Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

Jodavid de Araújo Ferreira

Classificadores Induzidos de Divergências para Modelos Vetoriais e para Distribuição Wishart Complexa: Experimentos em Imagens PolSAR

Recife
2017

Jodavid de Araújo Ferreira

Classificadores Induzidos de Divergências para Modelos Vetoriais e para Distribuição Wishart Complexa: Experimentos em Imagens PolSAR

Dissertação apresentada ao programa de Pós-graduação em Estatística do Departamento de Estatística da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre em Estatística. Área de Concentração: Estatística Aplicada.

ORIENTADOR: **Prof. Dr. Abraão D. C. do Nascimento**

Recife
2017

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

F383c Ferreira, Jodavid de Araújo
Classificadores induzidos de divergências para modelos vetoriais e para distribuição Wishart complexa: experimentos em imagens PolSAR / Jodavid de Araújo Ferreira. – 2017.
77 f.: il., fig., tab.

Orientador: Abraão D. C. do Nascimento.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2017.
Inclui referências e apêndices.

1. Estatística. 2. Teoria da informação. I. Nascimento, Abraão D. C. do (orientador). II. Título.

310 CDD (23. ed.) UFPE- MEI 2017-56

JODAVID DE ARAÚJO FERREIRA

CLASSIFICADORES INDUZIDOS DE DIVERGÊNCIAS PARA MODELOS VETORIAIS E PARA DISTRIBUIÇÃO WISHART COMPLEXA: EXPERIMENTOS EM IMAGENS POLSAR

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 17 de fevereiro de 2017.

BANCA EXAMINADORA

Prof. Abraão David Costa do Nascimento
UFPE

Prof. Getúlio José Amorim do Amaral
UFPE

Prof. Marcelo Rodrigo Portela Ferreira
UFPB

Agradecimentos

Em primeiro lugar agradeço a Deus por me abençoar a cada instante nesta jornada longa e difícil que enfrentamos em cada dia de nossas vidas.

Agradeço a minha esposa, Natália Santos, por estar ao meu lado a todo o tempo. Por seu amor companheirismo e palavras de encorajamento. À minha família, meus pais José Targino Ferreira e Maria José de Araújo Ferreira, que sempre acreditou, investiu e incentivou em meus estudos. Ao meu irmão Jobson de Araújo Ferreira e sua esposa Andréa Freitas. À minha segunda família (da minha esposa), em especial a minha sogra Maria de Lourdes por todo cuidado e preocupação para comigo.

Agradeço ao professor Abraão Nascimento por ter acreditado e confiado em mim, por ser exemplo de profissional, professor e amigo. Agradeço ao professor Hemílio Fernandes Campos Coêlho, por toda colaboração nestes trabalhos e por confiar em mim tanto profissional como pessoal. No meu crescimento acadêmico e científico levo um pouco de vocês comigo, obrigado por serem exemplos.

Aos meus amigos e colegas do mestrado e da graduação: Saul Azevedo, Ramon Lima, Pedro Almeida, Marina Travassos, Maizza, Alisson Santos, Camila Ravena, Aldine (na graduação), Yuri, Alejandro, Rodney, Walmir, Jucelino, Camila Ribeiro, Elisandra, Olivia, Camila Lopes, Vinícius (Mestrado) e tantos outros, obrigado por estarem próximos nos momentos de estudo e de descontração. À Valéria, secretária do programa, por sempre orientar em burocracias acadêmicas além de suas palavras de incentivo. Agradeço também a Luana Meireles, o Alisson Silva e a Wanessa Weridiana por toda amizade e apoio.

Aos professores do DE-UFPB, por terem colaborado no meu aprendizado na graduação em estatística. A todos os professores do DE-UPPE, pelos ensinamentos que contribuíram para minha formação acadêmica e científica a nível de mestrado, em especial aos professores Cribari Neto, Getúlio Amorim, Alex Dias, Abraão Nascimento e Audrey.

Ao Professor Ronei Marcos de Moraes, por ter me incluso neste mundo das pesquisas científicas e ter me cedido um pouco de suas experiência adquiridas.

À Fundação de Amparo a Ciência e Tecnologia de Pernambuco-FACEPE, pelo apoio financeiro. E aos membros da banca examinadora, por suas valiosas sugestões.

