados para melhorar o desempenho. Várias opções são possíveis para a seleção dos componentes dos classificadores (por exemplo classificadores diferentes treinados com os mesmos dados, o mesmo classificador treinado com dados diferentes ou diferentes características de entrada) e para a estratégia de combinação (de simples critérios heurísticos da regra voto maioritário até técnicas mais complexas que envolvem o treinamento de um classificador adicional para a decisão final).

Na Tabela 2.1 são mostrados alguns trabalhos de classificação de impressões digitais conjuntamente com as técnicas de classificação que foram utilizados em cada uma delas e o ano de publicação.

Abordagem	características				classificador						ano
	O	S	R	G	Rb	Sy	Str	Sta	Nn	Mc	
Moayer e Fu	x	-	-	-	x	-	-	-	-	-	1975
Moayer e Fu	x	-	-	-	x	-	-	-	-	-	1976
Rao e Balk	x	-	-	-	x	-	-	-	-	-	1980
Kawagoe e Tojo	-	x	x	-	x	-	-	-	-	-	1984
Hughes e Green	x	-	-	-	-	-	-	-	x	-	1991
Bowen	x	x	-	-	-	-	-	-	x	-	1992
Kamijo, Mieno e Kojim	x	-	-	-	-	-	-	-	x	-	1992
Moscinska e Tyna	x	-	-	-	x	-	-	-	x	-	1993
Kamijo	x	-	-	-	-	-	-	-	x	-	1993
Wilson, Candela e Watson	x	-	-	-	-	-	-	-	x	-	1995
Omidvar, Blue e Wilson	x	-	-	-	-	-	-	-	x	-	1995
Candela, Grother, Watson e Wilson	x	-	x	-	x	-	-	-	x	x	1995
Maio e Maltoni	x	-	-	-	-	-	x	-	-	-	1996
Halici e Ongun	x	-	-	-	-	-	-	-	x	-	1996
Karu e Jain	-	x	-	-	x	-	-	-	-	-	1996
Chong, Ngee, Jun e Gay	-	-	x	-	x	-	-	-	-	-	1997
Ballan, Sakarya e Evans	-	x	-	-	x	-	-	-	-	-	1997
Senior	-	-	x	-	-	-	x	-	-	-	1997
Wei, Yuan, e Yie	x	-	-	-	x	-	-	-	x	-	1999
Cappelli, Lumini, Maio e Maltoni	x	-	-	-	-	x	-	-	-	-	1999
Lumini, Maio, e Maltoni	x	-	-	-	-	x	-	-	x	-	1999
Jain, Prabhakar, e Hong	-	-	-	x	-	-	-	x	-	-	1999
Hong e Jain	-	x	x	-	x	-	-	-	-	-	1999
Cappelli, Maio, e Maltoni	x	-	-	-	-	-	-	x	-	-	1999
Cappelli, Maio, e Maltoni	x	-	-	-	x	-	-	x	-	-	2000
Cho, Kim, Bae, Bae e Yoo	-	x	-	-	x	-	-	-	-	-	2000
Bartesagui, Fernandez e Gomez	-	x	-	-	-	-	-	-	-	-	2001
Bernard	x	-	-	-	-	-	-	-	x	-	2001
Pattichis, Panayi, Bovic e Hsu	x	-	-	-	x	-	-	-	x	x	2001
Marcialis, Roli e Frasconi	x	-	-	x	-	-	x	x	x	x	2001
Senior	x	-	x	-	x	-	x	-	x	x	2001
Yao, Frasconi e Pontil	-	-	-	x	-	-	-	x	-	x	2001
Cappelli, Maio e Maltoni	x	-	-	-	-	-	-	x	-	x	2002
Jain e Minut	-	-	x	-	x	-	-	-	-	-	2002
Samatelo	x	-	-	-	-	-	-	x	x	-	2007

Tabela 2.1: Trabalhos de classificação de impressões digitais, técnicas de extração de características (Orientação (O), Singularidades (S), fluxo de cristas (R) e Gabor (G)), técnicas de classificação utilizadas (regras (Rb), sintática (Sy), estruturais (Str) classificadores por redes neurais (Nn), múltiplos classificadores (Mc)) e ano de publicação. Informação extraída de (Maltoni et al., 2003) e (Samatelo, 2007)

## 2.7 Arquitetura PCASYS

PCASYS (*Pattern level Classification Automation SYStems*) é a arquitetura de um sistema de classificação completa de impressões digitais baseada em redes neurais proposta pelos pesquisadores do NIST (Candela et al., 1995) e (Wilson et al., 1994) e conforma um marco para sucessivos estudos sobre a classificação de impressões digitais, posto que tem documentação disponível e código fonte aberto. Esta arquitetura é uma das mais referenciadas em trabalhos de pesquisa de classificação de impressões digitais uma vez que os trabalhos que reportam resultados precisos podem ser reproduzidos usando bancos de imagens de impressões digitais conhecidas e disponíveis.

Esta arquitetura permite fazer a comparação do desempenho de diferentes tipos classificadores, a avaliação de diferentes técnicas de melhoria da imagem de impressões digitais e o melhoramento na exatidão da classificação. A Figura 2.8 mostra um esquema funcional do PCASYS. A sua arquitetura tem as seguintes etapas: cálculo de imagens em regiões, melhora da imagem, cálculo e registro da imagem de orientação, redução dimensional e a etapa de classificação.

Na etapa de **cálculo de imagens em regiões** se detecta a área da imagem de impressões digitais em escala de cinza que tem presença de cristas; na seguinte etapa se realiza a **melhora da imagem** no domínio da frequência e depois se realiza o **cálculo e registro da imagem de orientação** gerando assim uma matriz de  $28 \times 30$  elementos. Na etapa de **redução dimensional** a imagem de orientação é representada por um vetor de duas componentes, contendo 1680 elementos ( $= 28 \times 30 \times 2$ ). Este vetor é reduzido a 64 elementos pelo uso da técnica de PCA. Finalmente, a **etapa de classificação** tem uma rede neural probabilística que é usada para atribuir cada vetor de 64 elementos a uma das classes do esquema de classificação. Para melhorar a confiabilidade, os autores também implementaram um módulo

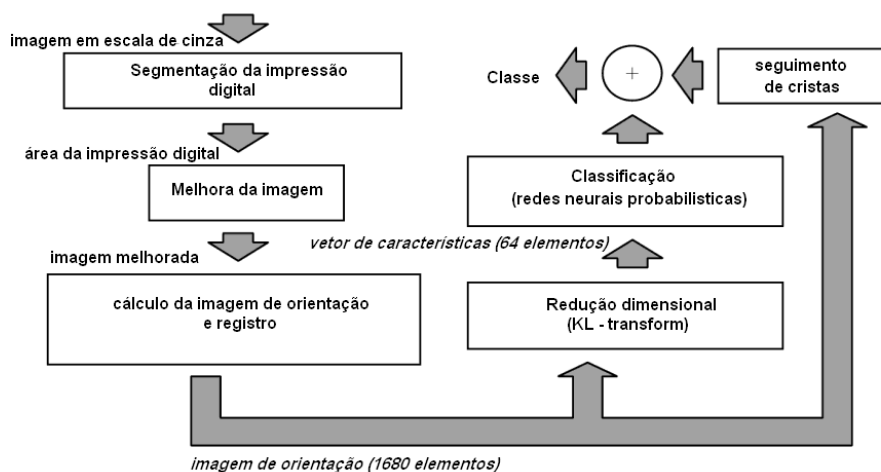


Figura 2.8: Esquema funcional do PCASYS (Candela et al., 1995)



adicional chamado seguidor de cristas, o qual determina o fluxo de cristas na parte superior da impressão digital; este módulo está projetado especificamente para detectar impressões digitais tipo *Whorl*.

## 2.8 Banco de dados DB4

Em 1992 e 1993 o NIST liberou duas bases de dados de imagens de impressões digitais bem adaptadas denominados DB4 (*NIST Special Database 4*) detalhada em (Watson e Wilson, 1992) e DB14 (*NIST Special Database 14*) detalhada em (Watson e Wilson, 1993) para permitir o desenvolvimento e avaliação de sistemas automatizados de classificação.

O banco de dados DB4 contém 4000 imagens de impressões digitais roladas, escolhidas aleatoriamente e digitalizadas a partir de cartões usando escalas de cinza de 8 bits e 500 pixels por polegada; foram 2000 impressões coletadas duas vezes cada uma, etiquetados de F000 a F2000 e de S0001 a S2000 com o prefixo F (*first*) na primeira coleta e S (*second*) na segunda coleta. As impressões digitalizadas foram comprimidas bidimensionalmente usando o algoritmo WSQ (*wavelet scalar quantization*), resultando cada imagem com  $512 \times 512$  pixels. As imagens foram analisadas, classificadas e rotuladas manualmente por um especialista forense segundo as cinco categorias de Galton-Henry (*Arch*, *Tented Arch*, *Left Loop*, *Right Loop* e *Whorl*) distribuídas uniformemente com a mesma quantidade de imagens das impressões digitais de cada classe.

Algumas impressões ambíguas têm uma referência adicional para uma segunda classe originando uma dupla categorização a causa da presença de impressões digitais com cicatrizes sobre a superfície dos dedos, ou pela qualidade do processo de coleta da impressão. DB4 e DB14 se tornaram bancos de dados padrão na classificação de impressões digitais e muitos trabalhos de pesquisa e artigos publicados foram testados com algum destes bancos.

## 2.9 Resultados atuais

O desempenho dos sistemas de classificação de impressões digitais é normalmente medido em termos de taxa de erro ou exatidão. A taxa de erro é calculada como a relação entre o número de impressões digitais erroneamente classificadas e o número total de amostras no conjunto de teste. A exatidão é simplesmente a porcentagem de impressões digitais corretamente classificadas (Maltoni et al., 2003). Assim;

$$T_{erro} = \frac{N_{ma}}{N_T} \times 100\%, \quad (2.1)$$

onde  $T_{erro}$  é a taxa de erro,  $N_{ma}$  é o número de impressões digitais classificadas em outras classes e  $N_T$  é o número total de impressões. A exatidão é expressa por:

$$Exatidao = (100 - T_{erro})\%. \quad (2.2)$$

A Tabela 2.9 mostra as taxas de erro obtidas num conjunto de trabalhos em que foram usadas distintas abordagens, se utilizando o banco de dados DB4, sendo que a maioria delas foi obtida usando 2000 imagens dos primeiros 1000 dedos para o treinamento, e as 2000 imagens restantes para o teste do sistema, no entanto, alguns métodos baseados em regras foram testados em todo o banco. Também reportam-se resultados de 542 impressões digitais selecionados aleatoriamente.

Quanto às impressões ambíguas com dois rótulos de classe, o resultado geralmente é assumido correto se a classe hipotetizada pelo classificador corresponde a qualquer uma das duas etiquetas. Na tabela pode-se ver que os experimentos foram realizados par os problemas de classificação de 5 e 4 classes e considerando as distribuições balanceada e natural.

Abordagem	5 classes		4 classes		Conjunto de teste	Ano
	B	N	B	N		
Blue et al.	-	7.2	-	-	segunda metade	1993
Candela et al.	-	-	11.4	6.1	segunda metade	1995
Karu e Jain	14.6	11.9	8.6	9.4	tudo	1996
Senior	-	-	-	8.4	542 aleatórias	1997
Jain et al.	10.0	7.0	5.2	-	segunda metade	1999
Cappelli et al.	-	12.9	-	-	segunda metade	1999
Cappelli et al.	7.9	6.5	5.5	-	Data set 2	1999
Hong e Jain	12.5	10.6	7.7	-	tudo	1999
Senior	-	-	-	5.1	segunda metade	2001
Yao et al.	10.7	9.0	6.9	-	segunda metade	2001
Marcialis et al.	12.1	9.6	-	-	segunda metade	2001
Jain e Minut	-	-	8.8	9.3	tudo	2002
Zhang e Yan	15.7	-	7.3	-	tudo	2004
Dass e Jain	-	-	5.6	-	tudo	2004
Tan et al.	-	-	6.1	-	segunda metade	2005
Park e Park	9.3	-	6.0	-	tudo	2005
(Samatelo, 2007)	12.3	8.34	6.75	7.19	segunda metade	2007
(Liu et al., 2007)	4.2	2.9	3.1	2.9	tudo	2007

Tabela 2.2: Taxas de erro de classificação sobre a base de dados DB4 considerando as distribuições natural (N) e balanceada (B). Dados baseados em informação fornecida em (Maltoni et al., 2003), (Maltoni et al., 2009), (Samatelo, 2007), (Liu et al., 2007).

# Capítulo 3

## Trabalho Prévio

Visto que o presente trabalho de pesquisa foi proposto levando em consideração os resultados, as conclusões e as sugestões apresentados no trabalho prévio denominado “Um Sistema de Classificação de Impressões Digitais” (Samatelo, 2007) neste capítulo é feita uma revisão dos principais aspectos daquele trabalho.

O trabalho prévio dividiu-se claramente numa etapa de processamento e numa etapa de classificação, porém, este capítulo focaliza-se na etapa de processamento e extração de características da imagem das impressões digitais propostas e utilizadas pelo autor, dessa forma na etapa de processamento, revisa-se as sub-etapas de segmentação e melhora da imagem assim como na determinação da imagem de orientação e o ponto de referência. Na etapa de classificação revisa-se a representação vetorial, a geração de características, assim como, a redução dimensional do vetor de características. Também são apresentados os resultados alcançados usando diversos classificadores, tais como o classificador normal, o classificador baseado em  $k$ -NN e os classificadores baseados em redes neurais MLP e PNN.

### 3.1 Processamento da Impressão Digital

Na etapa de processamento da impressão digital o autor propôs técnicas originais de segmentação, de melhora da impressão digital, de determinação da imagem de orientação e de determinação do ponto de referência; assim, os resultados da etapa de processamento estão dadas pela determinação da imagem em regiões, em que se indica qual parte de uma imagem de uma impressão digital é válida e qual outra é o fundo, pela determinação da imagem de orientação que representa a informação direcional da impressão digital e pela determinação do ponto de referência que permite realizar um registro dando a todas as impressões digitais um ponto de referência comum com o objetivo de eliminar as variações originadas pela traslação da imagem.

### 3.1.1 Determinação de Regiões na Imagem

No que se refere à determinação de regiões na imagem, o autor propôs uma técnica baseada na segmentação, que divide a imagem em blocos e depois extrai o valor mínimo da escala de cinza de cada bloco. Em seguida se faz uso de um filtro de mediana da inclusão de múltiplos limiares e de uma etapa de pós-processamento as quais permitem aumentar a capacidade de discriminação entre o objeto e o fundo. A Figura 3.1 mostra os resultados da segmentação de imagens de baixa qualidade quando as impressões digitais apresentam buracos, cortes, variação de contraste e anotações. As linhas brancas representam as fronteiras entre o objeto e o fundo. Pode-se encontrar uma explicação mais detalhada desta técnica em (Cho et al., 2000).

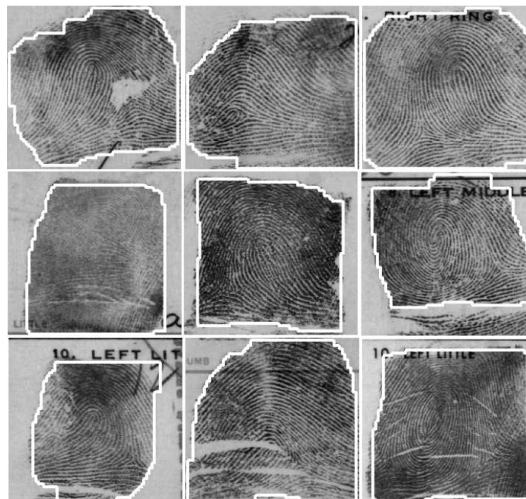


Figura 3.1: Resultados da segmentação. (Samatelo, 2007).

### 3.1.2 Melhora da Imagem

Para o melhoramento das imagens de impressões digitais foi proposta pelo autor uma técnica de normalização e melhora no domínio da frequência. A normalização da imagem foi baseada na transformada de *Top-hat* multi-escala (Mukhopadhyay e Chanda, 2000) para uniformizar e reduzir as variações dos níveis de cinza de cristas e vales, além de uniformizar o contraste entre as diferentes zonas da imagem. A melhora da imagem no domínio da frequência reduz o efeito do contraste e do ruído na imagem da impressão digital. O resultado deste processo pode-se observar na Figura 3.2.

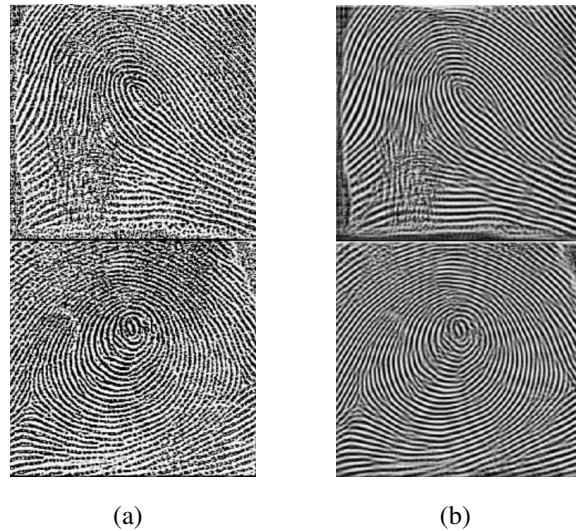


Figura 3.2: Imagens melhoradas; (a) Normalização. (b) Melhora no domínio da frequência (Samatelo, 2007).

### 3.1.3 Determinação da Imagem de Orientação

Com respeito à determinação da imagem de orientação, o autor testou diversas técnicas de estimação da orientação sendo a escolhida uma técnica que calcula a orientação dominante de cada bloco utilizando-se a média do quadrado do vetor gradiente (técnica padrão), mas que inclui uma segunda fase de média ponderada. Ou seja, usa-se a informação dada pela matriz de coerência para fazer a estimação local da orientação de maneira ponderada, suavizando adaptativamente a imagem de orientação.

A Figura 3.3 mostra as imagens de orientação obtidas usando tanto a técnica padrão quanto técnicas melhoradas.

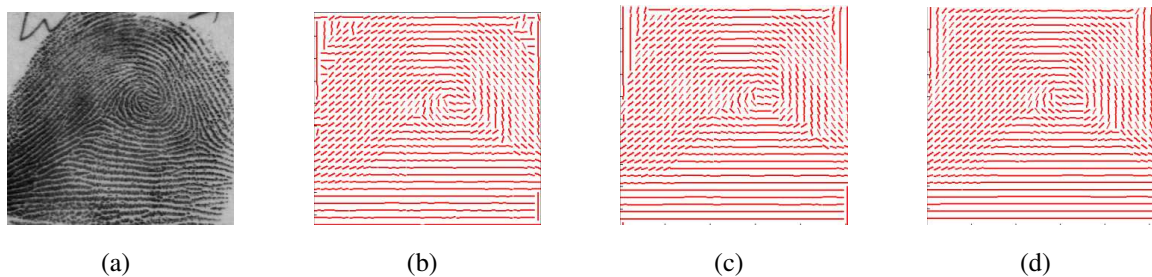


Figura 3.3: (a) Impressão digital de boa qualidade. (b) Imagem de orientação com técnica padrão, (c-d) com técnicas melhoradas (Jiang, 2005), (Wang et al., 2006).

### 3.1.4 Determinação do Ponto de Referência

No que se refere à determinação do ponto de referência, o autor propôs uma técnica baseada em múltiplos níveis de representação da imagem de orientação que permite usar as características morfológicas (diferença entre as orientações das linhas de fluxo das cristas das regiões superiores e inferiores) da imagem limiarizada para cada um dos seus valores de quantização. Uma explicação mais detalhada desta técnica pode ser encontrada em (Samatelo e Salles, 2009).

Na Figura 3.4 mostra-se a sequência de passos desta técnica. A Figura 3.4(a) mostra a quantização da imagem de orientação em 32 valores. As seguintes imagens mostram o cálculo das imagens da mistura binária (onde os pontos vermelhos indicam as esquinas mais internas do objeto binário superior de cada mistura binária e as linhas azuis indicam as rotas médias) mostradas nas Figuras 3.4(b-e); na Figura 3.4(f) mostra-se a imagem binária total onde os pontos vermelhos indicam todas as esquinas válidas e considera-se o ponto localizado na parte mais inferior como possível ponto de referência; na Figura 3.4(g) mostra-se o recálculo do ponto de referência através da matriz de coerência e na Figura 3.4(h) mostra-se o ponto de referência da impressão digital.

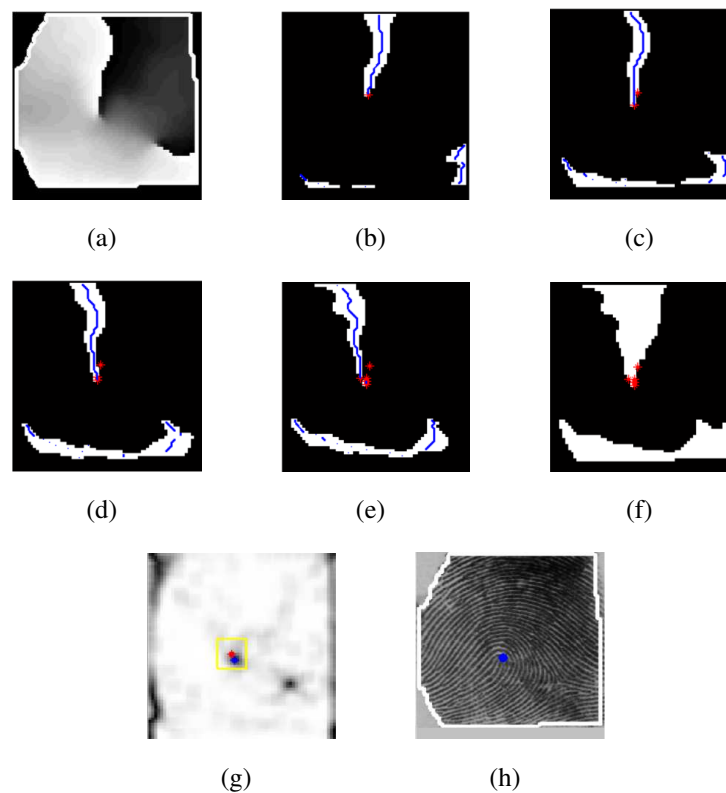


Figura 3.4: Sequência de passos no cálculo do ponto de referência. (Samatelo, 2007), (Samatelo e Salles, 2009)

## 3.2 Classificação

Nesta seção é revisada a etapa de classificação no qual realiza-se a tarefa de determinação da classe a que pertence uma impressão digital usando a informação obtida na etapa de processamento. Descreve-se brevemente as etapas próprias da classificação estatística as quais são: a geração de características a partir da imagem de orientação, a redução dimensional e também apresentam-se os resultados alcançados na classificação.

### 3.2.1 Geração de características

As características estão dadas pela representação vetorial da imagem de orientação, que é uma representação de maior dimensão gerada para eliminar as descontinuidades dos valores angulares cuja sequência de cálculo é mostrada na Figura 3.5. Dessa forma, na Figura 3.5(a) mostra-se a imagem de orientação, na Figura 3.5(b) mostra-se a imagem de orientação registrada segundo o ponto de referência, na Figura 3.5(c) mostra-se as duas componentes da representação vetorial da imagem de orientação registrada e na Figura 3.5(d) representa-se o resultado de vetorizar todos os elementos da representação vetorial da imagem de orientação.

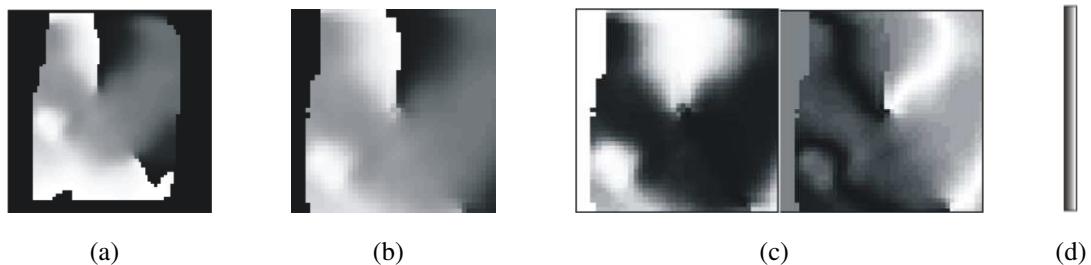


Figura 3.5: Sequência de obtenção do vetor de dados a partir da imagem de orientação. (Samatelo, 2007)

### 3.2.2 Redução dimensional

A representação vetorial tem grande quantidade de elementos, a qual, origina um espaço de características de alta dimensão o que faria, se fosse usado como vetor de características, que o tempo de treinamento ou classificação fosse muito grande. Por isso, com o objetivo de obter vetores de características com menor quantidade de elementos, o autor implementou uma etapa de redução dimensional baseada na técnica de análise das principais componentes ou coeficientes PCA, que é uma das técnicas mais importantes usadas na arquitetura PCASYS.

### 3.2.3 Classificação

Na etapa de classificação foram implementadas diversas técnicas como o classificador Bayesiano com distribuição normal (classificador normal multivariável), o classificador baseado nos  $k$ -vizinhos mais próximos ( $k$ -NN), o perceptron multi-camada (MLP) e a rede neural probabilística (PNN) todos eles treinados e testados utilizando o banco de imagens de impressões digitais DB4.

Na Tabela 7.2 são amostradas as menores taxas de erro alcançadas com o teste de cada técnica de classificação implementada no referido trabalho, também são amostrados a quantidade de coeficientes PCA no vetor de características com os quais se obtiveram tais resultados para os problemas de 5 e 4 classes e considerando as distribuições natural e balanceada. Percebe-se que as menores taxas de erro foram alcançadas com o Classificador Normal com 38 coeficientes PCA no vetor de características.

Técnica de Classificação	Problema de 5 classes			Problema de 4 classes		
	PCA	N	B	PCA	N	B
Classificador Normal	<b>38</b>	<b>8.34</b>	<b>12.30</b>	<b>38</b>	<b>7.19</b>	<b>6.75</b>
k-NN	32	14.28	16.40	26	13.36	9.85
MLP	112	9.33	10.68	112	9.24	7.27
PNN	44	10.00	21.30	48	9.56	15.7

Tabela 3.1: Menores taxas de erro de classificação alcançados no trabalho prévio (Samatelo, 2007) com distintos classificadores para os problemas de 5 e 4 classes considerando as distribuições balanceada (B) e natural (N).



# Capítulo 4

## Fundamentação Teórica

Neste capítulo apresenta-se as definições e algoritmos teóricos que foram empregados no desenvolvimento deste trabalho. São definidos os modelos de misturas, verosimilhança máxima, o algoritmo EM, o algoritmo SMEM, o algoritmo *Mean-Shift*, GMBC, entropia mutua e região de confiança. Porém, antes de fazer estas definições é feita a contextualização das mesmas em concordância com a proposta feita na seção 1.2.

### 4.1 Contexto

Neste trabalho propõe-se investigar se a função de densidade de probabilidade Gaussiana é um modelo adequado para afrontar o problema de classificação de impressões digitais usando GMM segundo as abordagens Não-Supervisionado e Supervisionado.

Na abordagem Não-Supervisionado estima-se os parâmetros do GMM que modela a distribuição dos vetores de características das impressões digitais de todas as classes. Usa-se o algoritmo EM (Dempster et al., 1977) para encontrar os parâmetros do GMM que maximize a função verosimilhança. Porém, o resultado deste processo iterativo depende fortemente dos parâmetros iniciais, já que poderiam fazer o algoritmo convergir num máximo local. Por isso, propõe-se três formas de inicializar os parâmetros: (a) com **parâmetros amostrais** (média e covariância dos vetores de características das impressões de cada classe), (b) com os **centroides** (estimados via o algoritmo *K-means* (Bishop, 2006)) e (c) com as **modas** das distribuições de cada classe (estimada via *Mean-Shift* (Fukunaga e Hostetler, 1975)). A classificação de impressões digitais com esta abordagem Não-Supervisionada é feita com o **classificador Bayesiano** (Duda et al., 2001) usando as funções de densidade de probabilidade Gaussianas que compõe a GMM estimada com o algoritmo EM.

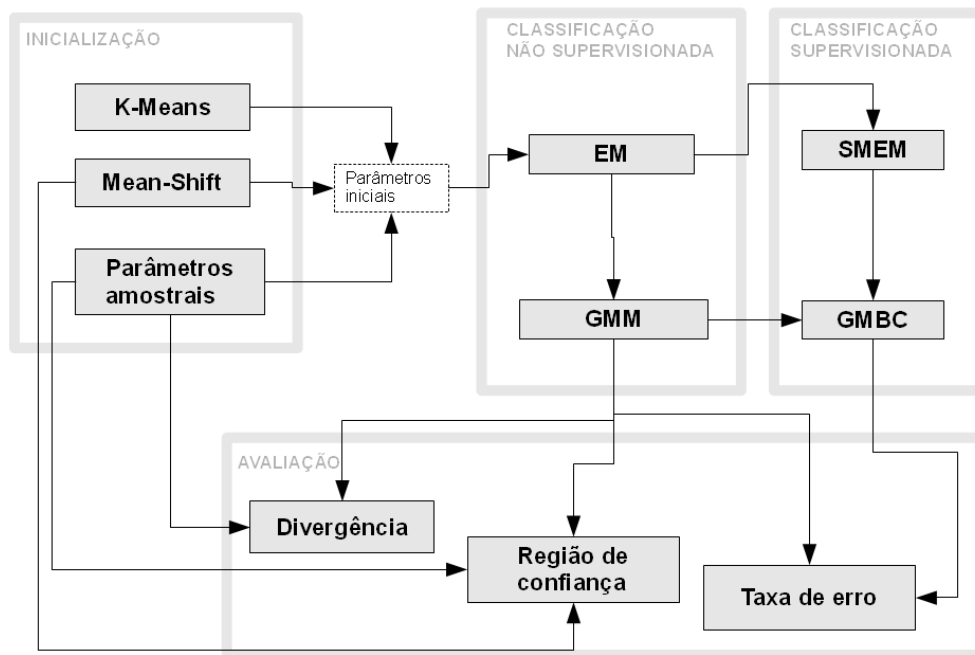


Figura 4.1: Resumo esquematizado dos algoritmos usados das neste trabalho.