Buscai em primeiro lugar o Reino de Deus e sua justiça e todas as coisas vos serão dadas em acréscimo. Não vos preocupeis, pois, com o dia de amanhã: o dia de amanhã terá suas próprias preocupações. A cada dia basta o seu cuidado.

Mt. 6: 33-34.

Resumo

Dentre as técnicas de sensoriamento remoto, o sistema de radar de abertura sintética polarimétrico (*Polarimetric Synthetic Aperture Radar-PolSAR*) tem assumido uma posição de destaque. Isto se deve a capacidade do PolSAR em (*i*) operar em várias condições atmosféricas (na ausência de luz, tempo nublado, entre outros), (*ii*) produzir imagens em alta resolução espacial e (*iii*) fornecer informações sobre uma cena na perspectiva de vários canais de polarização. Entretanto dados de imagens PolSAR são fortemente contaminados por um ruído multidimensional (chamado de *ruído speckle*), o que dificulta sua modelagem e processamento. Em particular, pesquisas indicam que o uso dos classificadores clássicos (tais como, análise discriminante linear-LDA, análise discriminante quadrática-QDA, K vizinhos mais próximos-KNN e máquina de vetores de suporte-SVM) comumente resultam em desempenhos insuficientes, o que motiva a proposta de classificadores adaptados à natureza de dados PolSAR. Neste trabalho, objetiva-se construir uma coleção de classificadores para dados PolSAR com base em distâncias e divergências entre medidas de probabilidade para vetores e matrizes aleatórios. São considerados os modelos vetoriais normal, skew-normal, t-Student e skew-t e a distribuição matricial Wishart complexa escalonada (WCE). Utilizando as divergências/distâncias de Rényi, Kullback-Leibler e Jeffreys combinadas aos modelos vetoriais e matricial na elaboração de classificadores para dados de intensidade SAR multivariados e "full PolSAR" (matricial hermitiano), são propostos quatorze classificadores. Em particular, oito deles são classes de classificadores incorporadas a um processo de otimização para um parâmetro de flexibilidade (induzido da ordem da divergência de Rényi). Através de um estudo de simulação Monte Carlo, a *performance* dos métodos propostos é quantificada e comparada com outras obtidas pela distância de Kullback-Leibler para a distribuição normal multivariada e pelos métodos LDA, QDA, KNN e SVM. Para este fim, imagens sintéticas foram simuladas obedecendo à física de formação de imagens PolSAR. O classificador com base na divergência otimizada de Rényi para a lei WCE pode superar os demais. Finalmente, experimentos com dados reais extraídos de duas imagens PolSAR são realizados. Resultados apresentam evidências em favor dos classificadores ótimos com base na divergência de Rényi para a imagem de São Francisco com base no sensor AIRSAR (Radar de

Abertura Sintética Airborne) e do classificador baseado na divergência Kullback-Leibler para a distribuição t-Student multivariada para a imagem de Foulum com base no sensor EMISAR (*Electromagnetics Institute SAR*).

Palavras-chave: PolSAR. Teoria Estatística da Informação. Divergências. Kullback-Leibler. Jeffreys. Rényi.

Abstract

Among the remote sensing techniques, the PolSAR (Polarimetric Synthetic Aperture Radar) system has received a prominent position. This is due to its capability in (*i*) operating in various weather conditions (in the light absence, cloudy scenarios, among others), (*ii*) producing images at high spatial resolution, and (*iii*) providing information on a scene from polarization channels. However, data obtained from PolSAR images are strongly contaminated by a multidimensional interference pattern (called speckle noise) and this effect hampers their modelling and processing. In particular, works have indicated the use of classic classifiers (Linear Discriminant Analysis-LDA, Quadratic Discriminant Analysis- QDA, k-Nearest Neighbour-KNN and Support Vector Machine-SVM) commonly result in bad performance, which motivates the proposal of classifiers tailored to the nature of PolSAR data. This study aims to furnish a collection of classifiers for PolSAR data based on distances and divergences between probability measures for random vectors and matrices. We consider normal, skew-normal, t-Student, skew-t vector models and the scaled complex Wishart (SCW) distribution as like probabilistic assumptions. Using the Rényi, Kullback-Leibler, and Jeffreys divergences/distances combined with the adopted vector and matrix models, we propose fourteen classifiers. In particular, eight of them are classes of classifiers equipped by an optimization procedure for the flexibility parameter (induced from the Rényi divergence order). Through a Monte Carlo simulation study, the performance of proposed methods is quantified and compared with those obtained by the Kullback-Leibler distance for multivariate normal distribution and by LDA, QDA, KNN and SVM methods. To that end, synthetic images are generated obeying the physics of PolSAR image formation. The classifier based on the optimized Rényi divergence for the SCW model may outperform all considered alternatives for synthetic data. Finally, experiments with real data from PolSAR images are considered. Results provide evidence in favor of optimal classifiers based on the Rényi divergence for the image of San Francisco based on the sensor AIRSAR (Airborne Synthetic Aperture Radar) and of the classifier in terms of the Kullback-Leibler divergence for the multivariate t-Student distribution for the image of Foulum based on the sensor EMISAR (Electromagnetics Institute SAR).