Na abordagem Supervisionada estima-se os parâmetros do GMM que modela a distribuição dos vetores de características das impressões digitais de cada classe como a combinação linear de funções de densidade de probabilidade Gaussiana. Para encontrar os parâmetros do modelo usa-se o algoritmo SMEM (Ueda et al., 2000) que realiza iterativamente a separação e a junção de componentes, esta modificação permite superar o problema do máximo local da função verosimilhança; no entanto, neste trabalho com o objetivo de determinar a quantidade de funções de densidade de probabilidade que modelam a distribuição das observações de cada classe usou-se a modificação do SMEM introduzida em (Wagemaar, 2000). A classificação de impressões digitais com esta abordagem foi feita com o GMBC (Kim e Kang, 2007) que é uma estrutura semelhante à usada em (Yiu et al., 1999) chamada PDBNN (*Probabilistic decision-based neural network*).

Na Figura 4.1 mostra-se o resumo esquematizado dos algoritmos usados neste trabalho. No bloco Inicialização mostra-se as três formas de inicialização utilizadas no EM. No bloco Classificação Não-supervisionada tem-se o modelo GMM e o algoritmo de estimação EM, analogamente no bloco Classificação Supervisionada tem-se o GMBC e o algoritmo SMEM. No bloco Avaliação tem-se a Divergência de Kullback-Leibler (Kullback e Leibler, 1951), e o teste de hipótese com Região de Confiança (Jonhson e Wichern, 1998) que avaliam os resultados da estimação da média. Usou-se a taxa de erro de classificação para avaliar os resultados da classificação das impressões digitais.

## 4.2 Modelos de Misturas

Seja  $\mathcal{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$  o conjunto de  $N$  variáveis aleatórias, onde  $\mathbf{x} = [x_1, x_2, \dots, x_d]^T$  representa uma variável particular de  $\mathcal{X}$ . O modelo de misturas finito  $p(\mathbf{x}|\Theta)$  é um modelo probabilístico dada pela combinação linear de funções de densidade de probabilidades básicas  $p(\mathbf{x}|\theta_m)$  (Duda et al., 2001) e (Bishop, 2006), que pode ser formulado no espaço  $d$ -dimensional  $\mathbb{R}^d$  da seguinte forma:

$$p(\mathbf{x}|\Theta) = \sum_{m=1}^M \alpha_m p(\mathbf{x}|\theta_m), \quad (4.1)$$

onde  $M > 1$  é a quantidade de componentes do modelo,  $\Theta = \{\alpha_1, \alpha_2, \dots, \alpha_M, \theta_1, \theta_1, \dots, \theta_M\}$  é o conjunto de parâmetros do modelo,  $\alpha_m \geq 0$  é o coeficiente de ponderação de cada componente  $m = 1, \dots, M$  tal que  $\sum_{m=1}^M \alpha_m = 1$  e  $p(\mathbf{x}|\theta_m)$  é a função de densidade com parâmetros  $\theta_m$ . No caso dos modelos de misturas Gaussianas  $p(\mathbf{x}|\theta_m)$  é a distribuição normal com parâmetros  $\theta_m = \{\mu_m, \Sigma_m\}$  onde  $\mu_m$  é a média  $d$ -dimensional e  $\Sigma_m$  é a matriz de covariância  $d \times d$ -dimensional. Assim pode-se escrever:

$$p(\mathbf{x}|\Theta) = \sum_{m=1}^M \alpha_m \mathcal{N}(\mathbf{x}|\theta_m) = \sum_{m=1}^M \alpha_m \mathcal{N}(\mathbf{x}|\mu_m, \Sigma_m), \quad (4.2)$$

onde  $\mathcal{N}(\mathbf{x}|\mu_m, \Sigma_m)$  é a função de densidade de probabilidade Gaussiana  $d$ -dimensional dado por:

$$\mathcal{N}(\mathbf{x}|\mu_m, \Sigma_m) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_m|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_m)^T \Sigma_m^{-1} (\mathbf{x} - \mu_m)\right). \quad (4.3)$$

Na Figura 4.2(a) se observa um exemplo de mistura de  $M = 4$  gaussianas de duas dimensões  $\mathbf{x} = [x_1, x_2]^T$  e na Figura 4.2(b) observa-se as curvas de nível correspondentes.

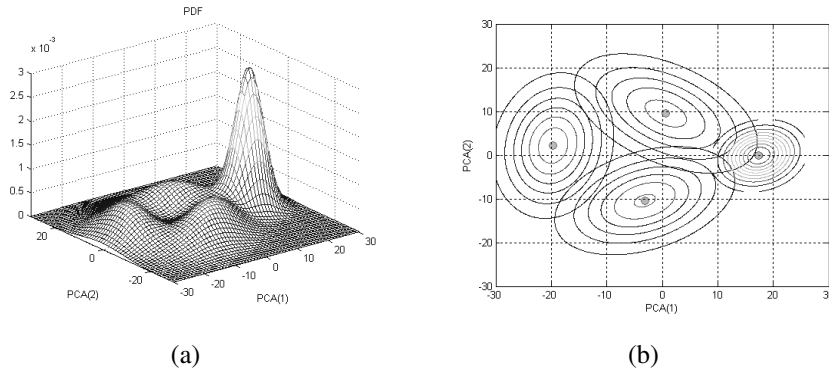


Figura 4.2: (a) Modelo de misturas gaussianas. (b) Curvas de nível.

### 4.3 Verossimilhança máxima

A função de verossimilhança  $\mathcal{L}(\Theta|\mathcal{X})$  é determinada a partir de dados coletados conjuntamente com uma função de densidade de probabilidade calculada por meio de inferência estatística. Diferentes parâmetros originam distintas funções de densidade de probabilidade, dessa forma, tem-se interesse em encontrar os parâmetros que correspondam a uma função de densidade de probabilidade desejada (Myung, 2003). É importante procurar um método eficiente de estimação de parâmetros para maximizar a função de verossimilhança.

Assim, a estimação da máxima verossimilhança (ML) consiste em estimar os parâmetros  $\Theta$  de uma função de densidade de probabilidade que se ajuste a um conjunto de observações  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ . Usa-se  $\hat{\Theta}$  para representar os parâmetros estimados. Assumindo que  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  são observações independentes identicamente distribuídas (i.i.d.), a função verossimilhança dos parâmetros  $\Theta$  com base nas observações é dada por:

$$p(\mathcal{X}|\Theta) = \prod_{n=1}^N p(\mathbf{x}_n|\Theta) \equiv \mathcal{L}(\Theta|\mathcal{X}). \quad (4.4)$$

$\mathcal{L}(\Theta|\mathcal{X})$  pode ser entendida como uma função dos parâmetros  $\Theta$ , onde  $\mathcal{X}$  é fixo. Já que no problema de maximização o objetivo é achar  $\hat{\Theta}$  que maximize a função de verossimilhança, tem-se:

$$\hat{\Theta} = \arg \max_{\Theta} \mathcal{L}(\Theta|\mathcal{X}). \quad (4.5)$$

Frequentemente, é conveniente maximizar o logaritmo da função de verossimilhança,  $\log(\mathcal{L}(\Theta|\mathcal{X}))$  ao invés do que a função em si, já que analiticamente pode-se resolver a expressão 4.4 mais facilmente assumindo-se a condição i.i.d.. Dessa forma, tem-se que para a  $m$ -ésima componente da mistura de  $N$  observações, sob as restrições  $\alpha_m \geq 0$  e  $\sum_{m=1}^M \alpha_m = 1$ ; o logaritmo da função verossimilhança, a partir de (4.1) e (4.4), é dado por:

$$\log(\mathcal{L}(\Theta|\mathcal{X})) = \log \prod_{n=1}^N p(\mathbf{x}_n|\Theta) = \sum_{n=1}^N \log \sum_{m=1}^M \alpha_m p(\mathbf{x}_n|\theta_m). \quad (4.6)$$

No caso em que  $p(\mathbf{x}|\Theta) = \mathcal{N}(\mathbf{x}|\mu, \sigma^2)$ , ou seja, uma única distribuição Gaussiana com  $\Theta = \{\mu, \sigma^2\}$ , é possível derivar  $\log(\mathcal{L}(\Theta|\mathcal{X}))$ , igualar a zero e resolver diretamente a equação para  $\mu$  e  $\sigma^2$ . Porém, há muitos problemas nos quais não é possível achar tais expressões analíticas, sendo necessário recorrer a técnicas estimativas (Duda et al., 2001).

## 4.4 Algoritmo EM

O algoritmo EM ou *Expectation-Maximization* é um método de estimação de parâmetros iterativo que busca maximizar a função verossimilhança a partir de um conjunto de  $N$  observações i.i.d.  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  (Ueda et al., 1998). Na estimação de parâmetros de modelos de misturas com algoritmo EM assume-se o número de componentes  $M$  conhecido.

O algoritmo EM é baseado na interpretação de que  $\mathcal{X}$  é um conjunto de dados incompletos; no caso de misturas finitas, os dados faltantes ou ocultos constituem um conjunto de  $M$  etiquetas  $\mathcal{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M\}$  associados com as  $N$  observações, se indicando qual componente produz cada observação, o que faz que cada  $\mathbf{z}_m$  seja um vetor  $N$ -dimensional binário da forma  $\mathbf{z}_m = [z_{1m}, z_{2m}, \dots, z_{Nm}]^T$ , onde:

$$z_{nm} = \begin{cases} 1 & n = p, \\ 0 & n \neq p. \end{cases}$$

Tendo em conta que o conjunto completo de dados vem dado por  $\mathcal{Y} = (\mathcal{X}, \mathcal{Z})$ , é possível representar o logaritmo da função de verossimilhança do conjunto de dados completo por:

$$\log p(\mathcal{Y}|\Theta) = \sum_{n=1}^N \sum_{m=1}^M z_{nm} \log[\alpha_m p(\mathbf{x}_n|\theta_m)]. \quad (4.7)$$

O algoritmo EM busca atingir um critério de convergência produzindo sequências de parâmetros estimados  $\hat{\Theta}^t$  para  $t = \{0, 1, 2, \dots\}$  iterações aplicando de maneira alternada os passos de *Expectation* e *Maximization*.

No passo **Expectation** calcula-se a expectância ou valor esperado condicional do logaritmo da função verossimilhança com respeito a  $\mathcal{X}$  e aos parâmetros estimados atuais  $\hat{\Theta}^t$ . Visto que  $\log p(\mathcal{X}, \mathcal{Z}|\Theta)$  é linear em relação ao conjunto de dados ocultos  $\mathcal{Z}$ , só tem-se que calcular a expectância condicional  $\mathcal{W} \equiv E[\mathcal{Z}|\mathcal{X}, \hat{\Theta}^t]$  e inseri-la em  $\log p(\mathcal{X}, \mathcal{Z}|\Theta)$ . O resultado é dado pela função  $\mathbf{Q}$  que se mostra na equação a seguir:

$$\mathbf{Q}(\Theta, \hat{\Theta}^t) = E[\log p(\mathcal{X}, \mathcal{Z}|\Theta)|\mathcal{X}, \hat{\Theta}^t] = \log p(\mathcal{X}, \mathcal{W}|\Theta). \quad (4.8)$$

Desde que os elementos de  $\mathcal{Z}$  sejam binários, a expectância condicional estaria dada por:

$$w_{nm} \equiv E[z_{nm}|\mathcal{X}, \hat{\Theta}^t] = P[z_{nm} = 1|x_n, \hat{\Theta}^t] = \frac{\hat{\alpha}_m^t p(\mathbf{x}_n|\hat{\theta}_m^t)}{\sum_{l=1}^M \hat{\alpha}_l^t p(\mathbf{x}_n|\hat{\theta}_l^t)}, \quad (4.9)$$

que pode ser interpretada como a regra de Bayes onde  $\hat{\alpha}_m^t$  seria a *probabilidade a priori*, enquanto  $w_{nm}$  seria a *probabilidade a posteriori* que  $z_{nm} = 1$  depois de se observar  $\mathbf{x}_n$ .

No passo *Maximization* considerando o critério de maximização da verossimilhança e as restrições  $\alpha_m \geq 0$  e  $\sum_{m=1}^M \alpha_m = 1$ , realiza-se a atualização dos parâmetros estimados no tempo  $t$  de acordo com:

$$\hat{\Theta}^{t+1} = \arg \max_{\Theta} \mathbf{Q}(\Theta, \hat{\Theta}^t). \quad (4.10)$$

No caso de misturas Gaussianas, o passo *maximization* consiste, tendo em conta que  $\hat{\theta}_m = \{\hat{\mu}_m, \hat{\Sigma}_m\}$ , em estimar  $\hat{\Theta} = \{\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_M, \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_M\}$ , para cada componente  $m$  da seguinte forma:

$$\hat{\alpha}_m^{t+1} = \frac{1}{N} \sum_{i=1}^N w_{nm}, \quad (4.11)$$

$$\hat{\mu}_m^{t+1} = \frac{\sum_{n=1}^N w_{nm} \mathbf{x}_n}{\sum_{n=1}^N w_{nm}} \quad \text{e} \quad (4.12)$$

$$\hat{\Sigma}_m^{t+1} = \frac{\sum_{n=1}^N w_{nm} (\mathbf{x}_n - \hat{\mu}_m^{t+1})(\mathbf{x}_n - \hat{\mu}_m^{t+1})^T}{\sum_{n=1}^N w_{nm}}. \quad (4.13)$$

É possível resumir o algoritmo EM da seguinte forma:

1. Escolhe-se uma inicialização  $\hat{\Theta}^{t=0}$  para  $\Theta$ .
2. Passo *Expectation*: Calcula-se a função  $\mathbf{Q}$ , resultado da expectância condicional, segundo a equação (4.8).
3. Passo *Maximization*: Procuram-se os novos parâmetros  $\hat{\Theta}^t$  que maximizem a função  $\mathbf{Q}$ , segundo a equação (4.10).
4. Se  $\hat{\Theta}^t$  satisfaz uma condição de convergência, o algoritmo para; caso contrário executa-se o passo 2 de novo.

No que se refere à convergência do algoritmo EM, na Figura 4.3 percebe-se que no passo *Expectation*, o algoritmo constrói uma função  $g_t$  que abaixa os limites da função objetivo  $\log p(\mathbf{x}|\theta)$  para o qual  $g_t(\hat{\theta}^t) = \log p(\mathcal{X}|\hat{\theta}^t)$ . No passo *Maximization*, o algoritmo desloca até um novo conjunto de parâmetros  $\hat{\theta}^{t+1}$  para maximizar  $g_t$ . Como o valor do limite abaixado  $g_t$  iguala-se com função objetivo em  $\hat{\theta}^t$ , segue-se que  $\log p(\mathcal{X}|\hat{\theta}^t) = g_t(\hat{\theta}^t) \leq g_t(\hat{\theta}^{t+1}) = \log p(\mathcal{X}|\hat{\theta}^{t+1})$ ; assim, a função objetivo cresce monotonicamente durante cada iteração. Porém, tal como acontece com a maioria dos métodos de otimização para funções que não são côncavas, o EM garante só a convergência para um máximo local da função objetivo.

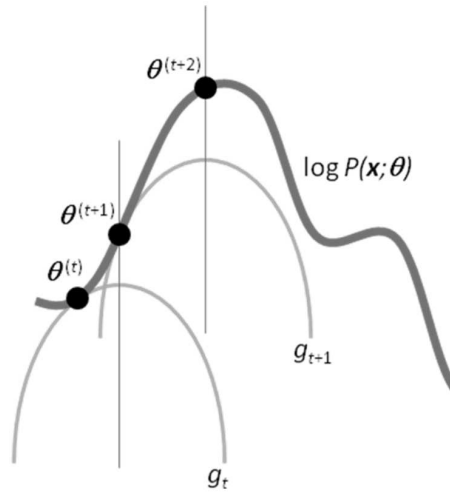


Figura 4.3: Convergência do algoritmo *Expectation-Maximization* (Chuong e Batzoglou, 2008).

## 4.5 Mean-Shift

O algoritmo *Mean-Shift* é uma técnica iterativa não-paramétrica de estimação de densidade que assume que a maior densidade dos pontos distribuídos no espaço de características corresponde à moda da distribuição de probabilidade. O algoritmo *Mean-Shift* foi proposto originalmente por (Fukunaga e Hostetler, 1975), sendo pesquisado depois por (Cheng, 1995); e tem sido empregado em áreas como visão computacional desde que se discutiram as características de convergência em (Comaniciu e Meer, 2002).

Suponha-se que o conjunto de observações  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  são amostras aleatórias de uma distribuição de densidade multivariada  $f$  no espaço  $d$ -dimensional  $\mathbb{R}^d$ . Define-se o *kernel*  $k$  multivariável de largura de banda  $h$  fixa estimada sobre o ponto  $\mathbf{x}$  como:

$$\hat{f}_{h,k}(\mathbf{x}) = \frac{c_{k,d}}{Nh^d} \sum_{n=1}^N k \frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{h^2}. \quad (4.14)$$

O termo  $c_{k,d}$  representa o coeficiente de normalização; então o gradiente da equação 4.14, depois de manipulações algébricas pode-se expressar por:

$$\hat{\nabla} f_{h,k}(\mathbf{x}) = \frac{2c_{k,d}}{Nh^d} \left[ \sum_{n=1}^n g \left( \frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{h^2} \right) \right] m_h(\mathbf{x}), \quad (4.15)$$

onde  $g(\cdot) = k'(\cdot)$  é proporcional ao gradiente da densidade normalizada e  $m_h(x)$  é o vetor *Mean-Shift* dado por:

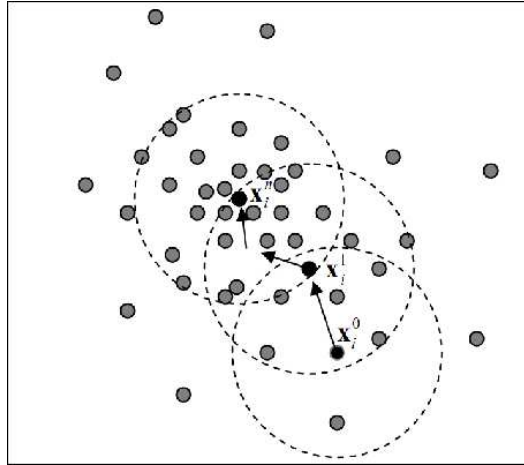


Figura 4.4: Busca de moda com *Mean-Shift* (Derpanis, 2005).

$$m_h(\mathbf{x}) = \frac{\sum_{n=1}^N g\left(\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{h^2}\right)\mathbf{x}_n}{\sum_{n=1}^N g\left(\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{h^2}\right)} - \mathbf{x}. \quad (4.16)$$

É garantido que o vetor *Mean-Shift* aponta em direção à região de maior densidade de pontos. Portanto, a partir de uma observação  $\mathbf{x}_i$  no espaço  $d$ -dimensional, *Mean-Shift* produz iterativamente uma trajetória que converge para um ponto estacionário que representa a moda de todo o espaço. Para a prova de convergência e mais detalhes do algoritmo veja (Comaniciu e Meer, 2002), (Fukunaga e Hostetler, 1975) e (Cheng, 1995).

A Figura 4.4 mostra a busca da moda usando-se a técnica de *Mean-Shift* a partir do ponto  $\mathbf{x}_i$  até achar os pontos estacionários da função de densidade. O índice superior denota as iterações e os círculos pontilhados denotam as janelas de densidade definidas pela largura de banda  $h$ .

## 4.6 Algoritmo SMEM

O algoritmo SMEM ou *Split and Merge Expectation Maximization* é um método de busca de máxima verosimilhança baseado no EM que tenta superar o problema do máximo local na estimação de parâmetros de modelos de misturas juntando componentes em regiões superpovoadas e dividindo componentes em regiões subpovoadas no espaço da distribuição.

A ideia de fazer operações de *Split* (divisão) e *Merge* (junção) tem sido aplicada com sucesso no algoritmo EM de estimação paramétrica do modelo de misturas gaussianas (Ueda et al., 2000) e (Ueda et al., 1998). No caso geral, na estimação de parâmetros de modelos



de misturas, a busca do máximo local usando o algoritmo EM tem o problema da escolha do número de componentes, já que no processo de estimação de parâmetros pode-se ter muitas componentes concentradas em algumas regiões e poucas componentes ou nenhuma em outras regiões. Assim, é possível evitar máximos locais através da introdução de uma operação de **junção**, que mescla elementos em regiões compostas por muitas componentes altamente semelhantes e de uma operação de **divisão** que divide os componentes em aquelas regiões cujos elementos não-semelhantes foram combinados numa única componente.

Visto que  $\hat{\Theta}$  denota os parâmetros estimados pelo algoritmo EM, depois da convergência do algoritmo, pode-se re-escrever a função  $Q$  de (4.8) na forma:

$$\hat{Q} = \hat{Q}_i + \hat{Q}_j + \hat{Q}_k + \sum_{m, m \neq i, j, k} \hat{Q}_m, \quad (4.17)$$

onde  $\hat{Q}_m = \sum_{n=1}^N P(m|\mathbf{x}_n; \hat{\Theta}) \log \hat{\alpha}_m P(\mathbf{x}_n | \hat{\theta}_m)$ .

A expressão do lado direito da equação (4.17) pode ser incrementada em três termos juntando os modelos  $i$ -ésimo e  $j$ -ésimo para produzir o novo modelo  $i'$ -ésimo e dividindo o modelo  $k$ -ésimo nos  $j'$ -ésimo e  $k'$ -ésimo modelo.

## Inicialização

Os valores iniciais dos parâmetros  $\hat{\Theta}'_i$  para realizar a junção de dois componentes (passo *Merge*) são dados pela combinação linear dos modelos originais  $\hat{\Theta}_i$  e  $\hat{\Theta}_j$ , como é indicado na seguinte expressão,

$$\alpha_{i'} = \hat{\alpha}_i + \hat{\alpha}_j \quad \text{e} \quad \theta_{i'} = \frac{\hat{\alpha}_i \hat{\theta}_i + \hat{\alpha}_j \hat{\theta}_j}{\hat{\alpha}_i + \hat{\alpha}_j}, \quad (4.18)$$

onde o coeficiente de ponderação  $\hat{\alpha}_l = \frac{1}{N} \sum_{n=1}^N P(l|\mathbf{x}_n, \hat{\Theta})$  é estimado como a média da probabilidade *a posteriori* segundo as equações (4.11) e (4.9). Percebe-se que a inicialização de  $\theta_{i'}$  é a combinação linear de  $\hat{\theta}_i$  e  $\hat{\theta}_j$  proporcionais aos coeficientes de ponderação  $\hat{\alpha}_l$ .

Os valores iniciais dos parâmetros para a divisão de uma componente  $\hat{\Theta}_k$  em duas  $\hat{\Theta}'_j$  e  $\hat{\Theta}'_{k'}$  (passo *Split*) podem ser definidos como:

$$\alpha_{j'} = \alpha_{k'} = \frac{\hat{\alpha}_k}{2}, \quad \theta_{j'} = \hat{\theta}_k + \varepsilon \quad \text{e} \quad \theta_{k'} = \hat{\theta}_k + \varepsilon', \quad (4.19)$$

onde  $\varepsilon$  e  $\varepsilon'$  são pequenas perturbações aleatórias tal que ( $\|\varepsilon\| \ll \|\hat{\theta}_k\|$ ).

No caso de GMM, os parâmetros estão dados por  $\hat{\theta}_i = \{\hat{\mu}_i, \hat{\Sigma}_i\}$  e  $\alpha_i$ . Para o passo *Split*,  $\Sigma_{j'}$  e  $\Sigma_{k'}$  devem ser inicializadas como positivas definidas tal que  $\Sigma_{j'} = \Sigma_{k'} = |\hat{\Sigma}_k|^{1/d} I_d$ , onde  $I_d$  é

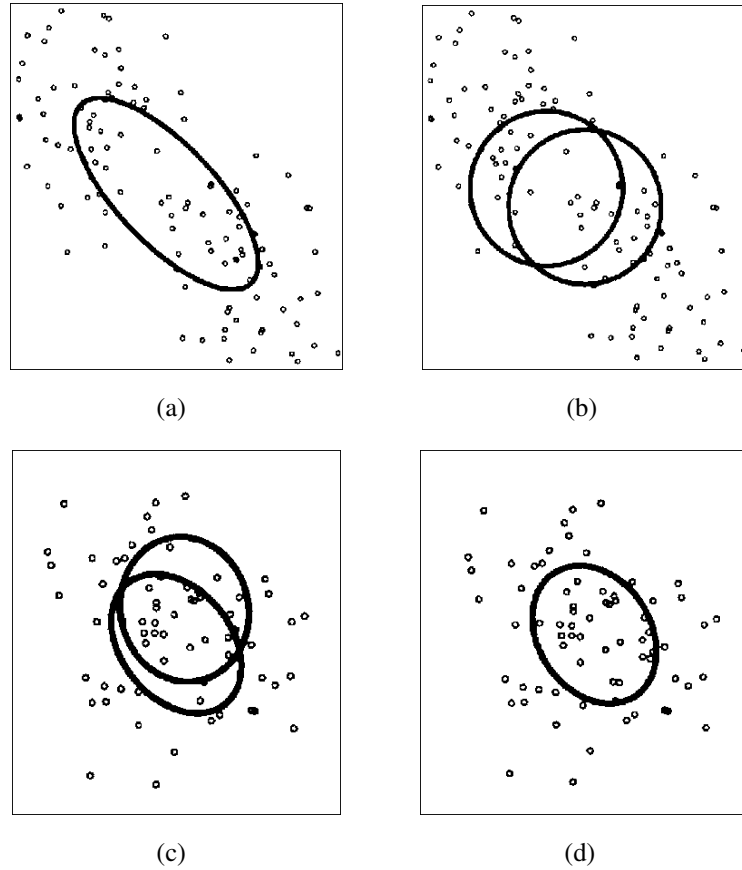


Figura 4.5: Inicialização para o caso de distribuição gaussiana bidimensional (a) e (b) Passo *Split*. (c) e (d) Passo *Merge*.

a matriz identidade  $d \times d$ -dimensional. No caso dos vetores  $\mu_{j'}$  e  $\mu_{k'}$  podem ser inicializados usando o algoritmo *K-means* ou simplesmente adicionado-se perturbações aleatórias  $\varepsilon_m$ , para  $m = 1, 2$  ( $\|\varepsilon_m\| \ll \|\hat{\mu}_k\|$ ) da forma  $\mu_{j'} = \hat{\mu}_k + \varepsilon_1$  and  $\mu_{k'} = \hat{\mu}_k + \varepsilon_2$ .

A Figura 4.5 mostra um exemplo simples dos passos de inicialização para o caso de distribuições gaussianas bidimensionais usando as equações (4.17), (4.18) e (4.19)

## EM parcial

A re-estimação de  $m' = i', j'$  e  $k'$  é realizada pelo algoritmo EM descrito na seção 4.4, mas substituindo a equação (4.9) por:

$$P(m'|\mathbf{x}, \Theta^{(t)}) = \frac{\alpha_m^{(t)} p_m(\mathbf{x}|\theta_m^{(t)})}{\sum_{l=i',j',k'} \alpha_l^{(t)} p_l(\mathbf{x}|\theta_l^{(t)})} \times \sum_{m=i,j,k} P(m|\mathbf{x}, \hat{\Theta}), \quad (4.20)$$

para  $m = i', j', k'$ . Usando (4.20), a soma das probabilidades *a posteriori* dos modelos  $i', j'$  e  $k'$  torna-se igual à soma das probabilidades *a posteriori*  $i, j$  e  $k$  logo antes de realizar os

passos *Split* e *Merge*; isto é

$$\sum_{m'=i',j',k'} P(m'|\mathbf{x}, \Theta^{(t)}) = \sum_{m=i,j,k} P(m|\mathbf{x}, \hat{\Theta}), \quad (4.21)$$

que é realizado durante todo o processo de re-estimação. Devido a isto é possível re-estimar os parâmetros dos modelos  $i'$ ,  $j'$  e  $k'$  sem afetar os outros modelos. O procedimento é chamado de **EM parcial** e faz com que o algoritmo total seja eficiente. Em seguida efetua-se o algoritmo EM descrito na secção 4.4 como pós processamento sendo este chamado de **EM completo**.

Depois de executar o EM parcial e o EM completo, se  $\mathbf{Q}$  melhorar, admite-se a nova estimativa e repete-se os passos descritos acima após definir o novo parâmetro  $\hat{\Theta}$ ; caso contrário, repete-se o procedimento com outros candidatos considerando o  $\hat{\Theta}$  anterior.

### O algoritmo *Split and Merge*

Os procedimentos descritos nas secções acima podem ser resumidos da seguinte forma:

1. Executa-se o algoritmo EM da secção 4.4. Seja  $\hat{\Theta}$  os parâmetros estimados e  $\hat{\mathbf{Q}}$  a estimativa da função  $\mathbf{Q}$  os resultados quando algoritmo tem convergido;
2. Escolhem-se e ordenam-se os candidatos para o *Split* e para o *Merge* seguindo o critério baseado em  $\hat{\Theta}$  (Ueda et al., 1998). Denota-se o  $c$ -ésimo candidato como  $\{i, j, k\}_c$ ;
3. Para  $c = 1, 2, \dots, C_{max}$ ; após inicializar os parâmetros baseado em  $\hat{\Theta}$ , executa-se o EM parcial para  $\{i, j, k\}_c$ , seguido do EM total até a convergência. Seja  $\hat{\Theta}^*$  a representação dos parâmetros obtidos e  $\hat{\mathbf{Q}}^*$  a função  $\mathbf{Q}$  depois do EM total; se  $\hat{\mathbf{Q}}^* > \hat{\mathbf{Q}}$ , então define-se  $\hat{\mathbf{Q}} \leftarrow \hat{\mathbf{Q}}^*$  e  $\hat{\Theta} \leftarrow \hat{\Theta}^*$  e vá-se para o passo 2, caso contrário  $\hat{\mathbf{Q}}^* \leq \hat{\mathbf{Q}}$ . vá-se ao passo 4.
4. Para-se o processo iterativo com  $\hat{\Theta}$  como parâmetros finais.

O algoritmo SMEM incrementa monotonicamente o valor da função  $\mathbf{Q}$ , no caso que o valor desta função não aumentar para todos os  $c = 1, 2, \dots, C_{max}$  o algoritmo para. Visto que os passos do EM total são equivalentes a executar o algoritmo EM original depois da convergência do EM parcial, o algoritmo *Split and Merge* mantém as propriedades de convergência global do algoritmo EM.

## 4.7 GMBC

O GMBC (*Gaussian mixture-based classification*) emprega uma estrutura de rede modular OCON (*one class one net*) na qual cada GMM que modela uma classe diferente é representada por uma sub-rede (Kim e Kang, 2007). Esta estrutura é semelhante à empregada por uma PDBNN (Probabilistic decision-based neural networks) mostrada em (Yiu et al., 1999). Para configurar uma estrutura GMBC (veja Figura 4.6) precisa-se conhecer o número de classes, a quantidade de componentes (funções de densidade de probabilidade Gaussiana) de cada classe e os parâmetros estimados com o algoritmo SMEM.

Assim, cada classe  $\omega_i$  terá associado um conjunto de observações  $\mathbf{x}_i$  e uma distribuição de misturas gaussianas dada por:

$$p(\mathbf{x}|\omega_i) = \sum_{m=1}^M P(\Theta_{m|i}|\omega_i)p(\mathbf{x}|\omega_i, \Theta_{m|i}), \quad (4.22)$$

onde  $\Theta_{r|i}$  representa os parâmetros da  $m$ -ésima componente da mistura,  $M$  é o número de componentes da mistura,  $p(\mathbf{x}|\omega_i, \Theta_{m|i}) \equiv \mathcal{N}(\mathbf{x}|\mu_{m|i}, \Sigma_{m|i})$  é a função densidade de gaussiana da  $m$ -ésima componente e  $P(\Theta_{m|i}|\omega_i)$  é a probabilidade *a priori* da  $i$ -ésima classe. A função discriminante da classe  $\omega_i$  tal que  $i = 1, 2, \dots, L$ , é projetada para modelar algumas funções

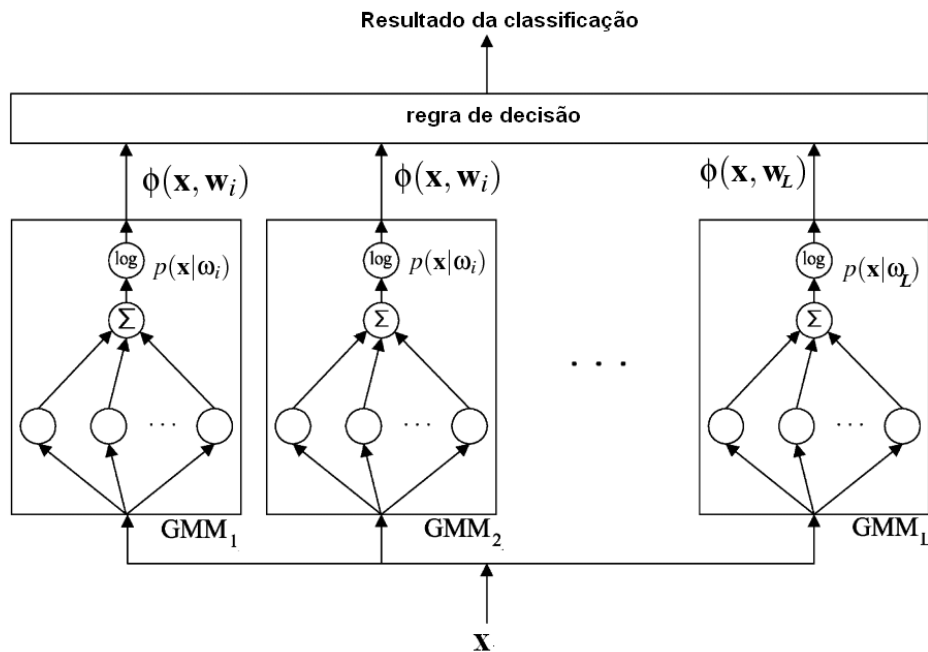


Figura 4.6: Estrutura do classificador GMBC com OCON. (Kim e Kang, 2007)

de verossimilhança da forma:

$$\begin{aligned}\phi(\mathbf{x}, \mathbf{w}_i) &= \log(p(\mathbf{x})|\omega_i) \\ &= \log \left[ \sum_{m=1}^M P(\Theta_{m|i}|\omega_i) p(x_i(t)|\omega_i, \Theta_{m|i}) \right],\end{aligned}\quad (4.23)$$

onde  $\mathbf{w}_i = \{\mu_{m|i}, \Sigma_{m|i}, P(\Theta_{m|i}|\omega_i)\}$ . A sub-rede com a máxima verossimilhança estimada é reivindicada como o vencedor. A eficiência do GMBC pode ser melhorada quando se encontra o número apropriado de componentes que descreva cada classe.

## 4.8 Divergência Kullback-Leibler

Provavelmente divergência Kullback-Leibler (KLD) é a medida de informação usado com mais frequência desde o ponto de vista teórico. KLD ou entropia mútua permite medir a divergência entre funções de densidade de probabilidade que têm parâmetros distintos.

Desde que  $p(\mathbf{x}|\theta_1)$  e  $q(\mathbf{x}|\theta_2)$  são duas funções de densidade de probabilidade com parâmetros  $\theta_1$  e  $\theta_2$  respectivamente, a divergência KLD é definida como:

$$\begin{aligned}\mathcal{D}(p(\mathbf{x}|\theta_1)||q(\mathbf{x}|\theta_2)) &= E_p \left[ \log \left( \frac{p(\mathbf{x}|\theta_1)}{q(\mathbf{x}|\theta_2)} \right) \right] \\ &= \int p(\mathbf{x}|\theta_1) \log \left( \frac{p(\mathbf{x}|\theta_1)}{q(\mathbf{x}|\theta_2)} \right) d\mathbf{x},\end{aligned}\quad (4.24)$$

onde  $E_p[\cdot]$  é o valor esperado com respeito à densidade de probabilidade  $p(\mathbf{x})$ .

A divergência KLD não é simétrica, mas, para se obter uma medição simétrica com respeito as duas densidades de probabilidade  $p(\mathbf{x})$  e  $q(\mathbf{x})$  aplica-se a divergência de Jefferys-Kullback-Leibler ou J-divergência introduzida em (Jeffreys, 1946), já que o resultado é igual à média da divergências KLD das distribuições  $p(\mathbf{x})$  e  $q(\mathbf{x})$ , assim:

$$\begin{aligned}\mathcal{J}(p(\mathbf{x}|\theta_1)||q(\mathbf{x}|\theta_2)) &= \mathcal{J}(p(\mathbf{x}|\theta_2)||q(\mathbf{x}|\theta_1)). \\ &= \frac{\mathcal{D}(p(\mathbf{x}|\theta_1)||q(\mathbf{x}|\theta_2)) + \mathcal{D}(p(\mathbf{x}|\theta_2)||q(\mathbf{x}|\theta_1))}{2}.\end{aligned}\quad (4.25)$$

Do ponto de vista teórico KLD mede quanto eficiente em media seria estimar os parâmetros de uma função de densidade de probabilidade a partir do conjunto de mostras. No caso da distribuição Gaussiana, segundo a equação (4.24) tem-se:

$$\begin{aligned} \mathcal{D}(\mathcal{N}(\mathbf{x}|\mu_1, \Sigma_1) || \mathcal{N}(\mathbf{x}|\mu_2, \Sigma_2)) &= E_{\mu_1, \Sigma_1} \left[ \log \left( \frac{\mathcal{N}(\mathbf{x}|\Sigma_1, \mu_1)}{\mathcal{N}(\mathbf{x}|\Sigma_2, \mu_2)} \right) \right] \\ &= \int \mathcal{N}(\mathbf{x}|\Sigma_1, \mu_1) \left[ \log \left( \frac{\mathcal{N}(\mathbf{x}|\Sigma_1, \mu_1)}{\mathcal{N}(\mathbf{x}|\Sigma_2, \mu_2)} \right) \right] d\mathbf{x}. \end{aligned} \quad (4.26)$$

Integrando a equação (4.26) e considerando que as duas funções de densidade de probabilidade Gaussianas estão no espaço  $d$ -dimensional (Tumminello et al., 2007), tem-se que a distância entre elas é dada por:

$$\mathcal{D}(\mathcal{N}(\mathbf{x}|\mu_1, \Sigma_1) || \mathcal{N}(\mathbf{x}|\mu_2, \Sigma_2)) = \frac{1}{2} \left[ \log \left( \frac{|\Sigma_2|}{|\Sigma_1|} \right) + \text{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2 - \mu_1)\Sigma_2^{-1}(\mu_2 - \mu_1) - d \right]. \quad (4.27)$$

## 4.9 Região de Confiança

Em estatística, a região de confiança é a generalização multidimensional do intervalo de confiança. Dessa forma, a região de confiança é um conjunto de pontos num espaço  $d$ -dimensional, representada por uma hiper-elipsoide ao redor do ponto que é uma solução prevista para um problema. Seja  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  um conjunto de  $N$  observações onde  $\mathbf{x} = [x_1, x_2, \dots, x_d]^T$  representa uma observação particular de  $\mathcal{X}$ ; define-se a região de confiança  $100(1 - \alpha)\%$  de uma distribuição normal  $d$ -dimensional como a hiper-elipsoide determinada por todas as  $\mu$  tal que:

$$N(\bar{\mathbf{x}} - \mu)^T \mathbf{S}^{-1}(\bar{\mathbf{x}} - \mu) \leq \frac{d(N-1)}{N-d} F_{d, N-d}(\alpha), \quad (4.28)$$

onde  $N$  é o número de observações,  $\bar{\mathbf{x}}$  é a média amostral e  $S$  é a matriz de covariância amostral e são dados por:

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n, \quad \text{e} \quad \mathbf{S} = \frac{1}{N-1} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T, \quad (4.29)$$

$F_{d, N-d}$  denota a função de densidade acumulativa de uma variável aleatória com distribuição de Fisher (ou *F-distribution*) com  $d$  e  $N-d$  graus de liberdade e nível de significância  $\alpha$ . A expressão  $N(\bar{\mathbf{x}} - \mu)^T \mathbf{S}^{-1}(\bar{\mathbf{x}} - \mu)$  é chamada de distância de Hotelling ou distância  $T^2$  e mede a distância entre  $\mu$  e  $\bar{\mathbf{x}}$ . O teste estatístico de Hotelling é utilizado quando tem-se o caso em que o vetor da média é conhecida e a matriz de covariância não é conhecida (Kanungo e Haralick, 1995). Para determinar se alguma  $\mu_0$  cai na região de confiança, tem-se que comparar a distância  $T^2$  com a expressão do lado direito da equação (4.28). Se  $T^2$  é maior diz-se

que  $\mu_0$  não fica na região de confiança. Isto é análogo a fazer o teste de hipótese com nível de significância  $\alpha$ . Para mais detalhes veja (Jonhson e Wichern, 1998), e (Chew, 1966).

A hiper-elipsoide (ou elipse) de confiança é dada pelas orientações e os comprimentos dos eixos determinados a partir dos auto-valores ( $\lambda_i$ ) e auto-vetores ( $\mathbf{e}_i$ ) da matriz de covariância amostral ( $\mathbf{S}$ ). Assim, tomando como centro  $\bar{\mathbf{x}}$ , os eixos são:

$$\pm \sqrt{\lambda_i} \sqrt{\frac{d(N-1)}{N(N-d)} F_{d,N-d}(\alpha)} \mathbf{e}_i, \quad (4.30)$$

onde  $\mathbf{S}\mathbf{e}_i = \lambda_i \mathbf{e}_i$ , para  $i = 1, 2, \dots, d$ . No caso bidimensional, as proporções entre os  $\lambda_i$ 's ajudam a identificar as quantidades relativas de alongamento ao longo dos pares de eixos. A Figura 4.7 mostra a elipse de confiança para este caso. A inferência sobre  $\mu$  quando há grande quantidade de amostras  $N \gg d$  é baseada na distribuição  $\chi^2$ , portanto tem-se:

$$\chi_d^2(\alpha) \approx \frac{d(N-1)}{N-d} F_{d,N-d}(\alpha), \quad (4.31)$$

por conseguinte, substituindo (4.31) na equação (4.30), os eixos da elipse de confiança podem ser calculados com a seguinte expressão:

$$\pm \sqrt{\lambda_i} \sqrt{\frac{\chi_d^2(\alpha)}{N}} \mathbf{e}_i. \quad (4.32)$$

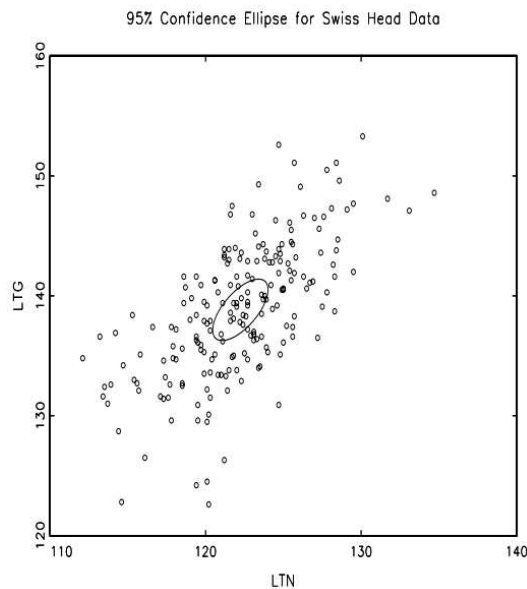


Figura 4.7: Elipse de confiança para um nível de significância de 95%

# Capítulo 5

## Avaliação dos parâmetros estimados

Neste trabalho foram propostas duas abordagens (Supervisionado e Não-Supervisionado) para estimar os parâmetros das funções de densidade de probabilidade Gaussianas a fim de modelar a distribuição dos vetores de características das imagens das impressões digitais. Na abordagem Não-Supervisionado propõe-se modelar a distribuição dos vetores de característica de cada classe mediante uma componente (função de densidade de probabilidade Gaussiana) do GMM; pelo contrário, na abordagem Supervisionado propõe-se modelar as distribuições de cada classe mediante a combinação linear de um conjunto de Gaussianas (dado por um GMM em si), cujos parâmetros são estimados via o SMEM.

Neste capítulo, são apresentados os resultados da avaliação dos parâmetros das funções de densidade de probabilidade Gaussianas estimadas via a abordagem Não-supervisionado versus os parâmetros amostrais das distribuições das impressões digitais para cada classe. Já nos dois capítulos sub-seguintes serão apresentados os resultados da classificação sob a abordagem Não-Supervisionada (Capítulo 6) e sob a abordagem Supervisionada (Capítulo 7).

Primeiramente avaliou-se o teste de hipótese, com respeito á média amostral das distribuições dos vetores de características das impressões digitais de cada classe, da média estimada via o algoritmo EM. Assim mesmo, incluiu-se também o teste do centroide com o algoritmo *K-means* e da moda estimada com o algoritmo *Mean-Shift*. A avaliação foi realizada para vetores de características com distinto número de coeficientes PCA.

Complementou-se a análise dos parâmetros estimados com a comparação, via a entropia mútua, da função de densidade de probabilidade Gaussiana de média e covariância amostral frente a função de densidade de probabilidade Gaussiana de média e covariância estimada via o algoritmo EM.



## 5.1 Avaliação da média via teste de hipótese

Na Figura 5.1 mostra-se a distribuição dos vetores de características não rotuladas das impressões digitais de todas as classes de treinamento do banco de dados DB4 para o **problema de classificação de 4 classes** quando foram considerados apenas os dois primeiros coeficientes PCA ( $d = 2$ ); também são amostrados as médias amostrais de cada classe juntamente com a elipse de confiança, com nível de confiança  $\alpha = 0.01$ , desenhado se considerando a distribuição  $\chi^2$  (Equação 4.31). Assim mesmo são amostradas as médias, centroides e modas estimadas via os algoritmos de estimação EM (etiquetado na figura como GMM), *K-means* e *Mean-Shift* respectivamente.

No caso da estimação da média usando o algoritmo EM, devido à dispersão dos resultados obtidos quando utilizados valores aleatórios na inicialização dos parâmetros, propus-se inicializar os parâmetros usando o vetor de média e a matriz de covariância amostrais da distribuição dos vetores de característica de cada classe com o objetivo de avaliar se a média resultante da estimação Não-Supervisionada, ainda que a inicialização não fosse aleatória, poderia ser considerada como média da distribuição de tais classes com um nível de confiança  $\alpha = 0.01$ . Assim, no plano bidimensional da Figura 5.1, gerado pelos dois primeiros

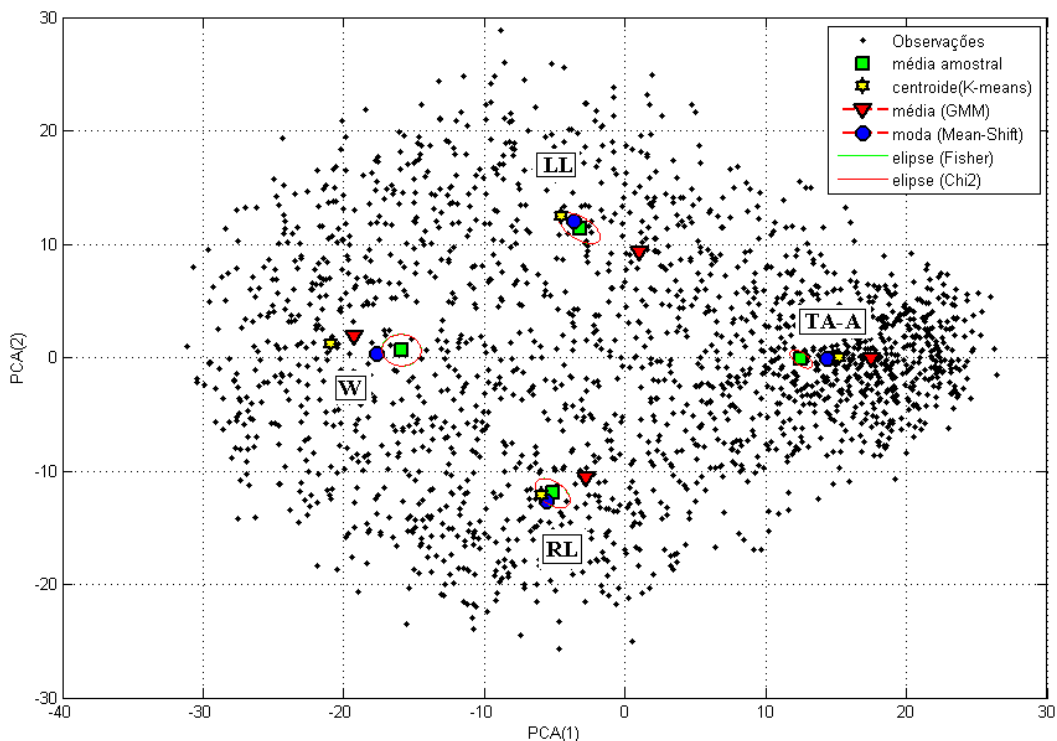


Figura 5.1: Distribuição dos vetores de características, médias amostrais, centroide estimado com *K-means*, média estimada com EM, moda estimadas com *Mean-Shift* e elipses de confiança  $F$  e  $\chi^2$  correspondentes a cada classe para o problema de 4 classes.

coeficientes PCA (note-se que a dimensão  $d$  do espaço de características é dado pelo número de coeficientes PCA do vetor de características) percebeu-se que a média estimada via EM ficou fora da elipse de confiança da média amostral, por tanto, poder-se-ia concluir que com dois coeficientes PCA no vetor de características a função de densidade de probabilidade Gaussiana bidimensional estimada via o EM não pode ser considerado como modelo da distribuição das impressões digitais de nenhuma das 4 classes.

No caso do *K-means*, a estimação Não-Supervisionado (com pontos iniciais dados pelas médias amostrais de cada classe) fez com que os centroides estimados das classes W e TA-A ficassem fora da elipse de confiança da média amostral e só os centroides correspondentes às classes LL e RL ficassem dentro da elipse. Já no caso do *Mean-Shift*, a estimação das modas da distribuição dos vetores de característica de cada classe foi feita a partir pontos iniciais aleatórios do espaço bidimensional; porém, percebeu-se que apesar deste tipo de inicialização a moda estimada na classe LL ficou dentro da elipse de confiança. Este resultado é importante para a distribuição desta classe, desde que o ponto inicial foi escolhido aleatoriamente ainda que nas outras classes as modas estimadas ficaram fora da elipse de confiança.

Os resultados da estimação de parâmetros obtidos a partir de dois coeficientes PCA não foram os esperados; porém, seria possível obter um resultado distinto com um número maior de elementos no vetor de características. Assim, sugere-se incrementar o número de coeficientes PCA no vetor de características e avaliar, ainda no espaço bidimensional por médio de projeções, a localização da média estimada com respeito à média amostral. Nas Figuras 5.2 e 5.3 são amostrados os resultados obtidos quando foram considerados  $d = \{3, 6, 10, 38, 61, 64\}$  coeficientes PCA.

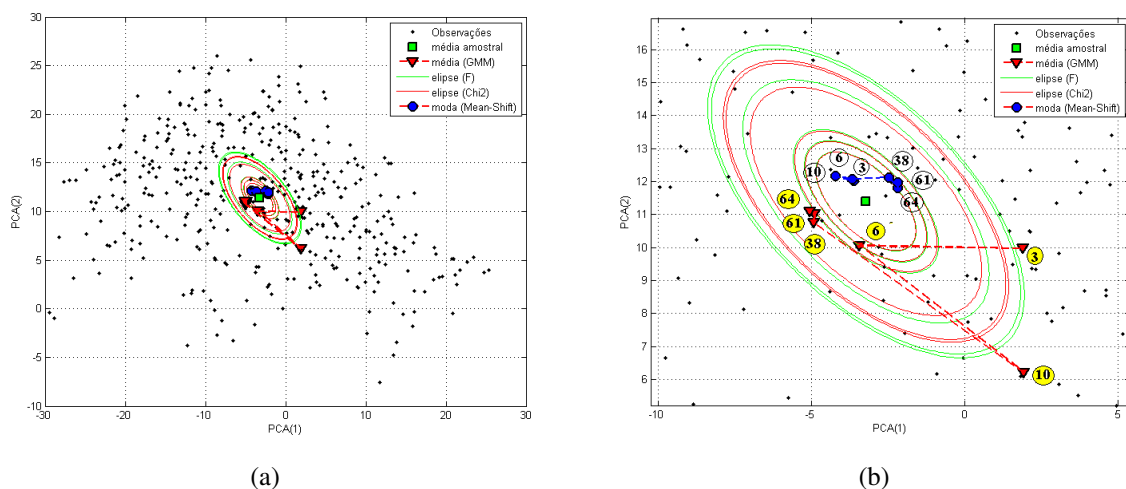


Figura 5.2: (a) Projeção bidimensional da distribuição de vetores de características da classe LL para o problema de 4 classes. (b) Elipses de confiança da média amostral (F e  $\chi^2$ ), moda estimada com *Mean-Shift* e média estimada com EM para  $d = \{3, 6, 10, 38, 61, 64\}$ .

Na Figura 5.2(a) são mostrados os resultados correspondentes aos vetores de características das imagens das impressões digitais da classe LL e na Figura 5.2(b) se observa com mais detalhe o entorno da elipse de confiança da média amostral. Tendo em conta que, tanto os comprimentos quanto as direções dos eixos da elipse de confiança, segundo a Equação (4.28) estão em relação com graus de liberdade  $d$  e  $N - d$  (onde  $N$  é o número de elementos da distribuição), pode-se ver que as elipses concêntricas dependem do número de coeficientes PCA e do número de elementos da distribuição; dessa forma na Figura 5.2(b), a elipse que fica na parte interna foi desenhada quando  $d = 3$ , a elipse imediatamente maior corresponde a  $d = 6$  e dessa forma a elipse localizada na parte externa corresponde a  $d = 64$ , pois os comprimentos dos eixos vão ficando maiores de acordo com o aumento dos elementos do vetor. As elipses verdes correspondem à distribuição  $F$  e as vermelhas correspondem às regiões de confiança desenhadas com  $\chi^2$ .

Em relação à média estimada com EM pode-se observar que quando o vetor de características teve  $d = 3$  coeficientes PCA, a média estimada ficou fora da elipse de confiança e quando  $d = 6$  a média estimada também ficou fora, mas, se aproximou da média amostral. Contudo, quando  $d = 10$  a média estimada afastou-se da média amostral. Na situação em que  $d = 38$  a média estimada voltou se aproximar da média amostral, neste caso ela já localizou-se dentro elipse de confiança correspondente (a quarta a partir do centro). Veja que quando  $d = [38, 61, 64]$  as médias estimadas para cada caso quase não sofreram variação, deste modo as três ficaram dentro das elipses de confiança correspondente. Assim sendo, salvo o caso de  $n = 10$ , a média estimada tendeu se aproximar mais da média amostral na medida que o número de coeficientes PCA aumentou. No referente às modas estimadas a partir do algoritmo *Mean-Shift*, observa-se que, para todos os casos de  $d$ , elas ficaram dentro da elipse de confiança.

Na Figura 5.3 são mostrados os parâmetros de média e moda estimados correspondentes às imagens das impressões digitais das classes W, RL e TA-A. Nos três casos mostrados nas Figuras 5.3(c) e (d), 5.3(c) e (d) e 5.3(c) e (d) respectivamente, da mesma forma do que no caso da classe LL, a média estimada tendeu-se aproximar da média amostral conforme a número de coeficientes PCA aumentou salvo o caso  $d = 10$ .

No que respeita à moda estimada com *Mean-Shift*, no caso da classe W (Figuras 5.3 (a)-(b)) e a classe RL, (Figuras 5.3 (c)-(d)) pode-se ver que a moda tendeu a se afastar da média amostral conforme o valor de  $d$  aumentou. No caso da classe TA-A (Figuras 5.3 (e)-(f)), observa-se que a moda ficou fora da região de confiança em todos os casos e tendeu se afastar da média amostral. Em todos os casos a moda estimada tendeu se afastar da média amostral quando foram incrementados os coeficientes no vetor de características enquanto a media estimada com o EM aproximou-se, porém, tendo em conta que as modas foram estimadas a partir de valores aleatórios, seria um boa alternativa inicializar o algoritmo EM com as modas estimadas ao invés do que inicializar com os parâmetros amostrais.

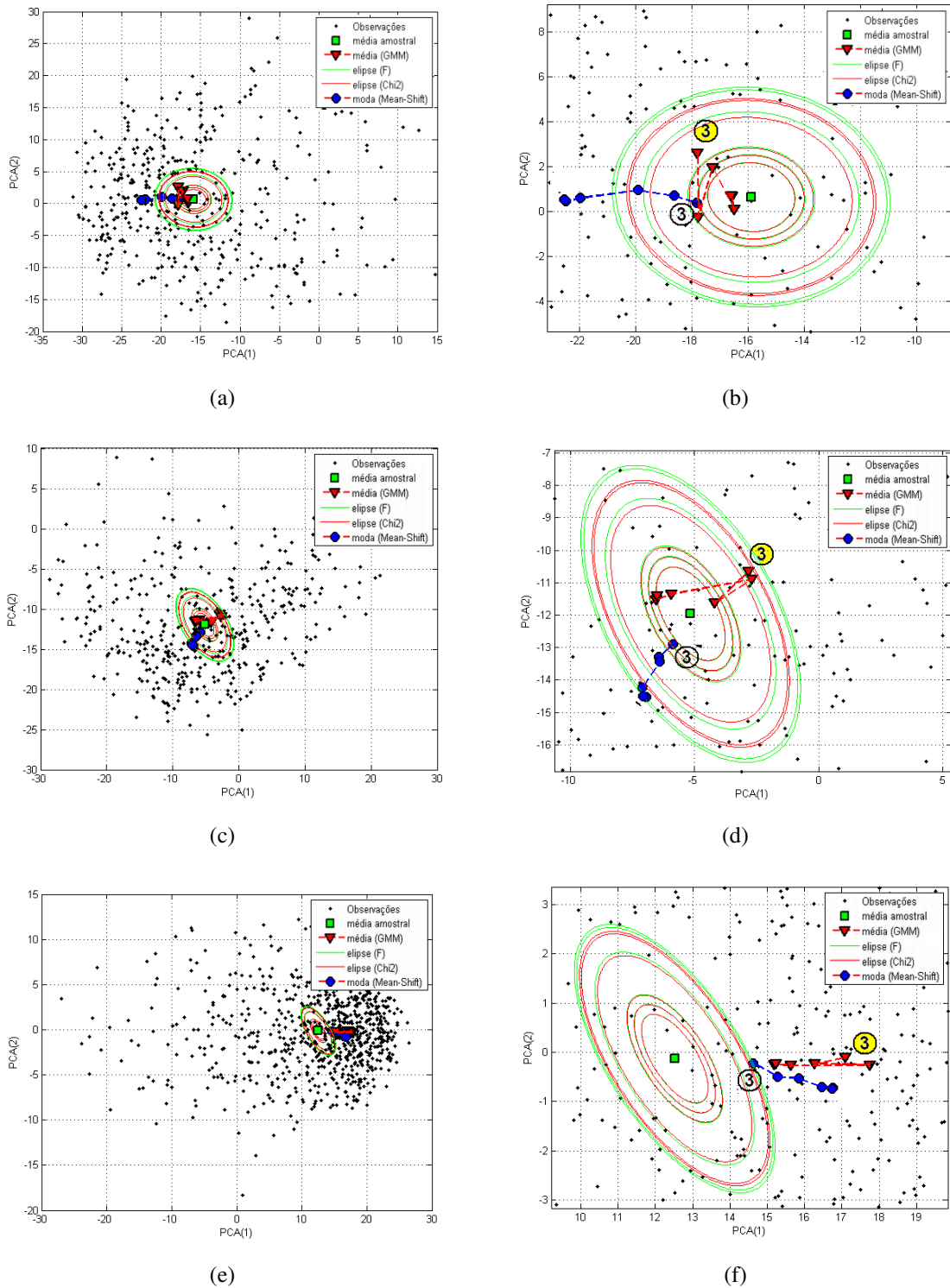


Figura 5.3: Projeção no espaço bidimensional da distribuição de vetores de características, elipses de confiança, da média amostral ( $F$  e  $\chi^2$ ), da moda estimada com *Mean-Shift* e da média estimada com EM para  $d = \{3, 6, 10, 38, 61, 64\}$  para o problema de 4 classes. (a) e (b) Classe **W**. (c) e (d) Classe **RL**. (e) e (f) Classe **TA-A**

No que tange ao **problema de classificação de 5 classes**, a Figura 5.4 apresenta a distribuição dos vetores de características das impressões digitais de treinamento no espaço bi-dimensional. A partir desta figura, pode-se avaliar as médias amostrais, as médias estimadas com EM, os centroides estimados com *K-means*, as modas estimadas com *Mean-Shift* e as elipses de confiança em torno das médias amostrais das classes LL, W, RL, TA e A. Observa-se que em nenhum dos casos a média estimada com o EM ficou dentro da elipse de confiança; já no caso do centroide, apenas o correspondente à classe A a que ficou dentro da elipse. No caso das modas, as correspondentes às classes LL e A foram as que ficaram dentro da elipse; a moda e o centroide da classe A se sobrepueram ainda que os pontos iniciais do algoritmo *Mean-Shift* foram aleatórios.