Keywords: *PolSAR. Statistical Information Theory. Divergences. Kullback-Leibler. Jeffreys. Rényi.*

Lista de Figuras

2.1 Reflexividade em diferentes rugosidades. 21
2.2 Definição de dados *single-look* e *multilook*. 22
2.3 Diferentes rugosidades através do valor de β na distribuição $\mathcal{G}_p^0$. 24

3.1 Densidade normal bivariada padrão. 26
3.2 Densidade skew-normal bivariada. 27
3.3 Densidade t-Student bivariada. 28
3.4 Densidade skew-t bivariada. 29
3.5 Comportamento da divergência KL para a distribuição skew-normal para diferentes coeficientes de assimetria. 33
3.6 Comportamento da divergência KL para a distribuição t-Student para diferentes graus de liberdade. 34

5.1 Imagem do sistema AIRSAR. 45
5.2 Imagem do sistema EMISAR. 46
5.3 Imagem sintética. 47
5.4 Classificação com base nas divergências/distâncias de Rényi para escolha do α ótimo. 49
5.5 Imagem de São Francisco. 52
5.6 Imagem de Foulum. 53

A.1 Faixas do comprimento de onda e da frequência para um espectro eletromagnético (Fonte: Freitas et al. 2003, p. 2). 61

B.1 Classificação da imagem sintética para *looks* = 3, 4 e 8. 67
B.2 Classificação de São Francisco 70
B.3 Classificação de Foulum 73

Lista de Tabelas

3.1 Distâncias e divergências $h - \phi$ e funções ϕ e h relacionadas. 31

5.1 Matriz de Classificação. 47
5.2 Interpretação do coeficiente kappa. 48

B.1 *Grid* utilizado para KNN. 62
B.2 Parâmetros utilizados no SVM. 62
B.3 Classificação para imagem sintética com número de *looks* 3. 63
B.4 Classificação para imagem sintética com número de *looks* 4. 64
B.5 Classificação para imagem sintética com número de *looks* 8. 65
B.6 Classificação para a imagem San Francisco (USA). 68
B.7 Classificação para a imagem Foulum (DK). 71

C.1 Resultados das estimações para α ótimo. 74
C.2 Resultados para Viés e EQM. 75
C.3 Resultados das estimações utilizando Bootstrap. 76
C.4 Resultados Viés e EQM utilizando Bootstrap. 77

Lista de Abreviaturas e Siglas

CKLN:	Classificador de Kullback-Leibler para distribuição normal
CJN:	Classificador de Jeffreys para distribuição normal
CKLSN:	Classificador de Kullback-Leibler para distribuição skew-normal
CJSN:	Classificador de Jeffreys para distribuição skew-normal
CKLt:	Classificador de Kullback-Leibler para distribuição t-Student
CJt:	Classificador de Jeffreys para distribuição t-Student
CKLSt:	Classificador de Kullback-Leibler para distribuição skew-t
CJSt:	Classificador de Jeffreys para distribuição skew-t
CRN1:α:	Classificador da divergência de Rényi para distribuição normal
CRN2:α:	Classificador da divergência de Rényi simétrico para distribuição normal
CRND1:α:	Classificador da distância de Rényi para distribuição normal
CRND2:α:	Classificador da distância de Rényi 2 para distribuição normal
CRWC1:α:	Classificador da divergência de Rényi para distribuição Wishart complexa escalonada
CRWC2:α:	Classificador da divergência de Rényi simétrico para distribuição Wishart complexa escalonada
CRWCD1:α:	Classificador da distância de Rényi para distribuição Wishart complexa escalonada
CRWCD2:α:	Classificador da distância de Rényi 2 para distribuição Wishart complexa escalonada