Nas Figuras 5.5(a) e 5.5(b) verificam-se as projeções dos vetores de características, as elipses de confiança e os parâmetros estimados sobre o espaço bidimensional formado pelos dois primeiros coeficientes PCA para os casos das classes TA e A respectivamente. Na classe TA, as médias estimadas aproximam-se da média amostral enquanto o número de características aumentou e ficaram dentro da elipse de confiança a partir de  $d = 38$ . A moda estimada, como nos casos anteriores, tendeu-se afastar da média amostral; pelo contrário, na classe A todas as modas estimadas ficaram dentro das elipses de confiança correspondentes.

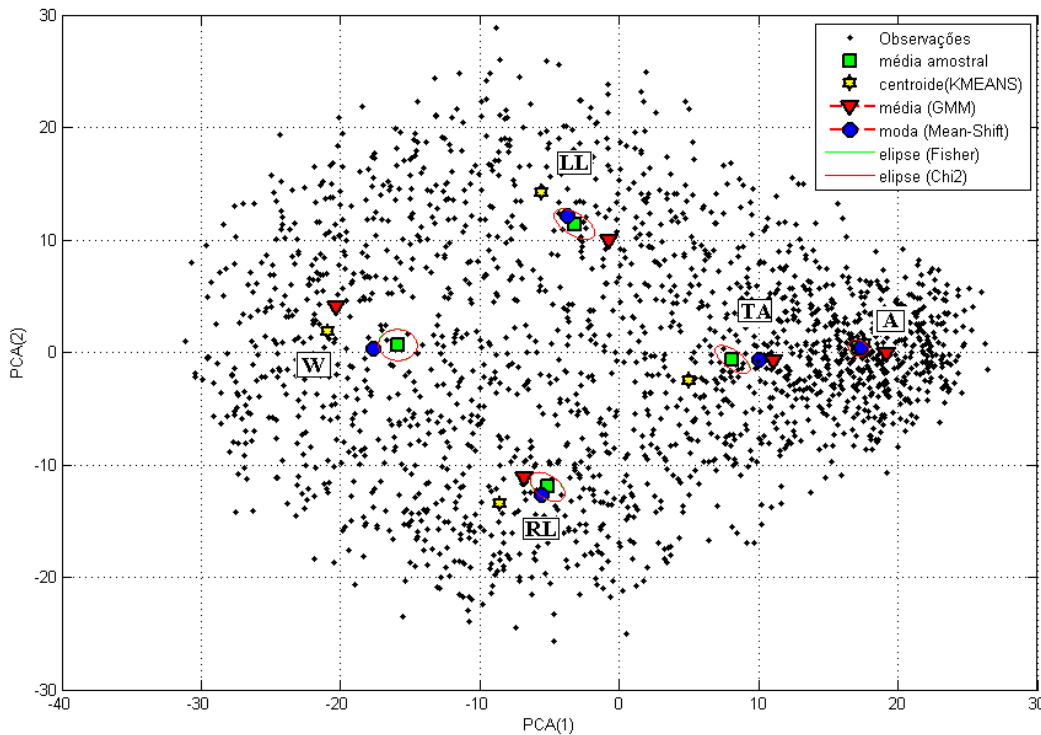


Figura 5.4: Distribuição dos vetores de características, médias amostrais, centroide estimada com *K-means*, média estimada com EM, moda estimadas com *Mean-Shift* e elipses de confiança  $F$  e  $\chi^2$  correspondentes a cada classe para o problema de 5 classes.

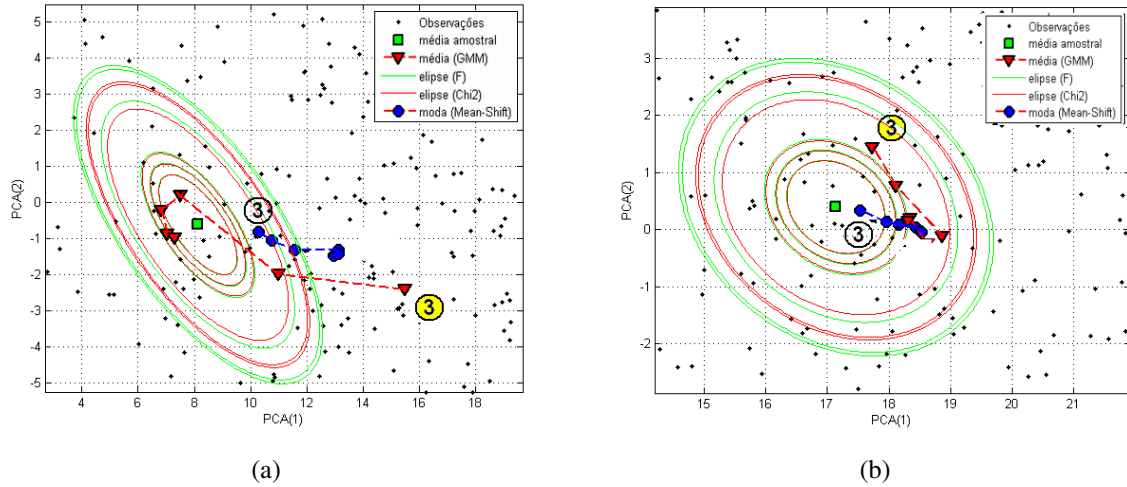


Figura 5.5: Projeção no espaço bidimensional da distribuição de vetores de características, elipses de confiança, da média amostral ( $F$  e  $\chi^2$ ), da moda estimada com *Mean-Shift* e da média estimada com EM para  $d = \{3, 6, 10, 38, 61, 64\}$  e para o problema de 5 classes. (a) Classe TA. (b) Classe A.

No que respeita às classes LL, W e RL no problema de 5 classes os resultados são amostradas nas figuras B.1(a), (b) e (c) do Apêndice B. Observa-se que os resultados estimados foram semelhantes aos obtidos no problema de 4 classes quando as dimensões  $d$  dos vetores de características foram incrementados. Assim, salvo o caso de  $d = 10$ , a média estimada tendeu se aproximar da média amostral enquanto o número de coeficientes PCA cresceu, do mesmo modo, nesta relação, todas as modas estimadas estiveram propensas a se afastarem da média amostral.

## 5.2 Avaliação da entropia mutua

Nesta seção realiza-se a medida da divergência de Kullback-Leibler entre dois de funções de densidade de probabilidade Gaussiana; a distribuição Gaussiana cujos parâmetros estão dados pela media amostral e pela covariância amostral frente à distribuição Gaussiana cujos parâmetros estão dados pela média e pela covariância estimadas com o algoritmo EM. Esta medição é feita com os resultados obtidos na estimação de parâmetros Não-Supervisionado do GMM na qual cada componente da misturas modela a distribuição dos vetores de características de cada classe no espaço  $d$ -dimensional usando a Equação (4.27) onde  $d$  é o número de coeficientes PCA do vetor de características.

A Figura 5.6(a) mostra o resultado da divergência de Kullback-Leibler simetrizada (chamada de Jeffreys-Kullback-Leibler) com respeito a cada classe para o problema de 4 classes.

No caso da classe LL pode-se ver que o valor da divergência tendeu se incrementar enquanto o número  $d$  de coeficientes PCA aumentou, mas, apresentou um incremento significativo quando o vetor de características teve  $d = 10$  coeficientes PCA e alcançou o máximo valor em  $d = 25$ ; também pode-se observar que na faixa entre  $6 < d < 10$  o valor da divergência teve um incremento considerável e na faixa entre  $24 < d < 28$  tal divergência diminuiu.

Na classe W os valores da divergência são pequenos em relação aos obtidos nas outras classes. Na curva correspondente à classe RL pode-se ver que na faixa entre  $6 < d < 9$  o valor da divergência teve um incremento considerável e na faixa entre  $20 < d < 25$  teve-se os valores mais altos e só em  $n = 27$  a divergência diminuiu e quase não apresentou variação enquanto o número de coeficientes PCA aumentou. Já no caso da classe TA-A os valores da divergência foram muito grandes (veja a Figura B.3 no Apêndice B), o que indicaria que os parâmetros estimados e os amostrais nesta classe foram muito diferentes.

Estes resultados têm coerência com a análise realizada a partir da elipse de confiança.

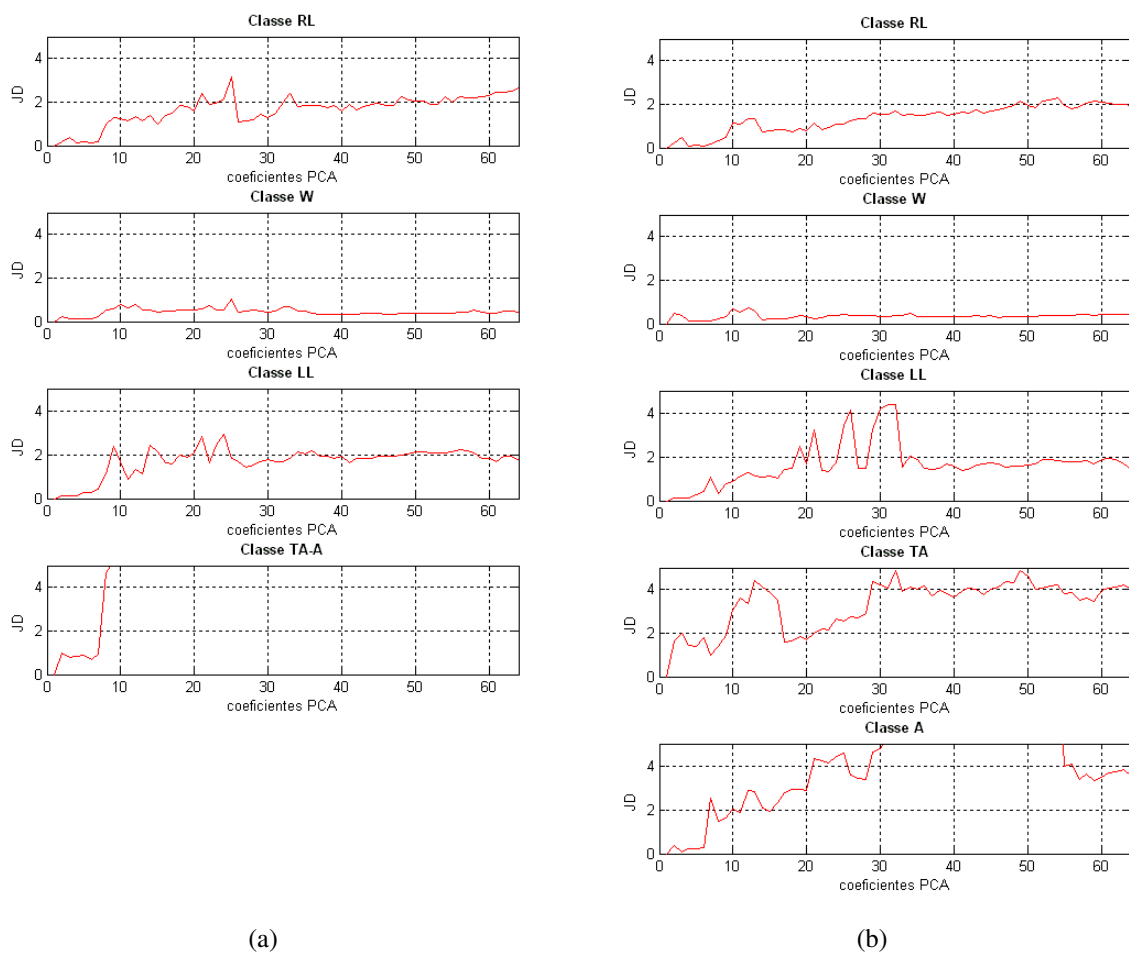


Figura 5.6: Divergência de Jeffreys-Kullback-Leibler (a) Problema de 4 classes. (b) Problema de 5 classes.

Assim, quando se observa que a localização da média estimada ficou longe da média amostral, para um determinado  $d$ , é possível que a divergência mútua seja maior. No caso da classe TA-A, na Figura 5.3(f) viu-se que as médias estimadas para  $d = \{3, 10, 38, 60, 61\}$  ficaram fora da elipse de confiança.

A Figura 5.6(a) mostra o resultado da divergência para o problema de 5 classes. Por meio da observação desta figura, percebe-se que a divergência para a classe LL na faixa  $20 < d < 30$  foi menor do que no problema de 4 classes. A classe RL apresentou valores de divergência muito grandes entre  $18 < n < 35$ . Já na classe classe W foram verificadas pequenas divergências. Os valores de divergência nas classes Ta e A foram grandes, porém viu-se que foram menores do que os obtidos na classe TA-A. Isto indica que existe menor divergência entre as gaussianas estimadas das classes TA e A por separado do que a gaussiana TA-A. Na Figura B.2 do apêndice B.3 são amostrados as divergências de Kullback-Leibler para os problemas de 4 e 5 classes.

### 5.3 Resumo

Na primeira parte do capítulo analisou-se os resultados da estimação Não-Supervisionada da média via o teste de hipótese. Viu-se para a maioria dos testes (salvo caso particulares como por exemplo  $d = 10$ ) que conforme a dimensão do vetor de características aumentou as médias estimadas via o algoritmo EM se aproximaram mais da média amostral. No entanto, isso não aconteceu com as modas estimadas com o *Mean-Shift*, já que a tendência foi se afastar da média amostral, ainda que para algum dos casos passaram o teste de hipótese. Tendo em conta que a busca da moda foi feita a partir de valores aleatórios poder-se-ia propor uma forma alternativa de se inicializar o algoritmo EM usando as modas estimadas com o *Mean-Shift*.

Na segunda parte, avaliou-se a divergência de Kullback-Leibler e viu-se que a função de densidade de probabilidade estimada com os vetores de característica da classe W foi a que apresentou a menor divergência. No caso das classes LL e RL, tiveram-se quando o vetor teve 10 coeficientes PCA o número de coeficientes, o que tem concordância com os resultados obtidos no teste de hipótese, pois com esta quantidade de coeficientes a média estimada com EM ficava mais longe da média amostral. No caso das classes TA e A, as funções de densidade de probabilidade estimadas foram muito distintos do que as obtidas com os parâmetros amostrais, tanto no problema do 4 classes quanto no problema de 5 classes.



## Capítulo 6

# Classificação Não-Supervisionada usando GMM e EM

Na seção 5.1 foi feito o teste de hipótese via região de confiança das médias estimadas (por meio de diversas técnicas tais como GMM, *K-means* e *Mean-Shift*) com relação às médias amostrais das distribuições dos vetores de características das imagens de impressões digitais de cada classe.

No presente capítulo são amostrados os resultados da classificação Bayesiana usando as funções densidade de probabilidade estimadas que formam parte do GMM, sendo que os parâmetros desses modelos de misturas foram estimadas através da abordagem Não-Supervisionada utilizando o algoritmo EM. Realizaram-se três experimentos de acordo com o tipo de inicialização de parâmetros do algoritmo:

1. Inicialização do EM com os parâmetros amostrais das distribuições das impressões digitais de treinamento de cada classe.
2. Inicialização do EM com os centroides estimados via *K-means*.
3. Inicialização do EM com modas estimadas via *Mean-Shift*.

Os resultados experimentais são apresentados via a taxa de erro de classificação alcançados usando cada tipo de inicialização do algoritmo EM sobre o conjunto de impressões digitais de teste do banco de dados DB4. Porém, antes de falar-se dos resultados, é pertinente apresentar os passos da metodologia utilizada para o treinamento e o teste de classificação.

## 6.1 Metodologia usada da classificação Não-Supervisionada

O treinamento Não-Supervisionado consiste em estimar os parâmetros do GMM que modela a distribuição de todas as impressões digitais sob a abordagem Não-Supervisionado. O teste consiste em determinar qual é a classe à que pertence uma impressão digital que não foi parte do treinamento usando o classificador Bayesiano. Dessa forma, para realizar o **treinamento Não-Supervisionado** seguiram-se os seguintes passos:

1. Define-se o problema de classificação (4 ou 5 classes).
2. Divide-se o banco de imagens de impressões digitais DB4 em duas metades (uma para treinamento e outra para o teste).
3. Escolhe-se a primeira metade.
4. Logo depois as imagens são pre-processadas a partir das técnicas desenvolvidas no trabalho prévio e descritas no Capítulo 3, assim:
  - faz-se a segmentação da impressão digital determinando as regiões da imagem conforme descrito na Seção 3.1.1,
  - faz-se a melhora da imagem conforme descrito na Seção 3.1.2,
  - determina-se a imagem de orientação conforme descrito na Seção 3.1.3,
  - determina-se o ponto de referência conforme descrito na Seção 3.1.4 para o registro, e,
  - acha-se o resultado de vetorizar todos os elementos da representação vetorial da imagem de orientação conforme descrito na Seção 3.2.1.
5. Após o pre-processado das imagens define-se o número de elementos do vetor de características.
6. Reduz-se a dimensão do vetor de características usando a técnica PCA.
7. Remove-se os rótulos de classes dos vetores de características.
8. Antes de executar o algoritmo EM de estimação de parâmetros, define-se o tipo de inicialização de parâmetros.
9. Estima-se os parâmetros do GMM usando o algoritmo de estimação EM conforme descrito na Seção 4.4.
10. Salva-se -como resultado do treinamento- os parâmetros do GMM estimados.

Na Figura 6.1(a) é mostrada a distribuição dos vetores de características -com dois coeficientes PCA- das imagens de treinamento rotuladas quando foi definido o problema de 4 classes (passo 1), quando foi dividido o banco de dados em duas metades (passo 2) e quando foi escolhida a primeira metade para o treinamento (passo 3). Estes vetores de características de dois elementos são o resultado de processar as imagens (passo 4) e obter os vetores a partir da imagem de orientação (passo 5) e reduzir a dimensão destes usando a técnica PCA (passo 6). Posto que até este ponto os vetores ainda têm os rótulos, assim, na Figura 6.1(a) atribuíram-se as cores vermelho, verde, ciano e azul para os pontos correspondentes a os vetores de características das classes LL, W, RL e TA-A respectivamente.

Na Figura 6.1(b) são amostrados os pontos da distribuição dos vetores de característica quando foi definido o problema de 5 classes; neste caso atribuíram-se as cores vermelho, verde, ciano, amarelo e azul para os pontos correspondentes a os vetores de características das classes LL, W, RL, TA e A respectivamente.

Na Figura 6.2(a) mostra-se o histograma, com 15 intervalos em cada dimensão dos pontos correspondentes aos vetores de características das impressões digitais de treinamento não rotuladas depois de aplicar o passo 7, quando foram considerados os dois primeiros coeficientes PCA. Na Figura 6.2(b) mostra-se o histograma dos mesmos pontos, mas desta vez com 31 intervalos em cada dimensão.

Na Figura 6.3(a) mostra-se a função de densidade de probabilidade do modelo de misturas estimada via o algoritmo EM a partir dos pontos correspondentes aos vetores de características das impressões digitais mostradas na Figura 6.1(a) quando foram considerados os parâmetros amostrais como parâmetros iniciais do algoritmo (passo 8), para o problema de 4 classes, a qual consiste de um modelo GMM com 4 componentes Gaussianas (passo 9). Percebe-se a semelhança do modelo estimado com o histograma mostrado na Figura 6.2(a). Na Figura 6.3(a) pode-se observar que o componente com maior coeficiente de ponderação corresponde á classe TA-A, pois no problema de classificação de 4 classes ela modela conjuntamente os pontos das classes TA e A.

Na figura 6.3(b) mostra-se a função de densidade de probabilidade estimada para o problema de 5 classes. Nela é evidente observar que a componente que corresponde à classe TA-A no problema de 4 classes, foi dividida em duas componentes, uma para a classe TA e outra para a classe A. O resultado da estimação também influencia nos parâmetros das componentes correspondentes às outras classes (LL, W e RL) e nesta figura também é possível ver que as médias e covariâncias destas classes estimadas no problema de 4 classes não forma iguais do que as estimadas no problema de 5 classes, como é possível conferir os valores em cada caso na seção B.1 do Apêndice B.

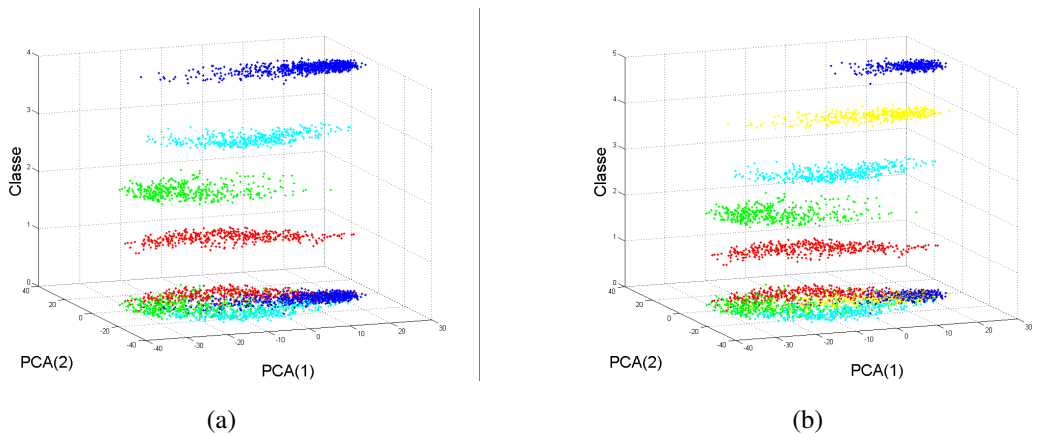


Figura 6.1: Vetores de características com dois coeficientes PCA das imagens de treinamento rotuladas. (a) Problema de 4 classes: LL (vermelho), W (verde), RL (ciano) e TA-A (azul), (b) Problema de 5 classes: LL, W, RL, TA (amarelo) e A (azul).

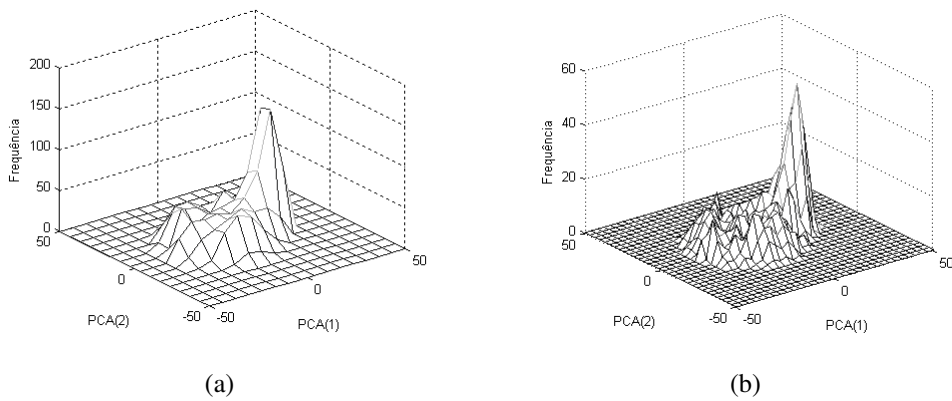


Figura 6.2: Histograma dos vetores de características com dois coeficientes PCA das imagens de treinamento sem rotular (a) 15 intervalos (b) 31 intervalos por dimensão.

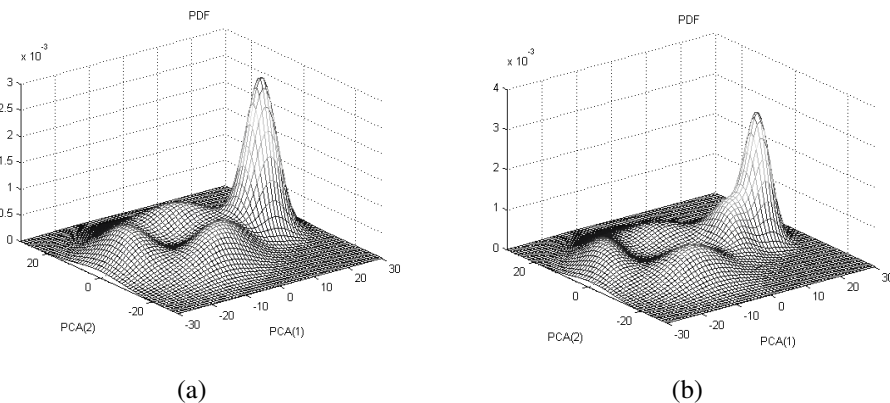


Figura 6.3: Função de densidade de probabilidade dos modelos de misturas gaussianas estimados para o problema de (a) 4 classes e (b) 5 classes.

Para realizar o teste da classificação Não-Supervisionada usando os parâmetros das componentes do GMM estimados via EM, faz-se:

1. Escolhe-se a segunda metade da base de dados DB4.
2. Repete-se o pre-processamento da imagem da impressão digital (o passo 4).
3. Implementa-se um classificador Bayesiano tendo em conta os parâmetros das funções de densidade de probabilidade Gaussiana estimados para cada classe.
4. Avalia-se o desempenho da técnica usando a taxa de erro de classificação.

Seguidamente apresentam-se os resultados da classificação Não-Supervisionada alcançados quando o algoritmo de estimação e parâmetros EM foi inicializado de três maneiras diferentes. Inicialização com parâmetros amostrais, inicialização com centroides estimados via *K-means* e inicialização com modas estimadas via *Mean-Shift*.

## 6.2 Resultados da classificação Não-Supervisionada iniciando o EM com parâmetros amostrais

Como mencionado nos capítulos 4 e 5, a inicialização do algoritmo EM com os parâmetros amostrais das distribuições dos vetores de características de cada classe permite avaliar se a estimação dos parâmetros a partir de dados não rotulados da como resultado parâmetros muito distintos aos iniciais. A classificação, da mesma forma que em todos os experimentos, foi testada para um conjunto de casos ( $d = \{2, 3, \dots, 64\}$ ) que permita observar o desempenho do classificador em função do número  $d$  de coeficientes PCA no vetor de características.

### Problema de 4 classes

Na Figura são mostrados os resultados da classificação de impressões digitais obtidos quando foram usados os parâmetros amostrais como parâmetros iniciais do algoritmo EM para o **problema de 4 classes**. Nos gráficos da coluna esquerda são amostrados os resultados quando considerou-se a distribuição natural; assim, na Figura 6.5(a) mostram-se as taxas de erro total de classificação e nas figuras, 6.5(c), 6.5(e), 6.5(g) e 6.5(i) mostram-se as taxas de erro das classes LL, W, RL e TA-A respectivamente.

Da mesma forma, nos gráficos da coluna direita mostram-se os resultados quando considerou-se a distribuição balanceada.

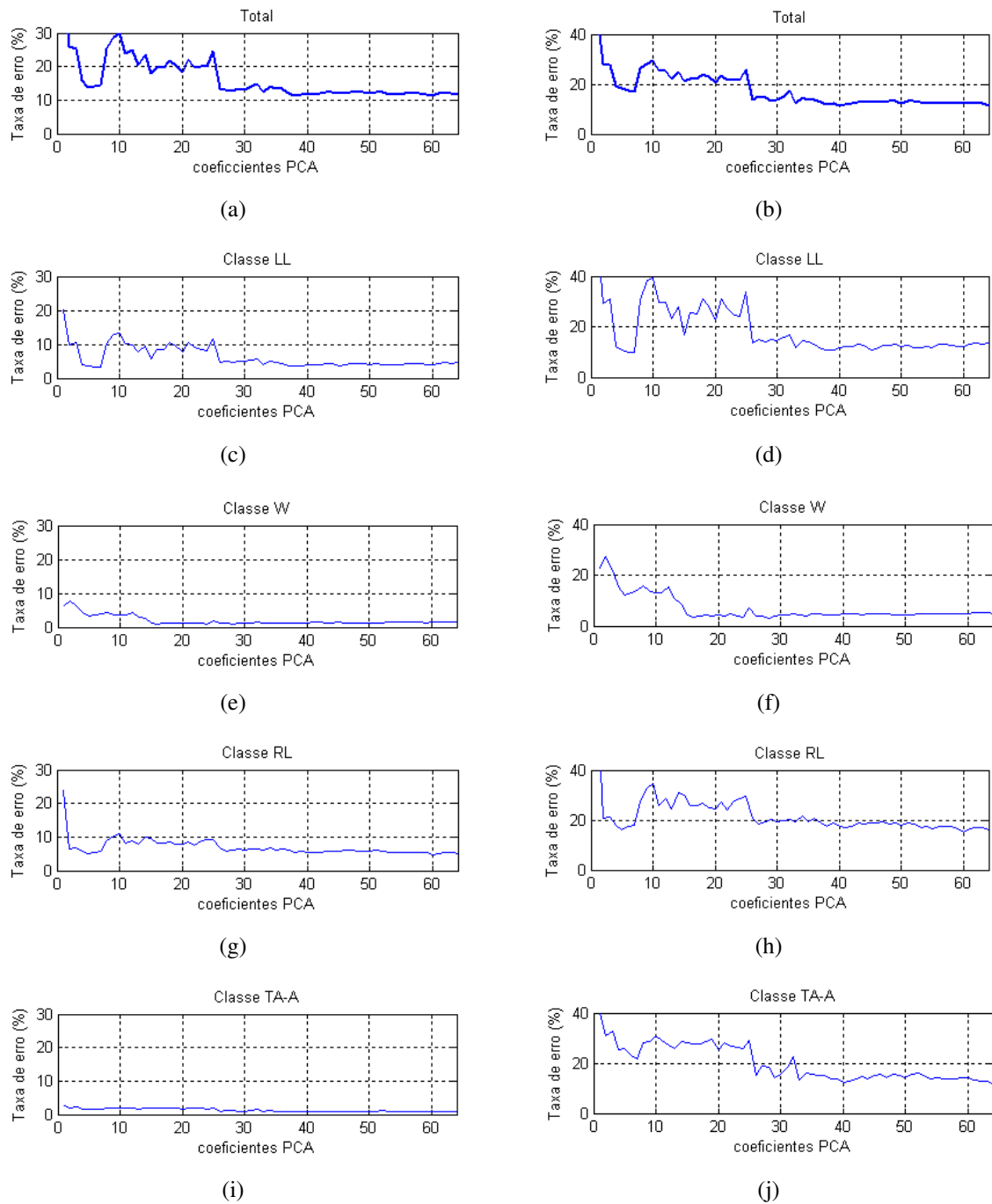


Figura 6.4: Curvas das taxas de erro de classificação natural ((a), (c), (e), (g) e (i)) e balanceada ((b), (d), (f), (h) e (j)) para o problema de 4 classes quando os parâmetros amostrais foram consideradas como parâmetros iniciais do EM.

No caso da **distribuição natural**, pode-se observar que:

- A taxa de erro total foi a soma das taxas de erro de todas as classes, já que o resultado de cada uma delas foi ponderado com as porcentagens de ocorrência natural.
- A menor taxa de erro total (11.29%) foi obtida quando tiveram-se  $d = 60$  coeficientes PCA no vetor de características das impressões digitais, no entanto, com apenas  $d = 38$  coeficientes PCA obteve-se uma taxa de erro de 11.35%.
- Apesar de se possuir um coeficiente PCA a mais no vetor de característica, a taxa de erro total com  $d = 8$  (25.38%) foi consideravelmente maior do que a taxa de erro alcançado com  $d = 7$  (14.45%). Isto indica que houve coeficientes que não contribuíram com a separabilidade entre classes distintas; sendo assim, a curva do erro total não decresceu monotonicamente quando o número de coeficientes PCA aumentou.
- A curva da taxa de erro total tornou-se estável em torno de 11.5% (não apresentando variações consideráveis) quando a classificação foi testada com o número de coeficientes PCA acima de  $d = 38$  por essa razão o valor obtido neste ponto não foi muito distinto do que o mínimo valor alcançado com  $d = 60$ .
- A classe LL apresentou a menor taxa de erro (3.59%) quando foram consideradas apenas os primeiros  $d = 5$  coeficientes PCA.
- Nas outras três classes precisou-se de um maior número de coeficientes no vetor de características para alcançar a mínima taxa de erro; assim, na classe W alcançou-se 1.18% com  $d = 38$ , na classe RL alcançou-se 4.85% com  $d = 60$  e na classe TA-A alcançou-se 0.74% com  $d = 64$ .
- Já que foram ponderadas com a porcentagem da ocorrência natural, a curva da taxa de erro da classe TA-A (veja a Figura 6.5(i)) não influenciou consideravelmente a taxa de erro total, só no caso da distribuição natural, pois teve valores pequenos em comparação com as outras classes.
- A curva da taxa de erro da classe W (veja a Figura 6.5(e)), apesar de não ter decrescido monotonicamente poder-se-ia considerar que sim, já que não apresentou incrementos significativos enquanto o número de coeficientes PCA do vetor características aumentou. A taxa de erro desta classe alcançou comportamento estável a partir de  $d = 16$ .
- As curvas das taxas de erro das classes LL e RL (veja as Figuras 6.5(c) e 6.5(g)) não decresceram monotonicamente, já que as taxas das ambas classes em  $d = 10$  foi maiores do que em  $d = 7$ . A curva da taxa de erro da classe LL não apresentou variações consideráveis e tornou-se estável com o número de coeficientes PCA acima de  $d = 38$ ; por outro lado, a curva da classe RL tornou-se estável apenas com coeficientes PCA acima de  $d = 40$ .

No caso da **distribuição balanceada**, ainda para o problema de 4 classes, pode-se observar que:

- A taxa de erro total foi a média das taxas de erro das classes separadas, já que no treinamento foram usadas a mesma quantidade de impressões digitais de cada classe.
- A menor taxa de erro total foi alcançada com  $d = 34$  (11.34%);
- A taxa de erro total alcançada com  $d = 8$  (26.1%) foi consideravelmente maior do que a taxa de erro total com  $d = 7$  (17.15%).
- A curva da taxa de erro total deixou de apresentar variações consideráveis, em torno de 12%, com o número de coeficientes PCA acima de  $d = 38$ ; assim, o resultado com  $d = 40$  (11.60%) não foi muito diferente do mínimo valor alcançado com  $d = 64$  (veja a Tabela 6.1).
- No referente às taxas mínimas de erro de cada classe, pode-se ver que na classe LL apresentou-se a mínima taxa de erro com apenas  $d = 5$  (10.62%); no entanto, nas classes restantes precisaram-se de vetores com mais coeficientes PCA, assim, na classe W alcançou-se a mínima taxa de erro com  $d = 38$  (4.23%), na classe RL com  $d = 60$  (15.31%) e na classe TA-A com  $d = 64$  (11.22%).
- Na classe LL (veja a Figura 6.5(d)) teve-se grande diferença, de quase 30%, entre as taxas de erro alcançadas com  $d = 6$  (9.84%) e com  $d = 10$  (39.11%). A curva da taxa de erro apresentou pequenas variações com o número de coeficientes PCA no vetor de característica acima de  $d = 37$  com valores entre 10% e 13%.
- A curva da taxa de erro da classe W (veja a Figura 6.5(f)) tendeu decrescer monotonicamente, apesar de demonstrar pequenos incrementos em  $d = \{7, 8, 12, 25, 30\}$ ; assim poder-se-ia considerar que os valores da taxa de erro começaram apresentar pequenas variações com o número de coeficientes PCA acima de  $d = 15$ , com valores entre 4% e 5%.
- Na curva da taxa de erro da classe RL (veja a Figura 6.5(h)), teve-se grande diferença, de quase 17%, entre as taxas de erro alcançadas com  $d = 6$  (17.60%) e com  $d = 10$  (34.69%). Os valores da curva da taxa de erro oscilaram entre 15% e 22% quando o número de coeficientes PCA esteve entre  $d = 30$  e  $d = 64$ .
- A distribuição balanceada permitiu fazer uma análise das taxas de erro da classe TA-A (veja a Figura 6.5(j)), já que nela apresentou-se a mesma quantidade de observações que as outras classes; assim, nesta classe a curva da taxa de erro oscilou entre 11% e 18% quando o número de coeficientes PCA esteve entre  $d = 33$  e  $d = 64$ .



PCA	Taxa de erro (Natural)					Taxa de erro (Balanceada)				
	Total	LL	W	RL	TA-A	Total	LL	W	RL	TA-A
2	26.13	9.89	7.63	6.55	2.05	27.95	29.27	27.36	20.66	31.10
3	25.41	10.51	6.04	6.71	2.16	27.90	31.09	21.64	21.17	32.68
5	13.78	<u>3.59</u>	3.40	5.09	1.69	18.20	<u>10.62</u>	12.19	16.07	25.73
10	29.95	13.22	3.67	10.99	2.05	29.75	39.11	13.18	34.69	31.09
32	14.88	5.60	1.39	6.39	1.50	17.45	16.58	4.98	20.15	22.68
38	11.35	3.68	<u>1.18</u>	5.58	0.91	12.05	10.88	<u>4.23</u>	17.60	13.78
40	11.56	3.94	<u>1.32</u>	5.49	0.80	11.60	11.65	4.72	17.34	12.19
60	<b>11.29</b>	4.12	1.39	<u>4.85</u>	0.93	12.15	12.18	4.98	<u>15.31</u>	14.15
64	11.78	4.64	1.39	5.01	<u>0.74</u>	<b>11.35</b>	13.73	4.98	15.82	<u>11.22</u>

Tabela 6.1: Taxas de erro (%) de classificação Não-Supervisionado para diferentes quantidades de coeficientes PCA no vetor de características quando foram considerados os parâmetros amostrais como parâmetros iniciais do EM para o problema de 4 classes.

- Nas curvas de erro das classes LL, RL e TA-A percebe-se que nas zonas que podem ser consideradas estáveis ou de pouca variação, acima de  $d = 35$ , os valores das taxas de erro da classe TA-A foram maiores do que os valores da classe LL e menores do que RL. Para o caso da distribuição W, os valores foram menores do que todas as taxas de erro das outras classes.

Na Tabela 6.1 mostram-se os valores das taxas de erro total de classificação e as taxas de erro por classes quando foram consideradas  $d = \{2, 3, 5, 10, 32, 38, 40, 60, 64\}$  coeficientes PCA no vetor de características.

Continuando com a análise dos resultados, na Tabela 6.2(a) mostra-se a **matriz de confusão** obtida quando foi testada a classificação Bayesiana com as funções de densidade de distribuição de probabilidade Gaussiana estimadas de forma Não-Supervisionada com o algoritmo EM considerando  $d = 38$  coeficientes PCA no vetor de características e na Tabela 6.2(b) mostra-se a matriz de confusão obtida com  $d = 60$  coeficientes PCA.

No caso da matriz obtida com  $d = 38$  pode-se ver que os resultados evidenciaram que:

- Tem-se pequena variação inter-classes entre as classes RL e TA-A, já que 41 impressões digitais de teste pertencentes à classe RL foram classificadas como impressões da classe TA-A e 26 impressões digitais da classe TA-A foram classificadas como impressões da classe RL; em ambos os casos as funções de densidade de probabilidade estimadas não conseguiram separar regiões com pontos correspondentes a vetores de características de distintas classes.
- No caso das classes RL e LL tem-se grande variação inter-classes entre elas apesar da semelhança das distribuições de crestas e vales das suas áreas padrão, já que apenas 3

(a)					(b)				
	LL	W	RL	TA-A		LL	W	RL	TA-A
LL	344	14	3	25	LL	339	21	2	24
W	7	385	4	6	W	9	382	4	7
RL	6	22	323	41	RL	6	11	332	43
TA-A	79	8	26	707	TA-A	79	9	28	704

Tabela 6.2: Matriz de confusão da classificação com misturas gaussianas para o problema de quatro classes com: (a) 38 ; e (b) 60 coeficientes PCA.

impressões digitais da classe LL foram classificadas como se fossem da classe RL e 6 impressões da classe RL foram classificadas como LL.

- No caso das classes RL e W apenas 4 impressões digitais da classe W foram classificadas como se fossem da classe RL, porém 22 impressões da classe RL foram classificadas como W.
- No caso da comparação da classe W com as outras classes, tem-se que apenas 7 e 6 impressões digitais desta classe foram reconhecidas como pertencentes às classes LL e TA-A, respectivamente.
- Pode-se afirmar que a classe W tem pequena variação intra-classe.
- No caso das classes LL e TA-A tem-se uma pequena variação inter-classes já que 25 impressões digitais da classe LL foram classificadas como impressões da classe TA-A e 79 impressões digitais da classe TA-A foram classificados como impressões da classe LL.
- Pode-se afirmar que as impressões digitais da classe TA-A tem grande variação intra-classe já que tem-se que 79, 8 e 26 impressões digitais desta classe foram reconhecidas como se fossem das classes LL e W e RL respectivamente.

No caso da matriz de confusão obtida quando se avaliou a classificação com o vetor de características de  $d = 60$  coeficientes PCA, pode-se observar que as variações inter-classes e as variações intra-classe são semelhantes às obtidas no caso de  $d = 38$ . Assim, tem-se pequena variação inter-classes entre as classes RL e TA-A; nas classes RL e LL tem-se grande variação inter-classes; na classe W tem-se menor variação intra-classe do que as outras classes; nas classes LL e TA-A tem-se pequena variação inter-classes e na classe TA-A tem grande variação intra-classe.

Mas, pode-se perceber que os resultados de classificação com  $d = 60$  foram melhores do que  $d = 38$  apenas no caso da classe RL, na qual classificou corretamente 332 impressões digitais frente a 323. Em todas as outras classes o número de acertos foi menor.

**Problema de 5 classes**

Na figura 6.5 são mostrados os resultados obtidos no caso do experimento em que foram considerados os parâmetros amostrais como parâmetros iniciais do algoritmo EM, para o **problema de 5 classes**. Assim, são amostrados os resultados da distribuição natural e balanceada nos gráficos da coluna da esquerda e da direita respectivamente; dessa forma, na Figura 6.5(a)-(b) mostra-se as taxas de erro total e nas figuras, 6.5(c)-(d), 6.5(e)-(f), 6.5(g)-(h), 6.5(i)-(j) e 6.5(k)-(l) as taxas de erro das classes LL, W, RL, TA e A respectivamente.

No caso da **distribuição natural**, pode-se observar que:

- A curva gerada pelas taxas de erro de classificação total avaliado com os diferentes valores de  $d$  não decresceu monotonicamente já que taxa de erro obtida com  $d = 7$  (16.56%) foi menor do que a obtida com  $d = 10$  (28.34%). A mínima taxa foi alcançada só com  $d = 64$  (12.51%) (veja a Figura 6.5(a)).
- A mínima taxa de erro total obtida foi de 12.51% quando foram considerados  $d = 64$  coeficientes PCA no vetor de características das impressões digitais.
- Os valores das taxas de erro da classe W foram pequenos (veja a Figura 6.5(e)) ainda que foram ponderadas com a porcentagem ocorrência natural (27.9%).
- Na classe W a taxa de erro em  $d = 1$  (7.21%) foi menor do que em  $d = 2$  (10.41%).
- Na classe W alcançou-se a mínima taxa de erro de 1.11 com  $d = 38$ .
- A taxa de erro da classe LL teve o valor mínimo em  $d = 6$ , porém, o valor em  $d = 10$  (12.78%) foi consideravelmente maior do que em  $d = 10$  (4.72%) (veja a Figura 6.5(c)).
- A taxa de erro da classe RL teve o valor mínimo em  $d = 64$  (5.18%) (veja a Figura 6.5(g)).
- Os valores das taxas de erro das classes TA e A foram pequenos se comparadas com as outras classes (veja as Figuras 6.5(i) e 6.5(k)) devido a que foram ponderadas com a porcentagem ocorrência natural. Na classe TA alcançou-se 0.81% com  $d = 64$  e na classe A alcançou-se 0.63% com  $d = 58$ .

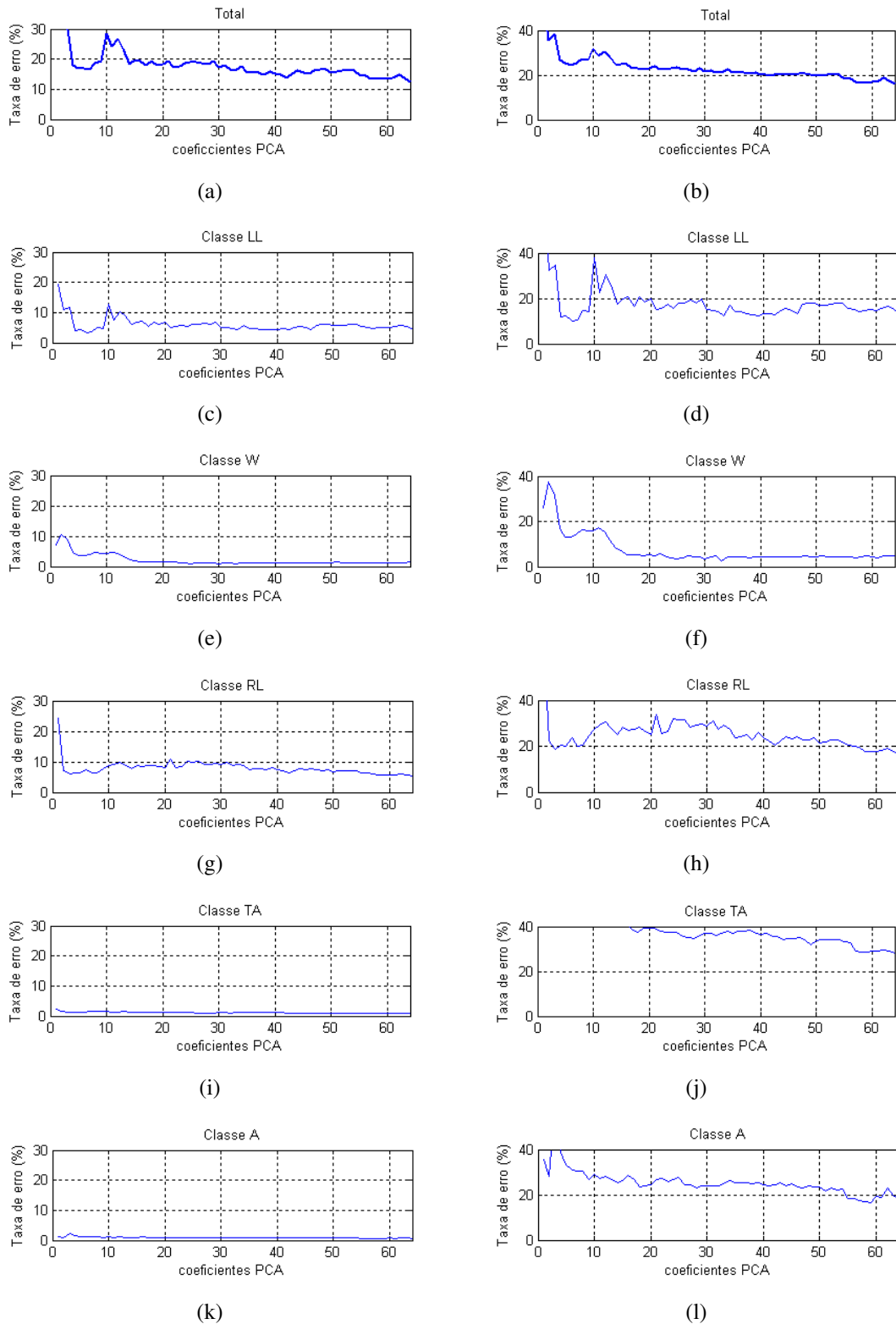


Figura 6.5: Curvas das taxas de erro de classificação natural ((a), (c), (e), (g), (i) e (k)) e balanceada ((b), (d), (f), (h), (j) e (l)) para o problema de 5 classes quando foram consideradas os parâmetros amostrais como parâmetros iniciais do EM.

No caso da **distribuição balanceada**, considerando o problema e 5 classes, pode-se observar que:

- A curva das taxas de erro total não decresceu monotonicamente pois o erro alcançado com  $d = 10$  (31.95%) foi consideravelmente maior do que com  $d = 9$  (26.8%) (veja a Figura 6.5(a)).
- A curva da taxa de erro total não alcançou um comportamento estável e tornou-se decrescente com valores acima de  $d = 16$  (28.57%). Alcançou-se o mínimo valor apenas em  $d = 64$  (16.15%).
- No referente às taxas de erro de cada classe, pode-se ver que na classe LL apresentou-se a mínima taxa de erro com apenas  $d = 6$  (10.09%), na classe W apresentou-se a mínima taxa de erro com  $d = 38$  (3.98%), na classe RL apresentou-se a mínima taxa de erro com  $d = 64$  (16.33%), na classe TA apresentou-se a mínima taxa de erro com  $d = 64$  (28.01%) e na classe A apresentou-se a mínima taxa de erro com  $d = 58$  (17.14%).
- A curva das taxas de erro da classe W não decresceu monotonicamente (veja a Figura 6.5(f)), pois apresentou os valores mais altos quando os vetores de característica tiveram  $d = 2$  coeficientes PCA (37.31%) e  $d = 11$  (17.62%); mas, a curva sofreu pequenas variações com coeficientes PCA acima de  $d = 16$ .
- Na classe LL a curva da taxa de erro não foi monotonicamente decrescente nem apresentou comportamento estável -pequenas variações das taxas de erro conforme o valor de  $d$  aumentou-, mas alcançou o mínimo valor em apenas em  $d = 6$  (10.09%) (veja a Figura 6.5(d)).
- Na classe RL a curva não foi monotonicamente decrescente nem apresentou comportamento estável em nenhum rango de coeficientes PCA. Por isso, alcançou o mínimo valor só com  $d = 64$  (veja a Figura 6.5(h)).
- A curva das taxas de erro da classe TA tiveram valores altos, mas, conseguiu-se ter pouca variação com valores de  $d$  acima de 58, assim, com  $d = 58$  alcançou-se a taxa de erro de 28.50% e com  $d = 64$  a taxa de erro de 28,01% (veja a Figura 6.5(j)).
- As taxas de erro da classe A também tiveram valores altos, sendo assim que a taxa de erro com  $d = 64$  foi de 18.10%, porém alcançou o valor mínimo com  $d = 58$  (17.14%) (veja a Figura 6.5(l)).

Na Tabela 6.3 mostram-se os valores das taxas de erro total de classificação e as taxas de erro de cada classe quando foram consideradas  $d = \{2, 6, 32, 38, 42, 58, 60, 64\}$  coeficientes PCA no vetor de características para o problema de 5 classes.

PCA	Taxa de erro (Natural)						Taxa de erro (Balanceada)					
	Total	LL	W	RL	TA	A	Total	LL	W	RL	TA	A
2	31.00	10.86	10.41	7.04	1.66	1.04	35.40	32.12	37.31	22.19	57.25	28.10
6	16.97	<u>3.41</u>	3.68	7.43	1.3	1.13	24.65	<u>10.09</u>	13.18	23.47	45.01	30.71
32	16.63	4.73	1.32	8.65	1.04	0.89	21.25	13.99	4.73	27.30	36.00	24.05
38	14.80	4.38	<u>1.11</u>	7.28	1.11	0.93	20.70	12.95	<u>3.98</u>	22.96	38.25	25.00
42	13.93	4.29	1.24	6.46	1.02	0.89	19.55	12.69	4.47	20.4	35.50	24.28
58	13.29	4.99	1.17	5.66	0.82	<u>0.63</u>	16.50	14.76	4.22	17.85	28.50	<u>17.14</u>
60	13.46	4.99	1.25	5.66	0.84	0.71	17.10	14.77	4.48	17.86	29.00	19.29
64	<b>12.51</b>	4.47	1.39	<u>5.18</u>	<u>0.81</u>	0.67	<b>16.15</b>	13.21	4.98	<u>16.33</u>	<u>28.01</u>	18.10

Tabela 6.3: Taxas de erro (%) de classificação Não-Supervisionada para diferentes quantidades de coeficientes PCA no vetor de características quando foram considerados os parâmetros amostrais como parâmetros iniciais do EM para o problema de 5 classes.

Na Tabela 6.4(a) mostra-se a matriz de confusão obtida quando se testou a classificação Bayesiana com as funções de densidade de distribuição de probabilidade Gaussiana estimadas de forma Não-Supervisionada com o algoritmo EM considerando  $d = 38$  coeficientes PCA no vetor de características e na Tabela 6.4(b) mostra-se a matriz de confusão obtida com  $d = 64$  coeficientes PCA. No caso da matriz obtida com  $d = 38$  pode-se evidenciar que:

- Nas classes LL e RL não tem-se pequena variação inter-classes já que apenas 2 impressões digitais de teste pertencentes à classe LL foram classificadas como impressões da classe RL e 3 impressões da classe RL foram classificadas como impressões da classe RL.
- A classe W não apresenta pequena variação inter-classes com a classe A, já que apenas 3 impressões da classe A foram classificadas como pertencentes à classe W e nenhuma impressão da classe W foi classificada como se fosse da classe A.
- No caso das classes LL e TA tem-se pequena variação inter-classes, já que 28 impressões digitais de teste pertencentes à classe LL foram classificadas como impressões da classe TA e 30 impressões da classe TA foram classificadas como impressões da classe LL.
- No caso das classes TA e A tem-se pequena variação inter-classes, pois 102 impressões digitais de teste pertencentes à classe TA foram classificadas como impressões da classe A e 97 impressões da classe A foram classificadas como impressões da classe TA. É devido à pequena variação interclasse entre TA e A que a taxa de erro total no problema de 5 classes é maior do que no problema de 4 classes, em que estas duas classes constituem uma classe só.

(a)						(b)					
	LL	W	RL	TA	A		LL	W	RL	TA	A
LL	336	15	2	28	5	LL	335	20	1	28	2
W	8	386	4	1	3	W	10	382	7	1	2
RL	3	20	302	45	2	RL	3	7	328	42	12
TA	30	9	12	247	102	TA	18	12	12	288	70
A	7	0	1	97	315	A	9	1	2	64	344

Tabela 6.4: Matriz de confusão da classificação com misturas gaussianas para o problema de quatro classes com: (a) 38 ; e (b) 64 coeficientes PCA.

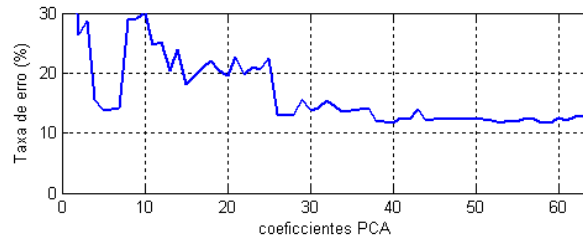
No caso da matriz de confusão obtida quando se avaliou a classificação com o vetor de características de  $d = 64$  coeficientes PCA pode-se observar que as variações inter-classes e as variações intra-classe são semelhantes aos obtidos no caso de  $d = 38$ . Percebe-se que no caso de  $d = 64$  apenas nos casos das classes RL, TA e A tiveram-se melhores resultados com mais acertos na classificação de impressões digitais.

### 6.3 Resultados da classificação Não-Supervisionada inicializando o EM com centroides estimados via K-means

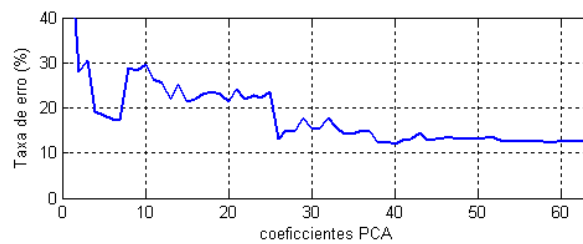
Na seção 5.1 fez-se a avaliação dos centroides estimados via K-means a partir das distribuições dos vetores de características das impressões digitais de treinamento não rotuladas usando como pontos iniciais as médias amostrais de cada classe com o objetivo de avaliar a distância entre os centroides estimados e os pontos iniciais. No problema de 4 classes viu-se que apenas os centroides estimados a partir das médias amostrais das classes LL e RL ficavam dentro da elipse de confiança respectivamente (veja Figura 5.1). No problema de 5 classes apenas a centroide da classe A ficou dentro da elipse de confiança e sobrepu-se com a média amostral e a moda estimada com *Mean-Shift*. Nesta seção apresentam-se os resultados obtidos na classificação Bayesiana das impressões digitais de teste com as funções de densidade de probabilidade estimadas via o algoritmo EM quando foram usados os centroides estimados via K-means como parâmetros iniciais de estimação.

Na Figura 6.6 mostram-se curvas das taxas de erro de classificação (para  $d = 2, ..64$ ) alcançadas quando foram consideradas a distribuição natural (veja a Figura 6.6(a)) e balanceada (veja a Figura 6.6(b)) para o problema de 4 classes; nessa figura pode-se ver que as curvas de taxa de erro resultantes são semelhantes aos obtidos no caso apresentado na seção anterior em que foram considerados os parâmetros amostrais na inicialização do algoritmo EM tanto para a distribuição natural quanto para a distribuição balanceada como se mostram nas Figuras 6.5(a) e 6.5(b). Obtiveram-se estes resultados semelhantes, ainda que na

Figura 5.1 observou-se que nas classes W e TA-A os centroides estimados com *K-means* e as médias estimadas com EM ficaram fora da elipse de confiança da média amostral.

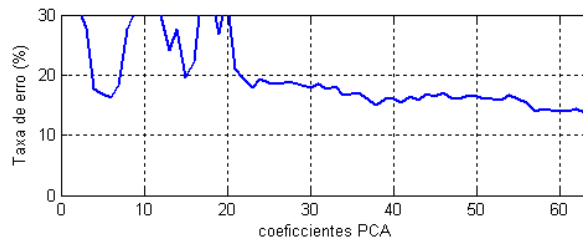


(a)

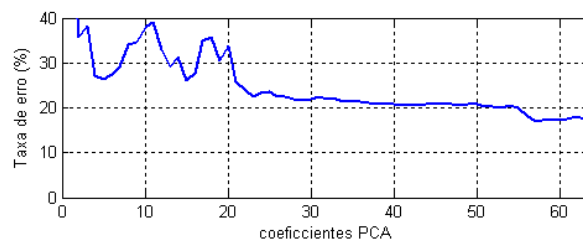


(b)

Figura 6.6: Curvas das taxas de erro de classificação natural (a) e balanceada (b) para o problema de 4 classes quando foram consideradas os centroides estimados com *K-means* como parâmetros iniciais do EM.