Lista de Símbolos

X:	Variável aleatória
x:	Realização da variável aleatória
$\mathbf{X}$:	Vetor aleatório
$\dot{\mathbf{X}}$:	Realização do Vetor aleatório
$\boldsymbol{X}$:	Matriz aleatória
$\mathrm{D}_{\mathrm{KL}}^{N}$:	Divergência de Kullback-Leibler para distribuição normal
J^{N}:	Divergência de Jeffreys para distribuição normal
D_{KL}^{SN}:	Divergência de Kullback-Leibler para distribuição skew-normal
J^{SN}:	Divergência de Jeffreys para distribuição skew-normal
D_{KL}^{t}:	Divergência de Kullback-Leibler para distribuição t-Student
J^{t}:	Divergência de Jeffreys para distribuição t-Student
D_{KL}^{St}:	Divergência de Kullback-Leibler para distribuição skew-t
J^{St}:	Divergência de Jeffreys para distribuição skew-t
$\mathrm{D}_{\mathrm{R}}^{\alpha}$:	Divergência de Rényi para distribuição normal
$\mathrm{D}_{\mathrm{R2}}^{\alpha}$:	Divergência de Rényi simétrico para distribuição normal
$\mathrm{d}_{\mathrm{R}}^{\alpha}$:	Distância de Rényi para distribuição normal
$\mathrm{d}_{\mathrm{R2}}^{\alpha}$:	Distância de Rényi 2 para distribuição normal
$D_{R_{\mathbb{C}}}^{\alpha}$:	Divergência de Rényi para distribuição Wishart complexa escalonada
$D_{R2_{\mathbb{C}}}^{\alpha}$:	Divergência de Rényi simétrico para distribuição Wishart complexa escalonada
$d_{R_{\mathbb{C}}}^{\alpha}$:	Distância de Rényi para distribuição Wishart complexa escalonada
$d_{R2_{\mathbb{C}}}^{\alpha}$:	Distância de Rényi 2 para distribuição Wishart complexa escalonada

Sumário

1 Introdução **16**
1.1 Contribuição desta dissertação 18
1.2 Plataforma computacional 18

2 Dados Polarimétricos *Multillook* **20**
2.1 Modelos para dados polarimétricos *multilook* 21

3 Distribuições Estocásticas para Dados de Intensidades *PolSAR* **25**
3.1 Distribuições de vetores aleatórios 26
3.1.1 Distribuição normal multivariada 26
3.1.2 Distribuição skew-normal multivariada 26
3.1.3 Distribuição t-Student multivariada 27
3.1.4 Distribuição skew-t multivariada 28
3.2 Distâncias estocásticas 29
3.2.1 Divergências de KL e J para distribuição normal multivariada . . . 31
3.2.2 Divergências KL e J para distribuição skew-normal multivariada . . 32
3.2.3 Divergências de KL e J para distribuição t-Student multivariada . . 33
3.2.4 Divergências de KL e J para distribuição skew-t multivariada 35
3.2.5 Divergência de Rényi para as distribuições normal multivariada e WCE 36

4 Classificadores Clássicos e Baseados em TEI para Dados *PolSAR* **38**
4.1 K vizinhos mais próximos - KNN 39
4.2 Análise discriminante linear - LDA 39
4.3 Análise discriminante quadrática - QDA 40
4.4 Máquina de vetores de suporte - SVM 40
4.4.1 SVM - Caso linear 40

4.4.2 SVM - Caso não linear 41
4.5 Classificadores baseados em divergências 42

5 Resultados Numéricos **45**
5.1 Análise com dados sintéticos 45
5.2 Estudo com dados reais 51

6 Considerações Finais **55**

Referências **57**

Apêndice A Faixas de Comprimento e Frequência de Sistemas de Radar **61**

Apêndice B Resultados das Classificações **62**
B.1 Resultados para a imagem sintética 63