(a)



(b)

Figura 6.7: Curvas das taxas de erro de classificação natural (a) e balanceada (b) para o problema de 5 classes quando foram consideradas os centroides estimados com *K-means* como parâmetros iniciais do EM.



Assim, a taxa de erro de classificação (30.02%) com o vetor de características de  $d = 10$  coeficientes PCA foi maior do que a taxa de erro (13.97%) com  $d = 6$  quando na distribuição natural. A menor taxa de erro (11.72%) foi obtida com  $d = 58$ . Da mesma forma, na distribuição balanceada a taxa de erro de classificação com  $d = 10$  (29.69%) também foi maior do que a taxa de erro com  $d = 6$  (17,45%) e a menor taxa de erro (16.81%) foi obtida com  $d = 64$ .

Por outro lado, na Figura 6.7 mostram-se curvas das taxas de erro de classificação (para  $d = 2, \dots, 64$ ) alcançadas quando foram consideradas a distribuição natural (Figura 6.7(a)) e balanceada (Figura 6.7(b)) para o problema de 5 classes. Nestas figuras observa-se que os valores nas curvas de taxa de erro resultantes foram maiores do que as mostradas nas figuras 6.5(a) e 6.5(b) respectivamente. Assim, as menores taxas só foram alcançadas quando o vetor de características teve  $d = 64$  coeficientes PCA; 13.24% na distribuição natural e 16.81% na distribuição balanceada.

Na Tabela 6.5 mostram-se os valores das taxas de erro total de classificação quando foram consideradas  $d = \{5, 6, 10, 11, 26, 40, 57, 58, 64\}$  coeficientes PCA no vetor de características.

PCA	Taxa de Erro (4 classes)		Taxa de Erro (5 classes)	
	Natural	Balanceada	Natural	Balanceada
5	19.79	18.25	16.92	26.25
6	13.97	17.45	16.13	27.45
10	30.02	29.69	38.36	38.05
11	24.78	26.29	39.66	39.01
26	13.71	13.15	18.57	22.35
40	11.73	<b>11.89</b>	16.09	20.79
57	12.26	12.59	13.90	16.95
58	<b>11.72</b>	12.45	14.29	17.35
64	12.04	12.21	<b>13.24</b>	<b>16.81</b>

Tabela 6.5: Taxas de erro (%) de classificação Não-Supervisionado para diferentes quantidades de coeficientes PCA no vetor de características quando foram usados os centroides como parâmetros iniciais do EM para os problemas de 4 e 5 classes.

## 6.4 Resultados da classificação Não-Supervisionada inicializando o EM com modas estimados via *Mean-Shift*

Nesta seção apresenta-se os resultados da classificação Não-Supervisionada de impressões digitais de teste com o classificador Bayesiano usando as funções de densidade de probabilidade do GMM estimados a partir das impressões digitais de treino via o algoritmo EM quando foram usados como parâmetros iniciais as modas das distribuições no espaço  $d$ -dimensional dos vetores de características de cada classe estimada via o algoritmo *Mean-Shift* para os problemas de 4 e 5 classes. O algoritmo *Mean-Shift* procura as modas das distribuições dos vetores de características das impressões digitais de cada classe no espaço  $d$ -dimensional a partir de pontos aleatórios; devido a este tipo de inicialização; neste caso modelou-se o resultado da classificação ( taxa de erro) como uma variável aleatória em que os experimentos foram repetidos um número de vezes; o que não aconteceu com os resultados da classificação descritos nas seções 6.2 e 6.3, nos quais só usaram-se como parâmetros iniciais do algoritmo EM os parâmetros amostrais e as centroides das distribuições de cada classe, respectivamente.

### Problema de 4 classes

Na Figura 6.8 mostram-se os resultados de média e a desviação estandar das taxas de erro de classificação, na qual os experimentos foram repetidos dez vezes; nas Figuras 6.8(a) e 6.8(b) mostram-se os resultados de classificação no problema de quatro classes para as distribuições natural e balanceada respectivamente; e nas Figuras 6.8(c) e 6.8(d), com o objetivo de observar melhor as desviações estandar das taxas de erro, mostram-se os resultados obtidos apenas com o número de coeficientes PCA acima de  $d = 25$  para as distribuições natural e balanceada respectivamente.

No caso da **distribuição natural** (veja as Figuras 6.8(a) e (c)) percebe-se que a curva das médias dos resultados obtidos para cada valor de  $d$  é muito semelhante à curva das taxas de erro obtidas quando foram considerados os parâmetros amostrais das distribuições como parâmetros iniciais do algoritmo EM, que foi amostrada na Figura 6.5(a). Assim, nos dois gráficos alcançou-se a taxa de erro maior quando o vetor de características teve  $d = 10$  coeficientes PCA, a taxa de erro total com  $d = 8$  foi consideravelmente maior do que a taxa de erro alcançado com  $d = 7$  apesar de ter tido um coeficiente a mais; a curva da taxa de erro total tornou-se estável -não apresentado variações consideráveis- quando a classificação foi testada com o número de coeficientes PCA acima de  $d = 38$ . As maiores desviações estandar dos resultados obtidos deram-se com  $d = 25$  e com  $d = 35$ ; em torno de 5% no primeiro caso e de 3% no segundo. Com o número de coeficientes PCA acima de  $d = 35$  as desviações estandar foram menores, não sendo nenhuma delas maior do que 1%.

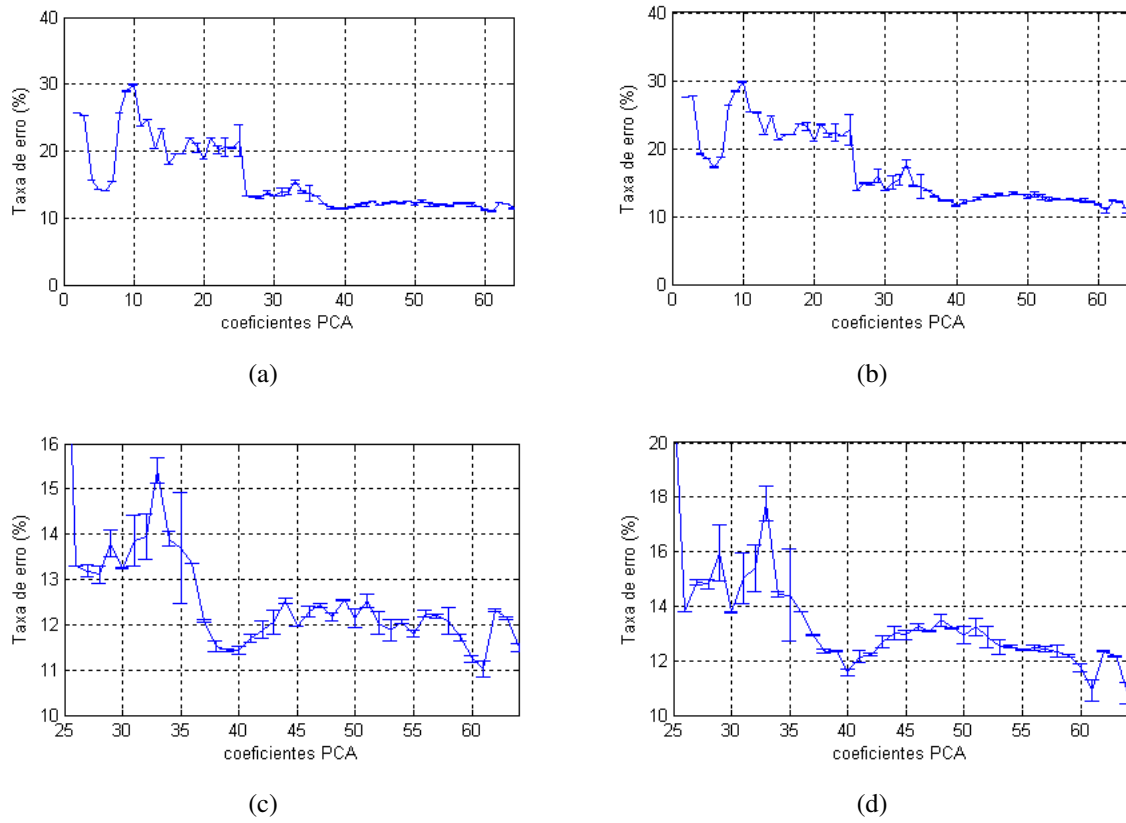


Figura 6.8: Taxas de erro modeladas como variáveis aleatórias para o problema de 4 classes quando foram consideradas as modas achadas via *Mean-Shift* como parâmetros iniciais do EM. (a) e (c) Distribuição natural. (b) e (d) Distribuição balanceada.

Assim, no problema de quatro classes com distribuição natural poderia-se considerar que os resultados obtidos quando foram usados os parâmetros amostrais como parâmetros iniciais do algoritmo EM (mostrados na Figura 6.5(a) da seção 6.2) constituem mais uma realização da variável aleatória amostrada na Figura 6.8(b). Da mesma forma, acontece com os resultados obtidos quando foram considerados os centroides (estimados via *K-means*) como parâmetros iniciais do algoritmo EM cujos resultados foram mostrados na Figura 6.6(a).

No caso do problema de 4 classes com **distribuição balanceada**, nas Figuras 6.8(b) e (d) pode-se ver que os resultados (curvas de taxas de erro) obtidos também foram semelhantes. Assim, alcançou-se a taxa de erro maior quando o vetor de características teve  $d = 10$  coeficientes PCA, a taxa de erro total com  $d = 8$  foi maior do que a taxa de erro alcançado com  $d = 7$ . A maior desviação estândar dos resultados obtidos deu-se com  $d = 35$  (em torno do 1.5%); com o número de coeficientes PCA acima de  $d = 35$  as desviações estândar foram menores, não sendo nenhuma delas maior do que 0.5%.

### Problema de 5 classes

Na Figura 6.9 mostram-se os resultados de média e desviação estândar das taxas de erro alcançadas quando os vetores de características tiveram  $d = \{2, \dots, 64\}$  coeficientes PCA para o problema de classificação de 5 classes; nas Figuras 6.9(a) e (b) mostram-se os resultados para as distribuições natural e balanceada respectivamente; e nas Figuras 6.9(c) e (d) mostram-se os resultados obtidos apenas com o número de coeficientes PCA entre  $d = 25$  e  $d = 64$  para as distribuições natural e balanceada respectivamente.

As curvas das médias das taxas de erro de classificação que são mostradas nas Figuras 6.9(a) e (b) são semelhantes às curvas das taxas de erro alcançado quando foram usados parâmetros amostrais das distribuições dos vetores de características de cada classe como parâmetros iniciais do algoritmo EM (mostradas nas Figuras 6.5(a) e (b) respectivamente). No caso das curvas obtidas quando foram usados os centroides achados via *K-means* como parâmetros iniciais do algoritmo EM (mostradas nas Figuras 6.7(a) e (b)) também são semelhantes, especialmente quando os vetores de caraterísticas tiveram o número de coeficientes PCA acima de  $d = 25$ .

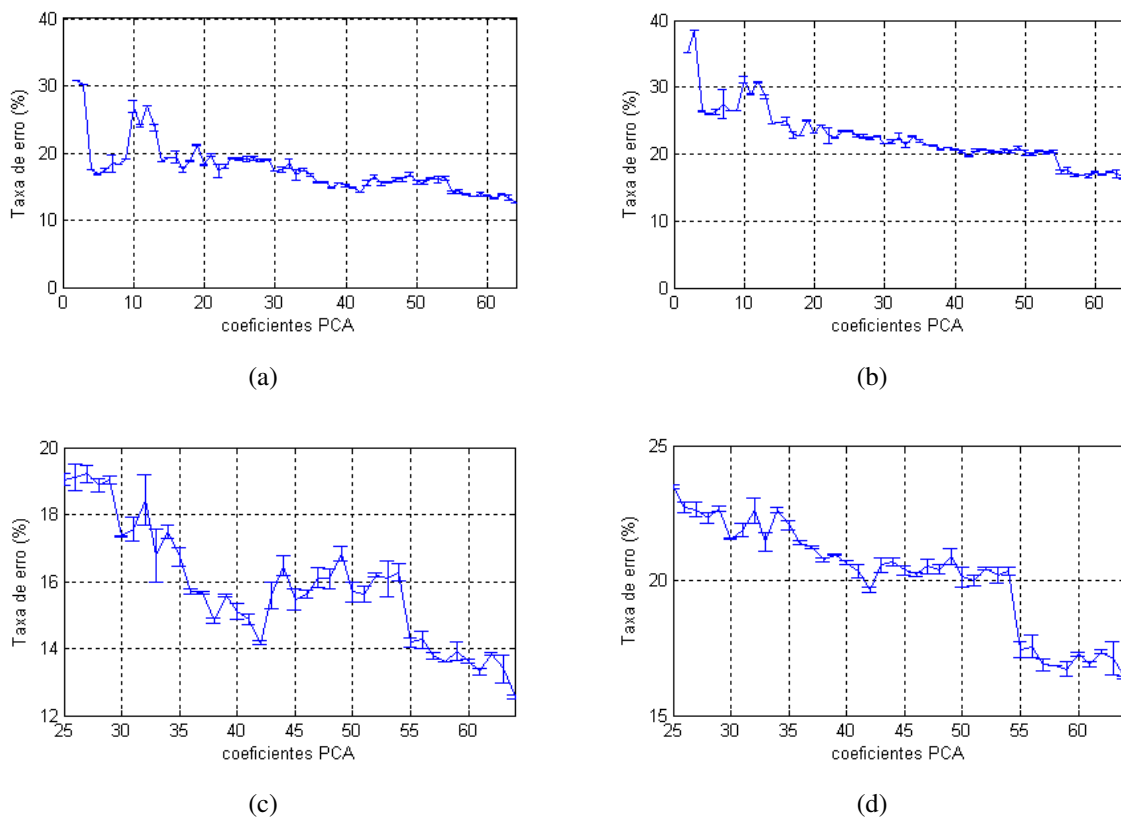


Figura 6.9: Taxas de erro modeladas como variáveis aleatórias para o problema de 5 classes quando foram consideradas as modas achadas via *Mean-Shift* como parâmetros iniciais do EM. (a) e (c) Distribuição natural. (b) e (d) Distribuição balanceada.

No que respeita às desviações estandard das taxas de erro de classificação, no caso da distribuição natural mostrada na Figura 6.9(a) alcançou-se a maior desviação estandard com  $d = 7$ . As desviações estandard obtidas com coeficientes PCA acima de  $d = 25$  são menores de 0.5 em quase todos os casos, sendo maiores apenas com  $d = \{32, 33, 63\}$  como pode-se ver na figura 6.9(c). No caso da distribuição balanceada, a maior desviação estandard foi ao redor de 1 obtida com  $d = 7$ . As desviações obtidas com coeficientes PCA acima de  $d = 25$  são menores do que 0.5 em quase todos os casos; sendo maior apenas com  $d = 63$ .

Da mesma forma do que no problema de classificação de 4 classes, no caso do problema de 5 classes também poderia-se afirmar que as curvas das taxas de erro de classificação mostrados nas Figuras 6.5(a) e (b) (para as distribuições natural e balanceada) obtidos quando foram considerados os parâmetros amostrais como parâmetros iniciais do algoritmo EM, cada uma de elas pode-se considerar mais uma realização das variáveis aleatórias amostradas nas Figuras 6.9 (a) e (b) respectivamente.

Na Tabela 6.6 mostram-se os valores das médias e as desviações estandard das taxas de erro de classificação Não-Supervisionado quando foram consideradas no treinamento e no teste vetores de características com  $d = \{6, 7, 8, 10, 25, 35, 38, 40, 60, 64\}$  coeficientes PCA.

PCA	Taxa de Erro (4 classes)				Taxa de Erro (5 classes)			
	natural		balanceada		natural		balanceada	
	média	desviação	média	desviação	média	desviação	média	desviação
6	14.02	0.0096	17.33	0.2910	17.29	0.0599	26.26	0.4157
7	15.53	0.0010	18.43	1.3479	18.65	0.0020	27.43	2.1815
8	25.71	0.0011	18.28	0.0524	26.30	0.0010	26.45	0.0433
10	29.93	0.0559	26.87	0.8984	29.75	0.0577	31.12	0.5579
25	21.57	2.4160	19.04	0.1772	22.81	2.2717	23.46	0.0808
35	13.69	1.2198	16.71	0.2724	14.40	1.6867	22.04	0.1828
38	11.53	0.1175	14.82	0.0706	12.35	0.0745	20.77	0.0791
40	11.44	0.0961	15.10	0.2411	11.57	0.1206	20.65	0.0682
60	<b>11.24</b>	0.0710	13.62	0.0707	11.74	0.1480	17.28	0.0712
64	11.48	0.0888	<b>12.55</b>	0.0462	<b>10.80</b>	0.4028	<b>16.39</b>	0.0417

Tabela 6.6: Médias e desviações estandard das taxas de de erro (%) de classificação Não-Supervisionado para diferentes quantidades de coeficientes PCA no vetor de características quando foram considerados as modas achadas via *Mean-Shift* como parâmetros iniciais do EM.

## 6.5 Resumo

Neste capítulo foram apresentados, para os três tipos de inicialização do algoritmo EM, os resultados dos testes de classificação Não-Supervisionada realizados com o classificador Bayesiano baseado nas funções de densidade de probabilidade Gaussiana estimadas sob a abordagem Não-Supervisionada.

Se observando os resultados alcançados nos problemas de classificação de 4 e 5 classes concluiu-se que as curvas das taxas de erro obtidas tanto com a inicialização do algoritmo EM com parâmetros amostrais quanto com centroides estimadas via *K-means*, podiam ser consideradas como realizações da curva da taxa de erro obtida quando o EM quando inicializado com modas estimadas via *Mean-Shift* que foi modelada como uma variável aleatória; tendo em conta que essas modas foram estimadas a partir de pontos aleatórios do espaço  $d$ -dimensional.

Tem coeficientes PCA que não contribuem com a separabilidade entre classes e tem outros que, além disso, pioram o desempenho do classificador como aconteceu com o coeficientes  $d = 10$  que fez que a taxa de erro das classes LL e RL não conseguiram melhorar os resultados obtidos por exemplo com  $d = 6$  em todos os experimentos realizados (veja Figuras (d) e (h)).

por isso, apesar de ter considerado os três tipos de inicialização no treinamento do sistema de classificação baseado na estimação de parâmetros com a abordagem Não-Supervisionada, não conseguiu-se melhorar os resultados obtidos com o classificador normal nem com o classificador baseado em MLP no trabalho prévio (veja Seção 3) como pode-se observar na Tabela 6.7.

Técnica de Classificação	Taxa de erro (5 classes)			Taxa de erro (4 classes)		
	PCA	Natural	Balanceada	PCA	Natural	Balanceada
Classificador Normal	38	7.19	6.75	38	8.34	12.3
k-NN	32	14.28	16.40	26	13.36	9.85
MLP	112	9.33	10.68	112	9.24	7.27
PNN	44	10.00	21.30	48	9.56	15.7
GMM (Parâmetros amostrais)	64	12.51	16.15	64	11.78	11,35
GMM (Centroides)	64	13.24	16.81	40	11.73	11.89
GMM (Modas)	64	10.80	16.39	64	11.48	12.55

Tabela 6.7: Comparação dos resultados obtidos com a classificação Não-Supervisionada baseada em GMM com os resultados obtidos no trabalho prévio.

# Capítulo 7

## Classificação Supervisionada usando GMBC e SMEM

Nos dois capítulos anteriores foram apresentados e avaliados os parâmetros das GMM estimados sob a abordagem Não-Supervisionado, assim, no Capítulos 5 avaliaram-se os resultados dos centroides, das modas e dos parâmetros de média e covariância estimados das distribuições dos vetores de características de cada classe. Já no Capítulo 6 foram apresentados os resultados do classificador Bayesiano implementado a partir das funções de densidade de probabilidade estimados se considerando a inicialização do algoritmo EM via os parâmetros amostrais, via os centroides e via as modas correspondentes às distribuições dos vetores de características de cada classe.

Na abordagem Supervisionado cada classe é modelada por um GMM, assim, no presente capítulo são apresentados os resultados da classificação realizada mediante o classificador GMBC, o qual foi implementado a partir dos parâmetros dos GMM de cada classe estimados via o algoritmo SMEM. Da mesma forma do que no caso Não-Supervisionado, os resultados da classificação estão dados pelas taxas de erro classificação, adicionalmente, nesta abordagem apresentou-se também o número de componentes do GMM com que o algoritmo conseguiu modelar a distribuição dos vetores de características de cada classe. Porém, nesta abordagem a inicialização do SMEM foi realizado a partir de parâmetros de média e covariância aleatórios de cada componente do GMM. Foi por esse motivo que foi preciso se modelar a taxa de erro de classificação como uma variável aleatória, tendo-se que repetir o experimento um número de vezes.

Nas seções seguintes, apresenta-se os resultados alcançados, sob esta abordagem, para os problemas de 4 e 5 classes para as distribuições natural e balanceada. No entanto, antes de se falar dos resultados, são apresentados os passos da metodologia utilizada para o treinamento e o teste de classificação.

## 7.1 Metodologia da classificação Supervisionada

O treinamento Supervisionado consiste em estimar os parâmetros dos GMM's que modelam a distribuição das impressões digitais de cada classe, e o teste consiste em determinar qual é a classe à que pertence uma impressão digital usando o classificador GMBC. Assim, para realizar o **treinamento Supervisionado** seguiram-se os seguintes passos:

1. Define-se o problema de classificação (4 ou 5 classes).
2. Divide-se o banco de imagens de impressões digitais DB4 em duas metades (uma para treinamento e outra para o teste).
3. Escolhe-se a primeira metade.
4. Depois as imagens são pre-processadas a partir do uso das técnicas descritas no Capítulo 3, assim:
  - faz-se a segmentação da impressão digital determinando as regiões da imagem conforme descrito na Seção 3.1.1,
  - faz-se a melhora da imagem conforme descrito na Seção 3.1.2,
  - determina-se a imagem de orientação conforme descrito na Seção 3.1.3,
  - determina-se o ponto de referência conforme descrito na Seção 3.1.4 para o registro, e,
  - acha-se o resultado de vetorizar todos os elementos da representação vetorial da imagem de orientação conforme descrito na Seção 3.2.1.
5. Após o pre-processado das imagens define-se o número de elementos do vetor de características.
6. Reduz-se a dimensão do vetor de características usando a técnica PCA, neste caso se conservam os rótulos.
7. Estima-se os parâmetros do GMM de cada classe separadamente usando o algoritmo SMEM conforme descrito na Seção 4.6.
8. Salva-se -como resultado do treinamento- os parâmetros do GMM estimados.

Na Figura 7.1 são apresentadas as funções de densidade de probabilidade estimadas no processo de treinamento Supervisionado usando SMEM a partir da distribuição dos vetores de características com três coeficientes PCA das imagens rotuladas para o problema de 4 classes. Naquela estimação as distribuições das classes LL, W, e RL foram modeladas por GMM's com duas componentes enquanto a classe TA-A foi modelada por um GMM de três componentes.



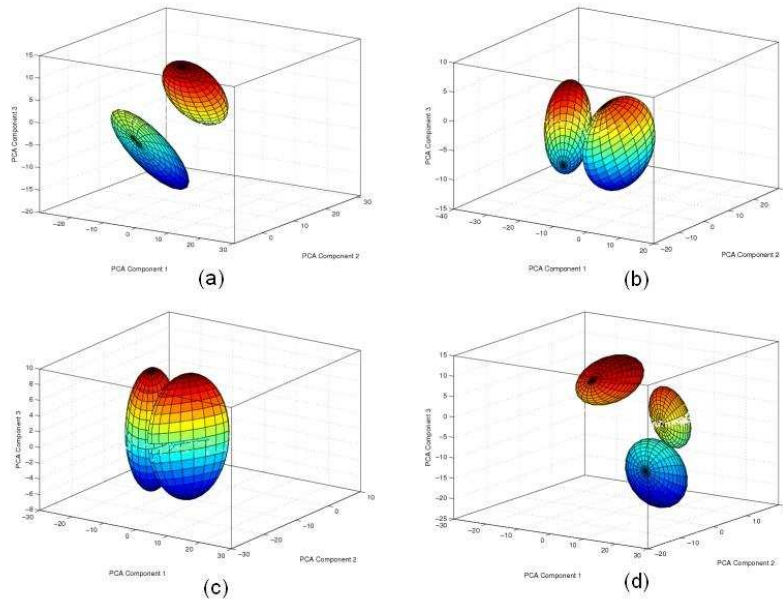


Figura 7.1: Modelos de misturas gaussianas estimadas correspondente a cada classe estimada com SMEM (a) Classe LL, (b) classe W, (c) classe RL (d) classe TA-A.

Para realizar o **teste da classificação supervisionada** usando os parâmetros dos GMM estimados para o caso de cada classe com o algoritmo SMEM:

1. Escolhe-se a segunda metade da base de dados DB4.
2. Repete-se os pre-processamento da imagem da impressão digital (passo 4).
3. Logo depois se implementa uma estrutura GMBC tendo em conta os parâmetros dos GMM estimadas para cada classe conforme descrito na Seção 4.7.
4. Finalmente avalia-se o desempenho da técnica usando a taxa de erro de classificação.

No caso correspondente aos resultados estimados, com três coeficientes PCA no vetor de características os quais são apresentados na Figura 7.1; para realizar o teste de classificação implementa-se uma estrutura GMBC (segundo a Figura 4.6) constituída por quatro sub-redes, a primeira segunda e terceira sub-redes constituídas por GMM de dois componentes que modelam as distribuições das classes LL, W e RL respectivamente. No caso da distribuição da classe TA-A, a quarta sub-rede teria o GMM com três componentes.

## 7.2 Resultados da classificação Supervisionada

Nesta seção são apresentados os resultados da classificação Supervisionada sobre as impressões digitais de teste usando o classificador GMBC descrito na Seção 4.7. Os parâmetros das misturas Gaussianas que modelam cada classe foram estimados via o algoritmo SMEM descrito na seção 4.6 o qual realiza a junção e divisão iterativamente dos componentes do GMM a partir de uma quantidade de componentes definido com parâmetros de media, covariância e coeficiente de ponderação aleatórios. Devido a este tipo de inicialização aleatória, no caso da classificação Supervisionada modelou-se o resultado da classificação (dada pela taxa de erro) como uma variável aleatória. Assim, os experimentos foram repetidos dez vezes, tanto no caso da distribuição natural quanto da distribuição balanceada para os problemas de 4 e 5 classes. Nas Figuras 7.2 e 7.4 são mostrados os resultados da média e a desviação estândar das taxas de erro de classificação obtidos usando vetores de características com  $d = \{1, \dots, 64\}$  coeficientes PCA para o problema de 4 e 5 classes respectivamente.

### Problema de 4 classes

No caso da apresentação dos resultados para o problema de 4 classes considerando a **distribuição natural** percebe-se na, Figura 7.2(a), que as médias das taxas de erro de classificação alcançaram o valor de 10.44% com apenas  $d = 6$  coeficientes PCA no vetor de características. Com coeficientes acima de  $d = 6$  as médias da taxa de erro de classificação ficaram com valores em torno de 10% e as desviações estândar foram menores do que 1. Porém, com  $d = 31$  a média das taxas de erro alcançou o valor de 8,63% e permaneceu sem maiores variações até  $d = 50$  e as desviações foram menores do que 0.5. O valor mínimo da média da taxa de erro (8.36%) foi alcançado com  $d = 53$ . Estes valores podem ser conferidos na Tabela 7.1.

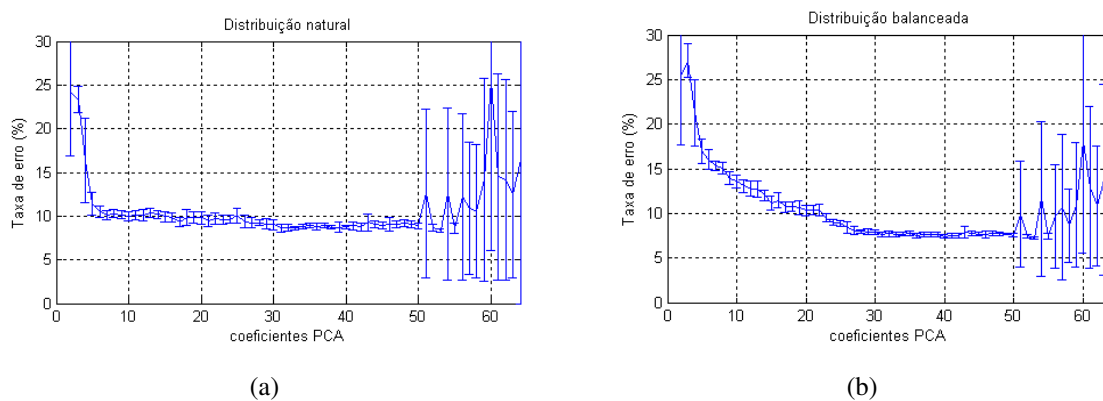


Figura 7.2: Taxas de erro de classificação Supervisionada modeladas como variáveis aleatórias para o problema de 4 classes considerando as distribuições: (a) natural e (b) balanceada.

Na Figura 7.2(a) também pode-se ver que os valores de média e desviação estandar obtidas com  $d = 51$  foram altas, isto indica que em pelo menos um dos dez experimentos realizados com esta quantidade de coeficientes a taxa de erro foi alta. Obteve-se resultados parecidos, muito altos, nos casos em que usaram-se na classificação vetores de características com  $d = \{2, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64\}$  coeficientes PCA.

No referente aos resultados obtidos para o problema de 4 classes considerando a **distribuição balanceada** pode-se ver, na Figura 7.2(b), que as médias das taxas de erro de classificação decresceram monotonicamente até  $d = 50$ . Assim, alcançou-se o mínimo valor (7.26%) com  $d = 53$  coeficientes PCA no vetor de características. As desviações estandar das taxas de erro foram menores no intervalo de  $28 < d < 50$ ; de igual forma do que na distribuição natural, os valores podem ser conferidos na Tabela 7.1; nesse gráfico também pode-se ver que no caso balanceado a média e a desviação estandar obtidas com  $d = \{2, 51, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64\}$  também foram altas.

Nos casos em que os valores de média e variância das taxa de erro de classificação foram altas, para o problema de classificação de 4 classes considerando a distribuição balanceada, pode-se inferir que o algoritmo SMEM não conseguiu modelar um GMM que aproxime à distribuição dos vetores de características das impressões digitais de uma ou mais classes. Por esse motivo, na Figura 7.3 são amostrados os resultados alcançados em cada classe. Nos gráficos da coluna da esquerda são mostradas as médias e desviações estandar das taxas de erro de classificação e na coluna da direita de forma correspondente são amostrados a quantidade de componentes do GMM. No caso das componentes estes não foram modeladas como variáveis aleatórias, senão, foram considerados o maior número de componentes estimado entre dos dez experimentos.

Nos resultados da classe LL mostrados na Figura 7.3(a) pode-se ver que as menores taxas de erro foram alcançadas quando as distribuições dos vetores de características dessa classe tiveram  $d = \{4, \dots, 17\}$  coeficientes PCA. Na 7.3(b) percebe-se que estes resultados foram obtidos quando a distribuição foi modeladas com um GMM de 3 componentes. Com coeficientes PCA acima de 17 a média da taxa de erro ficou em torno de 10%, precisando-se de 2 componentes para modelar tal classe, e acima de  $d = 28$  precisou-se de apenas uma componente em quase todos os casos (só em  $d = 30$  e  $d = 40$  precisaram-se 2 e 3 componentes respectivamente). Porém, quando  $d = \{51, 54, 57, 58, 59, 60, 61, 62, 63, 64\}$  as médias e as desviações estandar foram altas. A quantidade de componentes em estos casos foi igual do que à quantidade inicial, indicando que o algoritmo SMEM não conseguiu juntar nem separar nenhuma componentes e, por isso, ocasionando que o erro de classificação fosse alta.

No caso da classe W pode-se ver, nas Figuras 7.3(c)-(d) que as menores taxas de erro foram alcançadas quando as distribuições dos vetores de características dessa classe tiveram vetores com acima de 30 coeficientes PCA e foram modelas com apenas uma componente

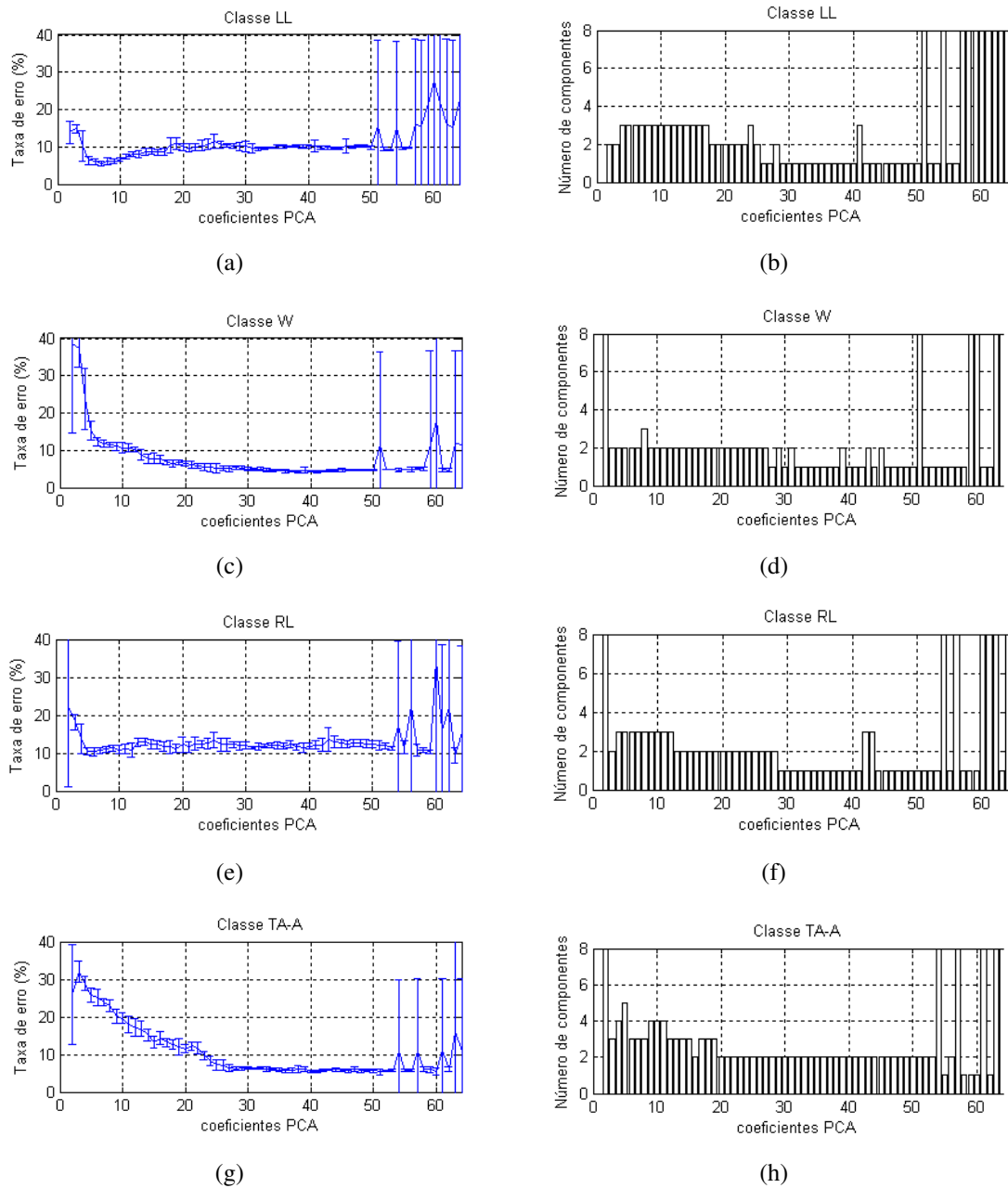


Figura 7.3: Resultados da classificação Supervisionada para o problema de 4 classes considerando a distribuição balanceada avaliadas para  $n = \{2, \dots, 64\}$  coeficientes PCA. Coluna da esquerda: médias e desvios estândar da taxas de erro. Coluna da direita: número de componentes gaussianas que modelam cada classe.

na maioria dos casos. Porém nos casos de  $d = \{2, 51, 59, 60, 63, 64\}$  os valores de média e desviação estândar foram altas e a quantidade de componentes nestes casos foi 8, o número inicial de componentes que utiliza o algoritmo.

Nos resultados da classe RL, nas Figuras 7.3(e) e (f), as médias das taxas de erro menores foram alcançadas quando as distribuições dos vetores de características tiveram  $d =$

$\{3, \dots, 12\}$  coeficientes PCA e foram modeladas com 3 componentes. Com o número de coeficientes PCA acima de 12 a média da taxa de erro ficou em torno de 12%, precisando-se de 2 funções de densidade de probabilidade Gaussiana para modelar tal classe e de apenas uma componente quando a quantidade de coeficientes PCA foi maior do que 28. Porém com  $d = 42$  e  $d = 43$  precisaram-se 3 funções de densidade de probabilidade Gaussianas.

Na classe TA-A, nos resultados mostrados nas Figuras 7.3(g)-(h) pode-se ver que a média da taxa de erro decresceu enquanto incrementou-se o número de coeficientes PCA no vetor de características; assim, o valor da média da taxa de erro, com 2 componentes na maioria dos casos, ficou em torno de 7% sem sofrer maior variação com valores de  $d$  acima de 26%. No gráfico também pode-se ver a média e a desviação estandard obtidas com  $d = \{54, 57, 61, 63, 64\}$  foram altas.

Os resultados obtidos nas classes RL e LL apresentaram semelhanças entre si tanto no gráfico das médias e desviações estandard das taxas de erro quanto no gráfico do número de componentes que modelaram estas classes. No caso das classes LL, W e RL, os experimentos com vetores de características com coeficientes PCA acima de 30 precisaram, geralmente, apenas uma componente para modelar as distribuições. No caso da distribuição das impressões digitais da classe TA-A precisaram-se de 2 componentes. Em todas as classes, o algoritmo SMEM teve problemas para modelar as classes quando a quantidade de coeficientes PCA no vetor de características foi alto ( $d > 50$ ).

Na Tabela 7.1 são apresentados os valores de média e desviações estandard das taxas de erro de classificação Supervisionada para os problemas de 4 e 5 classes.

PCA	Taxa de Erro (4 classes)				Taxa de Erro (5 classes)			
	natural		balanceada		natural		balanceada	
	média	desviação	média	desviação	média	desviação	média	desviação
3	23.33	1.4708	27.09	1.8639	23.57	0.5655	32.48	1.0250
5	11.43	1.2809	16.96	1.3564	11.54	0.3111	23.07	0.7363
6	10.44	0.6550	16.01	1.1215	11.25	0.3528	21.67	0.6902
10	9.95	0.5541	13.61	0.7256	10.71	0.3096	19.91	0.6543
20	9.82	0.6917	10.31	0.5559	11.27	0.9517	17.88	1.0862
28	9.25	0.2831	8.01	0.1951	9.73	0.6242	14.28	1.2950
31	8.63	0.4836	7.63	0.2846	<b>8.65</b>	0.1822	13.12	0.2449
40	8.76	0.1977	7.41	0.1669	9.32	0.0010	12.85	0.0030
53	<b>8.36</b>	0.2443	<b>7.26</b>	0.1574	9.13	0.0020	<b>12.75</b>	0.0025

Tabela 7.1: Médias e desviações estandard das taxas de erro (%) de classificação Supervisionada para diferentes quantidades de coeficientes PCA no vetor de características

## Problema de 5 classes

Os resultados da classificação para a **distribuição natural** são amostrados Figura 7.4(a), nela percebe-se na que as médias das taxas de erro de classificação alcançaram a valor de 10.71% com  $d = 10$  coeficientes PCA no vetor de características. Mas, com valores acima de  $d = 6$  coeficientes PCA as médias ficaram com valores em torno de 10% até  $d = 30$  e as desviações estândar foram menores do que 1. Com  $d = 31$  a média das taxas de erro alcançou o valor mínimo (8,63%) e permanece sem maiores variações até  $d = 53$ . Em  $d = 54$  percebe-se que os valores de média e desviação estândar foram altas; isto, da mesma forma do que no problema de 4 classes, é devido a que em pelo menos um dos experimentos realizados com esta quantidade de coeficientes a taxa de erro foi alta. pode-se ver que os resultados da classificação também foram altos com quase todos os valores de  $d$  acima de 54.

No caso da apresentação dos resultados para o problema de 5 classes considerando a **distribuição balanceada** pode-se ver, na Figura 7.4(b), que as médias das taxas de erro de classificação decrescem monotonicamente até  $d = 31$ . Acima de esta quantidade de coeficientes, as médias das taxas de erro ficaram com valores em torno de 13% com desviações estândar menores do que 1, e alcançou-se o valor mínimo da média com  $d = 53$  coeficientes PCA no vetor de características. Podem-se conferir os valores na Tabela 7.1.

De forma similar do que no caso da distribuição balanceada no gráfico também pode-se ver que a média e a desviação estândar obtidas com valores de  $d$  acima de 54 foram altas. Na Figura 7.5 são mostradas as médias e desviações estândar das taxas de erro de classificação e a quantidade de funções de densidade de probabilidade Gaussiana que foram estimadas pelo algoritmo SMEM das classes separadas para o problema de classificação de 5 classes considerando a distribuição balanceada. Pode-se observar nas Figuras 7.5(a)-(b) para o caso da classe LL, nas Figuras 7.5(c)-(d) para o caso da classe W e nas Figuras 7.5(e)-(f) para o caso da classe RL que os resultados foram muito semelhantes aos obtidos no problema

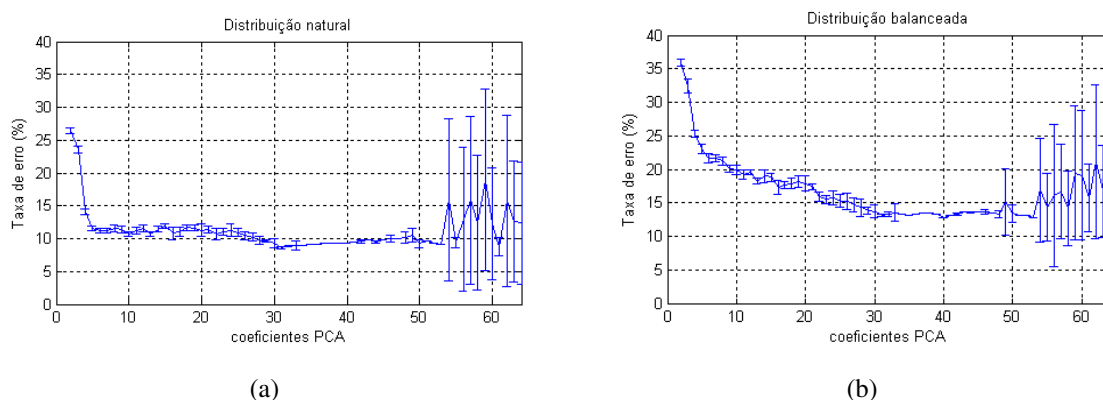


Figura 7.4: Taxas de erro de classificação Supervisionada modeladas como variáveis aleatórias para o problema de 5 classes considerando as distribuições: (a) natural e (b) balanceada..

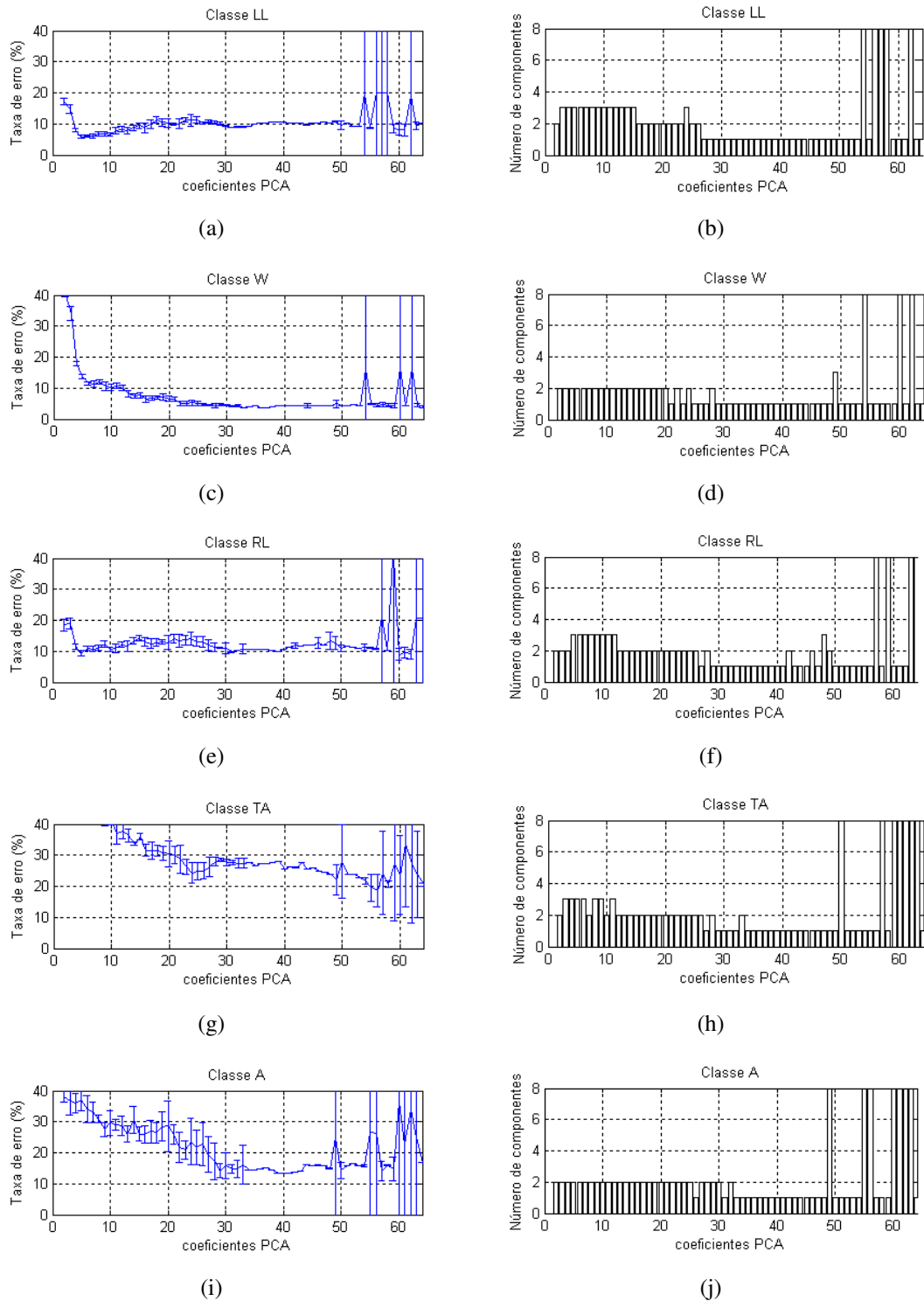


Figura 7.5: Resultados da classificação Supervisionada para o problema de 5 classes considerando a distribuição balanceada avaliadas para  $d = \{2, \dots, 64\}$  coeficientes PCA. Coluna da esquerda: Médias e desvios estândar da taxas de erro. Coluna da direita: Número de componentes gaussianas que modelam cada classe.

de 4 classes, tanto na média das taxas de erro e desvios estandard quanto na quantidade de componentes do GMM que modelam cada classe e são mostradas nas Figuras 7.2(a)-(b), 7.2(c)-(d), 7.2(e)-(f), respectivamente. No entanto, pode-se perceber que nas classes TA e A as médias das taxas de erro foram altas.

Assim, na classe TA os valores de média das taxas de erro em todos os casos ficaram acima de 20%, ainda que nos resultados dos experimentos obtiveram-se GMM's com 1, 2 ou 3 componentes. Comparando com as outras classes, esta foi a que apresentou maior taxa de erro. No caso da classe A os valores das taxas de erro ficaram em torno de 15% com o número de coeficientes PCA acima de  $d = 29$ . As desvios estandard foram consideravelmente altas com o número de coeficientes abaixo de 33, assim, as desvios com  $d$  acima desta quantidade de coeficientes foram menores e precisaram-se de apenas uma componente para modelar esta classe e obter uma taxa de erro alta.

Nestes dois últimos casos, poderia-se fazer questão do desempenho do algoritmo SMEM na estimação de parâmetros dos GMM que modela cada uma destas classes, mas, é preciso lembrar que as classes TA e A tem pequena variação inter-classe, o que origina que, apesar de se modelar de forma eficiente a distribuição das classes por separado, haveriam impressões digitais que pertencem com igual probabilidade a qualquer uma destas classes, o que influiriam no baixo desempenho de classificação no problema de 5 classes.

### 7.3 Resumo

As taxas de erro mínimas obtidas na classificação Supervisionada foram menores do que as taxas obtidas na classificação Não-supervisionada como pode-se ver, se compara-se as médias das taxas de erro obtidas, na Tabela 7.1 com as médias dos resultados mostrados na Tabela 6.6 e as taxas de erro mostradas nas Tabelas 6.1, 6.3 e 6.5 do capítulo 6. Isto acontece porque na abordagem Supervisionada a distribuição dos vetores de características de uma classe de impressões digitais no espaço  $d$ -dimensional foi modelada por um conjunto de funções de densidade de probabilidade Gaussiana, no entanto, na abordagem Não-Supervisionada uma classe foi modelada com apenas uma função de densidade de probabilidade Gaussiana.

A vantagem que pode-se perceber no classificador baseado no GMBC frente aos classificadores propostos no trabalho prévio é que alcançam-se taxas de erro pequenas com poucos coeficientes PCA, assim por exemplo na Tabela 7.2 observa-se que com apenas 10 coeficientes PCA obteve-se resultados comparáveis com os obtidos com o MLP (com 112 coeficientes PCA) e melhores do que os obtidos com o  $k$ -NN.

As taxas de erro das classes LL, W e RL tanto no problema de 4 quanto o problema de 5



classes alcançam valores baixos com vetores de características com poucos coeficientes PCA (menores do que 20), porém as classes TA e A, no problema de 5 classes, não conseguiram alcançar taxas de erro menores do que 20% no caso do TA nem menores do que 12% no caso do A. Isto ocorre porque, além de que as funções de densidade de probabilidade tenham sido estimadas com eficiência, essas classes têm pequena variação inter-classes.

Por outro lado, em todos os experimentos o algoritmo SMEM teve problemas para modelar as distribuições quando a quantidade de coeficientes PCA no vetor de características foi alto, geralmente maiores do que 50, porém o problema não tão crítico, já que nesta abordagem os resultados mais importantes foram obtidos quando o número de coeficientes PCA foi pequeno.

Técnica de Classificação	Taxa de erro (5 classes)			Taxa de erro (4 classes)		
	PCA	Natural	Balanceada	PCA	Natural	Balanceada
Classificador Normal	38	7.19	6.75	38	8.34	12.3
k-NN	32	14.28	16.40	26	13.36	9.85
MLP	112	9.33	10.68	112	9.24	7.27
PNN	44	10.00	21.30	48	9.56	15.7
GMBC (Supervisionado)	10	10.71	19.91	10	9.95	13.61
GMBC (Supervisionado)	31	8.65	13.12	31	8.63	7.63
GMBC (Supervisionado)	53	9.13	12.75	53	8.36	7.26

Tabela 7.2: Menores taxas de erro de classificação Supervisionada comparadas com os resultados alcançados no trabalho prévio para os problemas de 4 e 5 classes considerando as distribuições balanceada (B) e natural (N) e a se indicando a quantidade de coeficientes PCA. (Samatelo, 2007)

# Capítulo 8

## Conclusões e Trabalhos Futuros

No presente trabalho buscou-se avaliar, para a arquitetura de classificação de impressões digitais conhecida como PCASYS, se seria possível assumir que a distribuição dos vetores de características das classes do sistema de classificação Galton-Henry (LL, W, RL, TA e A) poderiam ser modeladas usando funções de densidade de probabilidade Gaussiana. Para tal fim foi proposto se empregar a técnica conhecida como GMM sob dos tipos de abordagens: (1º) Abordagem Não-Supervisionado, na qual buscou-se modelar a distribuição dos vetores de características das impressões digitais de treinamento pela combinação linear de um conjunto de funções de densidade de probabilidade Gaussiana (dada por um GMM) de forma que a distribuição de cada classe fosse modelada por uma função de densidade. (2º) Abordagem Supervisionado, na qual buscou-se modelar a distribuição dos vetores de características das impressões digitais de treinamento por um conjunto de GMMs, de forma que a distribuição dos vetores de características de cada classe fosse modelada por um GMM.

Na classificação com abordagem Não-Supervisionado os parâmetros das funções de densidade de probabilidade que compõem o GMM foram estimados usando o algoritmo de estimação EM e a partir destes resultados foi implementado um classificador Bayesiano. O desempenho do classificador foi avaliado mediante a taxa de erro de classificação. Devido a que a inicialização do algoritmo EM não foi aleatório (senão que foram propostas inicializações com parâmetros amostrais, com centroides estimados via o algoritmo *K-means* e com modas estimadas via *Mean-Shift*) foi preciso fazer a avaliação destes parâmetros via o teste de hipótese. Adicionalmente mediu-se a entropia mutua mediante a divergência (de Kullback-Leibler) entre as funções de densidade de probabilidade Gaussianas de parâmetros (média e covariância) estimadas com EM versus as funções de densidade de probabilidade Gaussianas de parâmetros amostrais.

Na classificação com abordagem Supervisionado os parâmetros do GMM de cada classe foram estimados usando o algoritmo SMEM e a partir destes resultados foi implementado

o classificador chamado GMBC. Os resultados da classificação também foram avaliados usando a taxa de erro de classificação, mas dessa vez foram considerados adicionalmente o número de componentes Gaussianas do GMM. Nesta abordagem a inicialização do algoritmo SMEM foi aleatória, salvo no caso do número de componentes iniciais do GMM de cada classe, que foi de 8 para todos os casos.

Com o objetivo de se comparar os resultados obtidos neste trabalho com outros resultados do estado da arte na classificação de impressões digitais, todos os experimentos, tanto para a abordagem Não-Supervisionada quanto para a abordagem Supervisionada, foram realizados para distintos casos nos quais os vetores de características tiveram 2,3,.. ou 64 coeficientes PCA para os problemas de classificação de 4 e 5 classes e considerando as distribuições natural e balanceada. Na Tabela 8.1 são amostrados os resultados obtidos (taxa de erro) neste trabalho juntamente com algum dos resultados obtidos por outros autores usando outras técnicas de processamento e classificação.

Abordagem	5 classes		4 classes		Conjunto de teste	Ano
	B	N	B	N		
Blue et al.	-	7.2	-	-	segunda metade	1993
Candela et al.	-	-	11.4	6.1	segunda metade	1995
Karu e Jain	14.6	11.9	8.6	9.4	tudo	1996
Senior	-	-	-	8.4	542 aleatórias	1997
Jain et al.	10.0	7.0	5.2	-	segunda metade	1999
Cappelli et al.	-	12.9	-	-	segunda metade	1999
Cappelli et al.	7.9	6.5	5.5	-	Data set 2	1999
Hong e Jain	12.5	10.6	7.7	-	tudo	1999
Senior	-	-	-	5.1	segunda metade	2001
Yao et al.	10.7	9.0	6.9	-	segunda metade	2001
Marcialis et al.	12.1	9.6	-	-	segunda metade	2001
Jain e Minut	-	-	8.8	9.3	tudo	2002
Zhang e Yan	15.7	-	7.3	-	tudo	2004
Dass e Jain	-	-	5.6	-	tudo	2004
Tan et al.	-	-	6.1	-	segunda metade	2005
Park e Park	9.3	-	6.0	-	tudo	2005
(Samatelo, 2007)	12.3	8.34	6.75	7.19	segunda metade	2007
(Liu et al., 2007)	4.2	2.9	3.1	2.9	tudo	2007
<b>Classificação Não-Supervisionada</b>	<b>16.4</b>	<b>10.8</b>	<b>12.5</b>	<b>11.5</b>	segunda metade	2010
<b>Classificação Supervisionada</b>	<b>12.7</b>	<b>9.1</b>	<b>7.3</b>	<b>8.4</b>	segunda metade	2010

Tabela 8.1: Comparação com resultados atuais, das Taxas de erro de classificação alcançadas neste trabalho. Informação baseada em (Maltoni et al., 2003), (Maltoni et al., 2009), (Samatelo, 2007), (Liu et al., 2007).

Em seguida são apresentadas as conclusões extraídas a partir dos resultados alcançados tanto na abordagem Não-Supervisionado quanto a abordagem Supervisionado que foram mostrados nos capítulos anteriores. Assim mesmo apresentam-se as propostas dos trabalhos futuros com o objetivo de obter melhores resultados de classificação de impressões digitais do que os alcançados neste trabalho.

## 8.1 Conclusões

1. Pode-se concluir com respeito à classificação de impressões digitais com a abordagem Não-Supervisionado que:
  - (a) Sob a abordagem Não-Supervisionado não conseguiu-se alcançar melhores resultados do que os obtidos no trabalho prévio (Samatelo, 2007), já que a menor taxa de erro obtida nesta abordagem foi de 10.8% (veja a Tabela 8.1) para o problema de 5 classes com distribuição balanceada.
  - (b) Isto indicaria que uma função de densidade de probabilidade Gaussiana não pode representar a distribuição dos vetores de características de impressões digitais de todas as classes tanto no problema de 4 classes quanto do 5 classes, apesar de que o algoritmo de estimação EM não foi inicializado com parâmetros aleatórios (senão com média e covariância amostrais, com centroides estimadas com *K-means* e com modas estimadas com *Mean-Shift*) conforme descrito nas seções 6.2, 6.3 e 6.4 respectivamente.
  - (c) No entanto, se observando os resultados alcançados em cada classe ao invés do que os resultados totais, tanto no problema de 4 classes quanto do 5 classes, pode-se ver que a distribuição dos vetores de características da classe W pode ser bem representada por uma função de densidade de probabilidade Gaussiana com respeito às outras classes, já que foi nesta classe que obteve-se menores taxa de erro (veja as Figuras 6.5 e 6.5) e menor divergência de Kullback-Leibler (veja a Figura 5.6) do que nas outras classes. Assim mesmo, no teste de hipótese a média estimada com EM ficou dentro da elipse de confiança da média amostral quando o vetor de características teve acima de 38 coeficientes PCA, como viu-se na Figura 5.2.
  - (d) Pelo contrario, no problema de classificação de 4 classes, foi na classe TA-A que obtiveram-se as maiores taxas de erro de classificação (veja a Figura 6.5(j)) e a média estimada via EM ficou fora da elipse de confiança para todos os casos de avaliados (veja a Figura 5.3(f)). Isto indicaria que o fato de se considerar os vetores de características das classes TA e A como se fossem de uma classe só, faz que uma função de densidade de probabilidade não possa modelar a distribuição

com eficiência. Percebe-se também que a entropia mútua nesta classe foi maior do que nas classes TA e A separadas (veja a Figura 5.6(a)).

- (e) Porém, o resultado da classificação Não-Supervisionada das impressões digitais das classes TA e A separadas, no problema de 5 classes, também não foi satisfatório, já que as taxas de erro (veja as Figuras 6.5(j) e 6.5(l)) foram altas e a entropia mutua ,ainda que foi menor do que na classe TA-A, foi alta se comparadas com as outras classes (veja a Figura 5.6(b)); o que levaria se concluir que ainda que uma função de densidade de probabilidade Gaussiana possa modelar estas duas classes com eficiência, o fato de ter pequena variação interclasses faz que o desempenho na classificação seja baixo.
- (f) Os resultados obtidos nas classes LL e RL são muito semelhantes entre eles, o que tem coerência com a semelhança das distribuições de cristas e vales nas áreas padrão das impressões digitais destas classes (veja a Figura 2.4). São nos resultados obtidos nestas classes que pode-se perceber claramente que a taxa de erro de classificação foi maior quando teve-se 10 coeficientes PCA no vetor de características do que quando teve-se apenas 6. Isto indicaria que aumentar a dimensão do vetor de características não implica necessariamente melhorar o desempenho na classificação de impressões digitais (veja as Figuras 6.5(d), 6.5(h), 6.5(d) e 6.5(h)).