Apêndice C Estimação do Parâmetro α para as Imagens de São Francisco e Foulum **74**

CAPÍTULO 1

Introdução

Radares foram inicialmente utilizados na vigilância durante a segunda guerra mundial (MAITRE, 2006). Gradativamente os radares adquiriram capacidades de reconhecimento com imagens de baixa resolução, o que tornou possível distinguir `frequências` em diferentes refletores (MAITRE, 2006). Imagens de radar de abertura sintética (*Synthetic Aperture Radar-SAR*) surgiram como resultado de uma pesquisa feita por Carl A. Wiley no ano de 1951, enquanto trabalhava em um sistema de orientação de correlação (LOVE, 1985).

Imagens de radar estavam intimamente ligadas às aplicações militares. Por sua complexidade e custo de aquisição o uso de imagens de radar era inviável para muitas aplicações industriais. Essas imagens além de muito difíceis de manipular eram na maioria das vezes produto de sensores experimentais (MAITRE, 2006). Com o investimento de institutos sérios como a NASA (*National Aeronautics and Space Administration*) e ESA (*European Space Agency*), estas imagens passaram a ser utilizadas em várias áreas, tais como, prevenção de desastres naturais, levantamentos geológicos e de mineração na monitoração de rotas polares em Cartografia, gestão florestal e em levantamentos marítimos.

Na área do sensoriamento remoto, os sistemas de radar de abertura sintética polarimétrico (*Polarimetric Synthetic Aperture Radar-PolSAR*) têm sido indicados como ferramentas importantes e eficientes. Isso pode ser justificado por pelo menos três razões: tal sistema pode operar em todas as condições meteorológicas (o que é particularmente útil em países frequentemente nublado como os localizados na faixa equatorial), fornecer alta resolução espacial e produzir várias informações sobre uma cena.

Uma cena *PolSAR* pode ser entendida como um mapa em que cada entrada está associada a uma matriz hermitiana definida positiva. Isso requer o uso de métodos de processamento multivariados (FRERY *et al.*, 2014). Além disso, as imagens *PolSAR* são fortemente afetadas por uma interferência particular multidimensional, chamada *ruído*

speckle. Este efeito impõe aos dados *PolSAR* que suas características escalares reais (conhecidas como `intensidades`) não sigam um comportamento gaussiano e, como consequência, utlizem-se técnicas adaptadas. Assim, ferramentas de classificação eficientes são procuradas no processamento de imagens *PolSAR*.

Vários trabalhos têm endereçado a proposta de vários classificadores para dados *PolSAR* com o objetivo de extração de informações a respeito da cobertura do solo. Hoekman e Vissers (2003) propuseram uma técnica de classificação aplicada a regiões heterogêneas, em que os retornos *PolSAR* são descritos pela distribuição Wishart complexa. Os classificadores foram avaliados separando e combinando as bandas C e L[1]. Observou-se que desempenho superior foi obtido para bandas combinadas. Doulgeris *et al.* (2007) adotaram um esquema de classificação bayesiano com base na distribuição K-Wishart, que pode ser usado tanto de forma supervisionada como de não supervisionada. Mishra (2008) apresentou evidências de que o uso da análise discriminante linear (*Linear Discriminant Analysis*-LDA) no processamento de imagens *SAR* pode fornecer um desempenho ineficiente devido à natureza dos dados resultantes desse sistema. Tao *et al.* (2015) utilizaram vários métodos de extração de recursos que combinados com os classificadores KNN (*k-Nearest Neighbour classifier*) e SVM (*Support Vector Machine classifier*) podem melhorar as taxas de classificação no processamento de imagens *PolSAR*.