É assim que pode-se ver que todas as classes não podem ser razoavelmente bem ser representadas por uma distribuição gaussiana multivariável como foi suposto no trabalho prévio, pelo menos nos casos avaliados neste trabalho. Apenas a distribuição dos vetores de características da classe W pode ser descrito e modelado com eficiência por uma função de densidade de probabilidade Gaussiana, as distribuições das classes LL e RL podem-se aproximar de forma razoável, mas, não foi obter bons resultados na classificação das impressões digitais das classes TA-A no problema de 4 classes nem das classes TA e A no problema de 5 classes.

Um outro aspecto importante na abordagem Não-Supervisionado foi que, visto nos resultados das classes LL e RL (que influenciam no resultado total) percebeu-se que tem coeficientes PCA que não contribuem com a separabilidade entre impressões digitais de distintas classes.

2. Com respeito à classificação de impressões digitais com a abordagem Não-Supervisionado pode-se concluir que:

- (a) Os resultados alcançados nesta abordagem foram melhores dos que os obtidos na abordagem Não-Supervisionado, pois a menor taxa de erro de classificação foi de 7.3% para o problema de 4 classes com distribuição balanceada. Mas, não

conseguiu-se melhorar os resultados da classificação obtidos no trabalho prévio (veja a Tabela 8.1).

- (b) No entanto, sob esta abordagem conseguiu-se obter taxas de erro relativamente pequenas com um número reduzido de coeficientes PCA no vetor de características como pode-se ver claramente na Figura 7.2(a). Na Tabela 7.1 pode-se ver, para o problema de 4 classes e distribuição natural, que com apenas 10 coeficientes PCA no vetor de características obteve-se uma taxa de erro (média) de 9.95%. Isto pode ser considerada uma vantagem desta abordagem visto que para obter resultados semelhantes com técnicas alternativas ( $k$ -NN, MLP e PNN) apresentadas no trabalho prévio precisaram-se de grande quantidade de elementos no vetor de características (veja a Tabela 7.2).
- (c) Observando-se os resultados obtidos em cada classe, pode-se ver que a distribuição dos vetores de características da classe LL pode ser aproximada de forma razoável com apenas uma função de densidade de probabilidade Gaussiana com um número de coeficientes PCA acima de 30. Isto tem coerência com os resultados da abordagem Não-Supervisionado, mas, nas Figuras 7.3(a) e 7.3(b) no problema de 4 classes, e nas 7.4(a) e 7.4(b) no problema de 5 classes, percebe-se que a menor taxa de erro nesta classe foi obtida quando a distribuição foi modelada com um GMM de três componentes e com uma quantidade muito reduzida de coeficientes PCA. Os resultados obtidos na classe RL foram muito semelhantes a estes resultados (veja as Figuras 7.3(e), 7.3(f), 7.4(e) e 7.4(f)).
- (d) Por outro lado, a distribuição dos vetores de características da classe W pode ser modelada com apenas uma função de densidade de probabilidade Gaussiana com coeficientes PCA acima de 30. Com menor quantidade de coeficientes precisou-se GMMs de mais de uma componente para modelar a distribuição, mas, o erro de classificação foi maior.
- (e) No problema de 4 classes, os resultados da classificação na classe TA-A foram muito melhores do que os obtidos na abordagem Não-Supervisionado, já que obteve-se as taxas de erro pequenas, inclusive menores do que os obtidos nas classes LL e RL, quando a distribuição dos vetores de características da classe TA-A foi modelada com um GMM de duas componentes (veja as Figuras 7.3(g) e 7.3(h)). Isto indica que esta classe TA-A é melhor representada com duas funções de densidade de probabilidade Gaussiana do que com apenas uma.
- (f) Poder-se-ia dizer que no problema de 5 classes, o algoritmo SMEM não conseguiu modelar com sucesso as classes TA nem A, já que observa-se que as taxas de erro de classificação da classe TA são muito altos (acima de 20%) apesar que foram modeladas com três, duas ou uma componente Gaussiana no GMM (veja as Figuras 7.4(g) 7.4(h)). No caso da classe A os erros foram menores do que na

classe TA, mas ainda permaneceram altas (em torno de 15%) com uma componente no GMM (veja as Figuras 7.4(i) 7.4(j)). Porém poder-se-ia inferir também que estas altas taxas de erro de classificação foram originados, ou pelo menos influenciados, pela pequena variabilidade inter-classe que existe entre as duas classes.

Nesta abordagem, é preciso anotar que o algoritmo SMEM modificado teve problemas para convergir quando o vetor de características teve uma quantidade alta de coeficientes PCA; porém a falha, ainda que possa ser corrigida, não é muito grave desde que a provas foram focalizadas sobre vetores de características com a menor quantidade possível de elementos. É assim que tem classes que podem ser modeladas de forma eficiente com uma quantidade pequena de coeficientes PCA no vetor de características.

## 8.2 Trabalhos futuros

Os trabalhos que se apresentam em seguida foram propostos com o objetivo de se desenvolver um sistema de classificação de impressões digitais eficiente a partir dos resultados alcançados neste trabalho e baseado na classificação de Galton-Henry, com arquitetura PCASYS, usando os coeficientes PCA como elementos do vetor de características e trabalhando com todas as impressões digitais do banco de dados DB4 sem considerar técnicas de seleção prévia nem rejeição de impressões digitais.

Devido a que foram alcançados melhores resultados com a classificação baseado em GMM sob abordagem Supervisionado, propõe-se:

1. Visto que as taxas de classificação alcançadas tiveram relação com o número de coeficientes PCA e com o número de componentes no GMM, propõe-se buscar uma combinação eficiente entre estes parâmetros que permitam obter a menor taxa de erro de classificação usando o banco de dados DB4 para cada classe. Para isto poder-se-ia empregar técnicas de baseadas em algoritmos evolutivos na qual a população estaria constituída por todas as combinações possíveis do número de coeficientes PCA e do número de componentes de cada mistura. A função objetivo estaria dada pela taxa de erro de classificação. Porém, no contexto de reconhecimento de padrões seria recomendável pôr um limite ao número de componentes, já que, ainda que um GMM com grande quantidade de componentes Gaussianas representaria melhor a distribuição de vetores de características de uma classe, este modelo poderia ficar fixada no conjunto de dados e particularizar o treinamento.

2. Por outro lado, visto que com um reduzido número de coeficientes PCA no vetor de características sob a abordagem Supervisionada obtiveram-se resultados que foram semelhantes aos obtidos com uma grande quantidade de elementos sob outras técnicas de classificação; poder-se-ia reduzir a taxa de erro de classificação obtida com o GMBC estimada a partir de vetores de característica de baixa dimensão (com um pequeno número de coeficientes PCA como por exemplo 6) se implementando uma rede neural probabilística baseada em decisão (*Probabilistic Neural Networks Decision Based*) (Tseng et al., 2004) tendo como pesos iniciais precisamente os parâmetros do GMBC. A vantagem desta proposta seria diminuir o tempo de processamento e principalmente de classificação.
3. Tendo em conta os resultados alcançados na abordagem Não-Supervisionado, poder-se-ia realizar a seleção e rejeição dos elementos do vetor de características -dado pelos coeficientes PCA- que não contribuem com a separabilidade entre impressões distintas de classes distintas. Isto, no caso da abordagem Não-Supervisionado, diminuiria a taxa de erro de classificação se precisando menor quantidade de coeficientes PCA; e no caso da abordagem Supervisionado diminuiria o tempo de estimação de parâmetros e possivelmente a quantidade de componentes do GMM de cada classe.

Acredita-se que se fosse possível desenvolver um sistema de classificação considerando todas essas propostas, conseguiria-se obter resultados na classificação de impressões digitais semelhantes aos alcançados atualmente no estado da arte.



# Apêndice A

## Interface de usuário

Os algoritmos propostos neste trabalho foram desenvolvidos usando uma linguagem de computação científica chamado MATLAB de MathWorks se aproveitando as facilidades de implementação que proporciona esta linguagem. MATLAB possui funções de leitura e escrita de arquivos de imagens que permitem administrar de forma eficiente um banco de dados de impressões digitais. Também possui funções de desenvolvimento de interfaces gráficas de usuário denominada GUIDE (*Graphic User Interface Development Enviroment*).

Na Figura A.1 mostra-se a interface gráfica de usuário desenvolvida neste trabalho se usando o GUIDE de MATLAB. A interface permite fazer a gestão dos algoritmos utilizados neste trabalho, e estão organizados da seguinte forma em quatro blocos de algoritmos: Ver impressão, Extração de características, Classificação e Classificação com GMM.

O bloco **Ver impressão** está constituído pelas funções que permitem ler as imagens e mostrar na tela do computador as impressões digitais do banco de dados DB4.

O bloco **Extração de características** está constituído pelos algoritmos de processamento e extração de características propostos no trabalho prévio (Samatelo, 2007). Os argumentos que precisam estes algoritmos estão dados pelo tipo de problema de classificação, o número de imagens para o treino (pre-estabelecido em 2000), o tipo de redutor (pre-estabelecido na técnica de PCA) e a dimensão do vetor de características que pode ser escolhida pelo usuário.

O bloco **Classificação** está constituído pelos algoritmo que realizam o treinamento e teste da classificação das impressões digitais de teste mediante técnicas de conhecidas. Os argumentos que precisam estes algoritmos estão dados pelo tipo de classificador (que pode ser BAYES, KNN, PNN, BPNN, MLP ou LVQ), o tipo de problema de classificação (que pode ser 4 ou 5 classes) e pela dimensão do vetor de característica que pode ser escolhida pelo usuário.



Figura A.1: Interface gráfica de usuário desenvolvido em MATLAB.

O bloco **Classificação com GMM** está constituído pelos algoritmos desenvolvidos e utilizados neste trabalho. Os argumentos que precisam estes algoritmos estão dados pelo tipo de estimação de parâmetros que podem ser, no caso da abordagem Não-Supervisionado, o EM, o *K-means*, ou o *Mean-Shift* e o SMEM no caso Supervisionado. Apenas no caso da abordagem Supervisionado tem-se um argumento adicional que é o número de componentes iniciais do algoritmo SMEM. Da mesma forma do que no bloco Classificação, neste bloco também precisa-se estabelecer o tipo de problema de classificação e a dimensão do vetor de características.

Os resultados da classificação, feitas tanto com técnicas conhecidas quanto com as técnicas baseadas em GMM, são mostrados mediante a taxa de erro de classificação total, pela taxa de erro de classificação em cada classe tanto para os problemas de classificação de 4 classes quanto para o problema de de 5 classes considerando as distribuições natural e balanceada, e pela matriz de confusão resultante do teste de classificação.

Na Figura A.2 mostra-se o esquema da organização dos controles da interface de usuário, nela os blocos brancos correspondem a programas e funções desenvolvidos no trabalho prévio (Samatelo, 2007) e os blocos cinza correspondem a algoritmos e funções desenvolvidas e usadas neste trabalho.

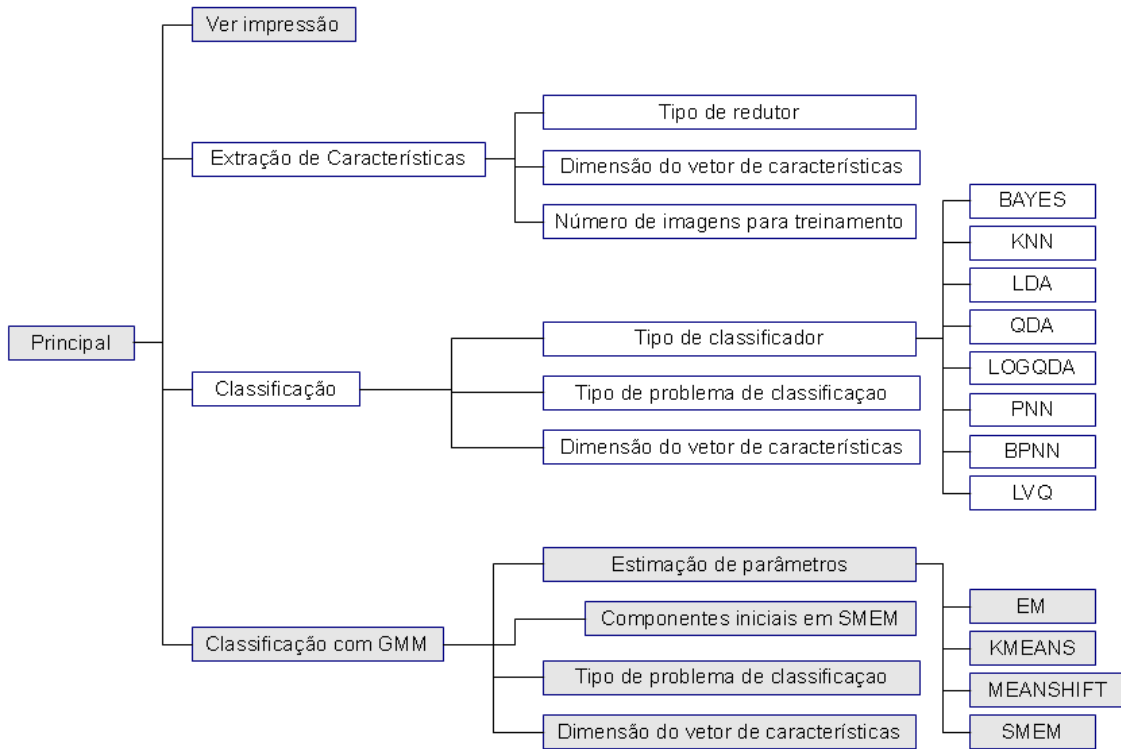


Figura A.2: Esquema da organização da interface de usuário.

# Apêndice B

## Resultados Adicionais

### B.1 Valores de media e covariância

A Tabela B.1(a) mostra os resultados obtidos ao realizar a clusterização com misturas gaussianas com dois coeficientes PCA, ou seja  $\mu$ ,  $\Sigma$  e  $\alpha$ , onde:  $\mu = [\mu_1, \mu_2]^T$  e  $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$  para o problema de quatro classes. A distribuição da mistura de quatro componentes Gaussianas correspondente a esses parâmetros foi mostrada na Figura 4.2(a).

(a)

	$\mu_{(1)}$	$\mu_{(2)}$	$\sigma_{11}$	$\sigma_{12} = \sigma_{21}$	$\sigma_{22}$	$\alpha$
<b>LL</b>	0.65	9.54	79.90	-31.27	54.83	0.24
<b>W</b>	-19.67	2.18	31.41	6.22	72.92	0.22
<b>RL</b>	-3.16	-10.42	72.55	19.61	42.20	0.26
<b>A/TA</b>	17.41	-0.11	16.86	2.46	14.95	0.28

(b)

	$\mu_{(1)}$	$\mu_{(2)}$	$\sigma_{11}$	$\sigma_{12} = \sigma_{21}$	$\sigma_{22}$	$\alpha$
<b>LL</b>	-1.01	9.89	60.10	-17.65	53.61	0.21
<b>W</b>	-20.70	5.38	27.09	19.70	63.89	0.17
<b>RL</b>	-8.21	-10.79	60.73	-3.65	40.46	0.25
<b>TA</b>	10.35	-1.06	16.14	7.79	35.94	0.15
<b>A</b>	18.93	-0.17	9.74	2.46	13.96	0.22

Tabela B.1: Parâmetros estimados da função de distribuição de probabilidades da mistura gaussiana para o problema de (a) quatro classes e (b) cinco classes.

A Tabela B.1(b) mostra os resultados para o problema de cinco classes. A distribuição da mistura de cinco componentes correspondente a esses parâmetros foi mostrada na Figura 4.2(b).

Veja que os parâmetros das funções de densidade de probabilidade que modelam as classes RL, LL e W são distintos nos dois tipos de problemas. Isto acontece porque o fato de juntar as amostras pertencentes às classes TA e A influenciam na estimação dos parâmetros de todas as amostras do espaço.

## B.2 Elipses de confiança

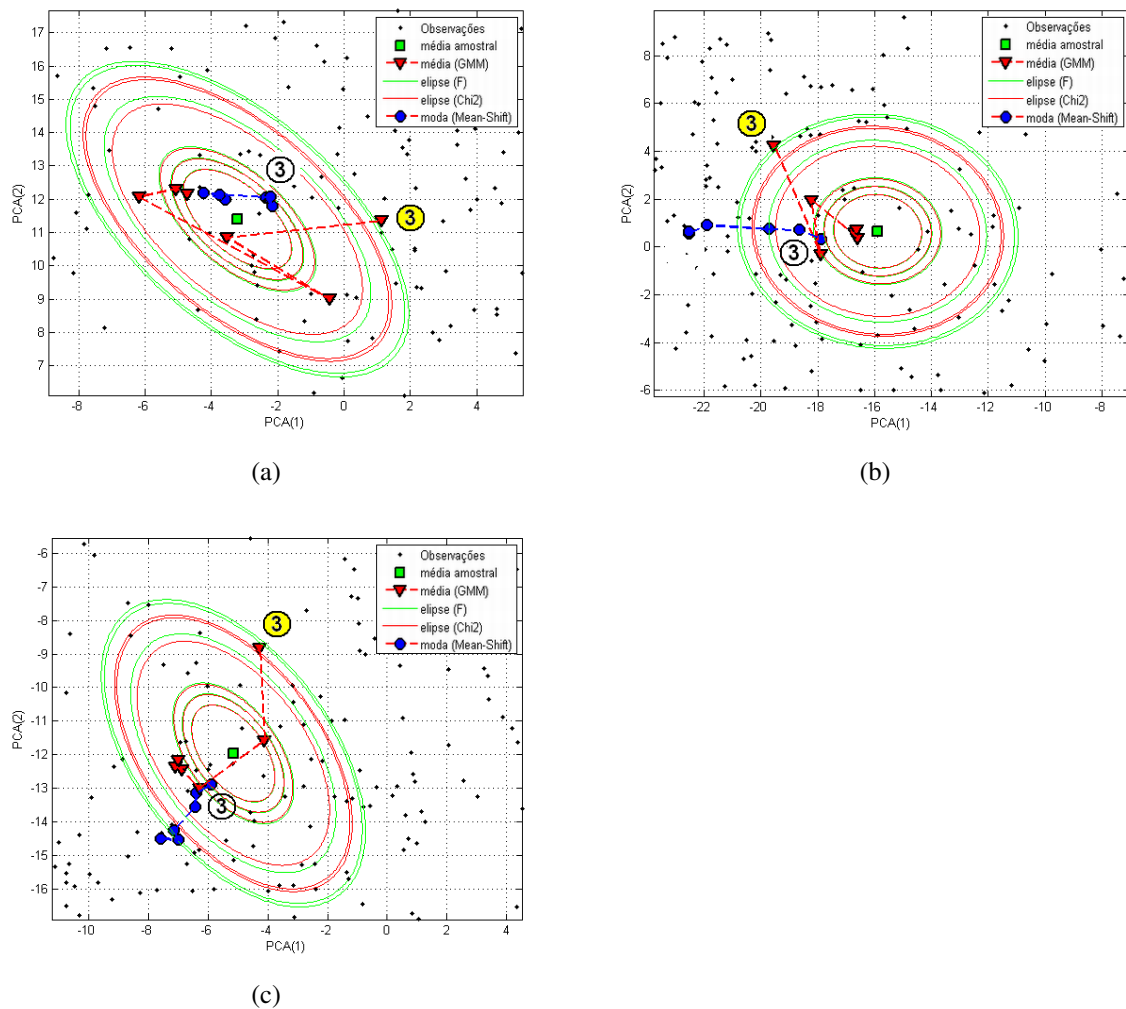


Figura B.1: Elipses de confiança da média amostral de Fisher e  $\chi^2$ , moda achada com *Mean-Shift* e média achada com EM para  $d = \{3, 6, 10, 38, 61, 64\}$ . (a) Classe **LL**. (b) Classe **W**. (c) Classe **RL**.

### B.3 Divergência de Kullback-Leibler

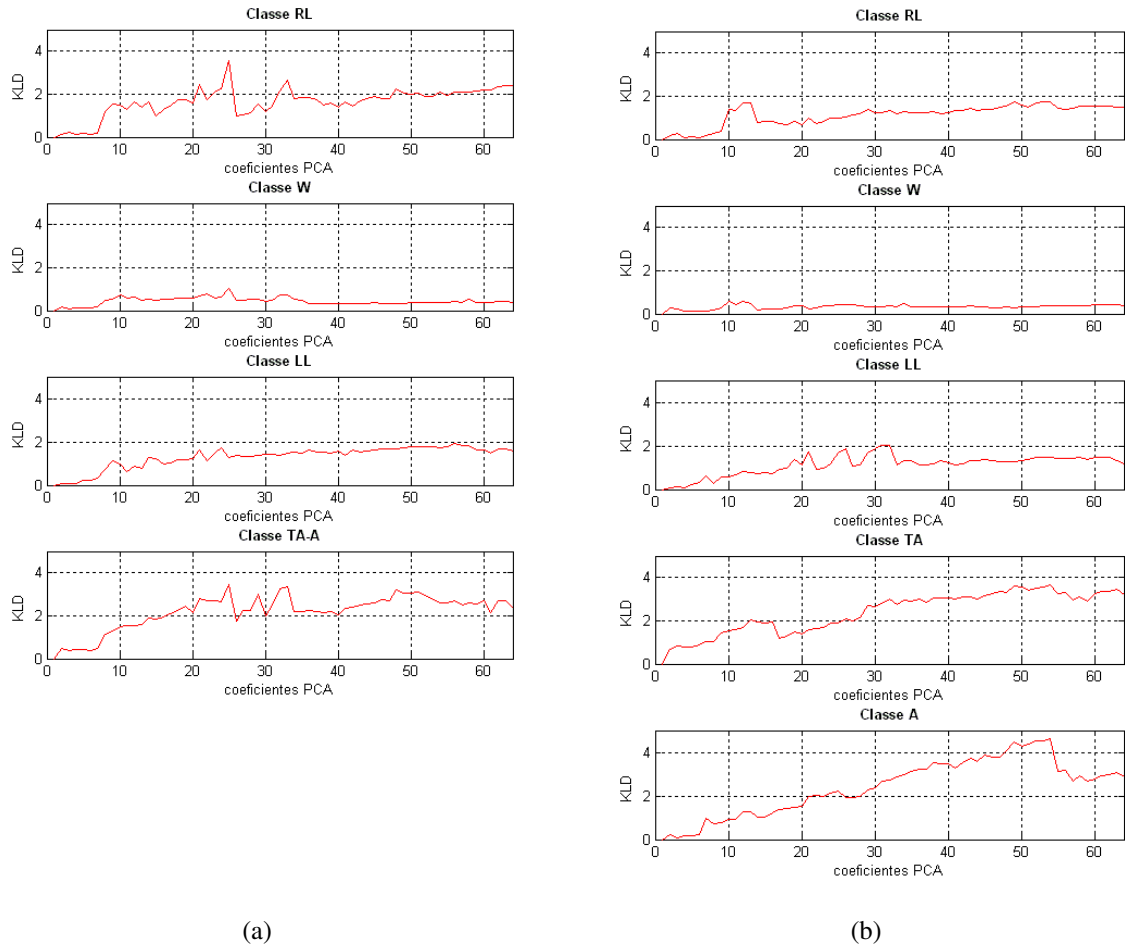


Figura B.2: Divergência de Kullback-Leibler (a) Problema de 4 classes. (b) Problema de 5 classes.

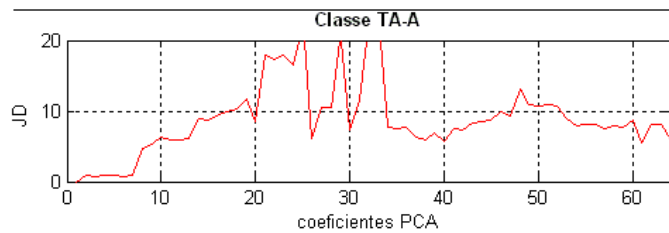


Figura B.3: Divergência de Jefferys-Kullback-Leibler entre distribuições geradas pelos parâmetros amostrais e os parâmetros estimados com EM da classe TA-A para o problema de quatro classes.

# Referências Bibliográficas

- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Candela, G. T., Grother, P. J., Watson, C. I., Wilkinson, R. A., e Wilson, C. L. (1995). Psasys: A pattern-level classification automation system for fingerprints. *Tech. Report 5647 National Institute of Standards and Technology*.
- Capelli, R., Lumini, A., Maio, D., e Maltoni, D. (1999). Fingerprint classification by directional image partitioning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, páginas 402–421.
- Cheng, Y. (1995). Mean shift mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Chew, V. (1966). Confidence, prediction and tolerance regions for multivariate distribution. *Journal of the American Statistical Association*.
- Cho, B. H., Kim, J. S., Bae, J. H., Bae, L. G., e Yoo, K. Y. (2000). Core based fingerprint image classification. *International Conference of Pattern Recognition*.
- Chuong, B. e Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature Biotechnology*.
- Comaniciu, D. e Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Daugman, J. (1993). High confidence visual recognition of person by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, páginas 1148–1161.
- Dempster, A. P., Laird, N. M., e Rubin, D. B. (1977). Maximum likelihood from incomplete data via em algorithm. *Journal of the Royal Statistical Society. Series B*, páginas 1–38.
- Derpanis, K. G. (2005). mean shift clustering. *Tech Rep*.
- Duda, R. O., Hart, P. E., e Stork, D. G. (2001). *Pattern Classification*. Wiley-Interscience, 2 edição.

- Fu, K.-S. e Booth, T. L. (1986). Grammatical inference: Introduction and survey-part ii. (3):360–375.
- Fukunaga, K. e Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Information Theory*, páginas 32–40.
- Hong, L., Wan, Y., e Jain, A. (1998). Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, páginas 777–789.
- Jain, A. K., Prabhakar, S., e Hong, L. (1999). A multichannel approach to fingerprint classification. 21(4):348–359.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond.*, 186:453–461.
- Jiang, Y. (2005). On orientation and anisotropy estimation for on line fingerprint authentication. *IEEE Transactions on Signal Processing*.
- Jonhson, R. A. e Wichern, D. W. (1998). *Applied multivariate statistical analysis*. Prentice Hall, 4 edição.
- Kanungo, T. e Haralick, R. (1995). Multivariate hypothesis testing for gaussian data: Theory and software. *Tech Report: ISL-TR-95-05*.
- Kawagoe, M. e Tojo, A. (1984). Fingerprint pattern classification. *Pattern Recognition*.
- Kim, S. C. e Kang, T. J. (2007). Texture classification and segmentation using wavelet packet frame and gaussian mixture model. *Pattern Recognition*, 40(4):1207 – 1221.
- Kullback, S. e Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.
- Liu, M., Jiang, X., e Kot, A. C. (2007). Efficient fingerprint search based on database clustering. *Pattern Recognition*, páginas 1793–1803.
- Maltoni, D., Maio, D., Jain, A. K., e Prabhakar, S. (2003). *Handbook of fingerprint recognition*. Springer-Verlag.
- Maltoni, D., Maio, D., Jain, A. K., e Prabhakar, S. (2009). *Handbook of fingerprint recognition*. Springer-Verlag.
- Mukhopadhyay, S. e Chanda, B. (2000). A multiscale morphological approach to local contrast enhancement. *Signal Processing*.



- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Jornal on Mathematical Psychology*, páginas 90–100.
- Samatelo, J. L. A. (2007). *Um sistema de classificacao de impressoes digitais*. MST dissertation in portuguese, PPGEE, Espírito Santo federal university.
- Samatelo, J. L. A. e Salles, E. O. T. (2009). Determination of the reference point of a fingerprint based on a multiple. *XXII Brazilin Symposium on Computer Graphics and Image Processing*, páginas 209–215.
- Tseng, C. L., Chen, Y. H., Xu, Y. Y., Pao, H. T., e Fu, H.-C. (2004). A self-growing probabilistic decision-based neural network with automatic data clustering. *Neurocomputing*, 61:21 – 38.
- Tumminello, M., Lillo, F., e Mantegna, R. N. (2007). Kullback-leibler distance as a measure of the information filtered from multivariate data. *Phys. Rev. E*, 76(3):031123.
- Ueda, N., Nakano, R., Grahramani, Z., e Hilton, G. E. (1998). Split and merge em algorithm for improving gaussian mixture density estimates. *Neural Networks for Signal Processing VIII, 1998.*, páginas 274–283.
- Ueda, N., Nakano, R., Grahramani, Z., e Hilton, G. E. (2000). Smem algorithm for mixture models. *Neural Computation*, páginas 2109–2128.
- Wagemaar, D. (2000). Fsmem for mog. *Term project for CS/CNS/EE 156b*.
- Wang, Y., Hu, J., e Han, F. (2006). Enhanced gradient-based algorithm for the estimation of fingerprint orientation fields. *Applied Mathematics and Computation*.
- Watson, C. I. e Wilson, C. L. (1992). Nist special database 4, fingerprint database. *Tech. Rep. U.S. National Institute Standards and Technology*.
- Watson, C. I. e Wilson, C. L. (1993). Nist special database 14, fingerprint database. *Tech. Rep. U.S. National Institute Standards and Technology*.
- Wilson, C. L., Candela, G. T., e Watson, C. I. (1994). Neural network fingerprint classification. *Artificial Neural Networks*, páginas 1,2.
- Yiu, K., Mak, M., e Li, C. (1999). Gaussian mixture models and probabilistic decision-based neural networks for pattern classification: A comparative study. *Neural Computing and Applications*, páginas 235–245.