Nesta dissertação, utilizam-se distâncias estocásticas derivadas de diferentes modelos de probabilidade como núcleo de novos classificadores. A distância estocástica é uma quantidade importante na área da Teoria da Informação, cujos fundamentos foram formalizados por Shannon (1948). Ele introduziu três conceitos básicos: informação, entropia e entropia relativa (ou divergência). Dos conceitos introduzidos por Shannon, é possível definir distância entre duas medidas de probabilidade. O primeiro conceito de divergência na Estatística foi proposto por Kullback-Leibler (KL) (COMMENGES, 2015) e, a partir de então, várias classes de divergências têm sido propostas. As medidas de divergência foram submetidas a um tratamento formal por Csiszár (1967) e, subsequentemente, por Salicrú *et al.* (1994) resultando nas classes de divergência ϕ e $h-\phi$, respectivamente. Através deste formalismo, pode-se obter diferentes distâncias entre as distribuições de probabilidades já conhecidas da literatura. Vários autores têm derivado a divergência de Kullback-Leibler para distribuições univariadas e multivariadas; por exemplo, Contreras-Reyes e Arellano-Valle (2012) derivaram a divergência de Kullback-Leibler para a distribuição skew-normal multivariada e Contreras-Reyes (2014) propôs uma expressão assintótica da divergência de KL entre distribuições multivariadas assimétricas de caudas pesadas (em particular, os modelos multivariados t-Student e skew-t). Recentemente, o interesse em adaptar técnicas da Teoria Estatística da Informação para contextos de processamento de imagens tem crescido (NASCIMENTO *et al.*, 2010). Nascimento *et al.* (2010) utilizaram o tratamento formal desenvolvido por Salicrú *et al.* (1994) para obter oito medidas de divergências e

[1]As bandas de um sistema de radar definem em qual frequência e comprimento o radar está sendo utilizado. A banda C trabalha em uma frequência que varia de 3,9 a 5,75 GHz com um comprimento que varia de 3 a 10cm. A banda L possui uma frequência entre 0.39 e 1,55 GHz com comprimento variando de aproximadamente 25 a 90cm. Para mais detalhes ver a imagem A.1.

distâncias entre distribuições $\mathcal{G}_I^0$ para descrever intensidades *SAR* (FRERY *et al.*, 1997).

Técnicas de processamento de imagens costumam impor a suposição gaussiana. Entretanto, dados *PolSAR* são representados por matrizes hermitianas cujas diagonais principais não são descritas pela distribuição normal multivariada. Nesta dissertação, objetiva-se quantificar o efeito da não normalidade e propor critérios robustos em problemas de classificação de *pixels PolSAR*. Para este fim, as divergências KL e Jeffreys induzidas do modelo normal e das distribuições multivariadas t-Student, skew-t e skew-normal. Considerou-se também a classificação baseada na divergência de Rényi para dados *PolSAR* descritos pela distribuição Wishart complexa escalonada (WCE).

1.1 Contribuição desta dissertação

Este estudo objetiva construir uma coleção de classificadores para dados *PolSAR* induzidos de distâncias e divergências entre as medidas de probabilidade para vetores e matrizes aleatórias. Duas aplicações a dados reais são consideradas. Os resultados fornecem evidências a favor de classificadores ótimos baseados na divergência de Rényi para a imagem AIRSAR de San Francisco e do classificador baseado na divergência de Kullback-Leibler com a distribuição de t-Student multivariada para a imagem EMISAR de Foulum. Os resultados desta dissertação foram submetidos para o Periódico *Pattern Recognition*.

Também foi apresentado um trabalho no *"2nd Brazilian Workshop on Polarimetric SAR Data Processing"*, realizado na Universidade Federal de Pernambuco no mês de agosto de 2016, intitulado "Classificadores Paramétricos de Regiões PolSAR com base em Divergências para Modelos Normal e Não Normal Multivariados". Neste caso, propuseram-se esquemas de classificação de vetores de intensidades associados a três canais de polarização SAR com base em divergências e distâncias entre modelos não normais. O comportamento das divergências foi quantificado na classificação de dados PolSAR, adotando duas medidas de avaliação: o acerto percentual e o coeficiente kappa.

1.2 Plataforma computacional

A utilização de pacotes estatísticos para a análise de dados é de grande importância. Dentre os *software* de domínio público utilizados para análise de dados, encontra-se o ambiente R. Este ambiente apresenta código aberto e tem abertura para publicação de pacotes fundamentados por novas metodologias estatísticas. Em um futuro próximo, objetiva-se que os classificadores desenvolvidos sejam publicados como um pacote em R. Além disso, o R possui um grande número de colaboradores das mais diversas áreas do conhecimento (R Development Core Team, 2015).

Com o intuito de trabalhar com processamento de imagens no *software* R, foram criadas rotinas (funções) adequadas para tais tipos de dados. A função `gerarpar()` foi criada para receber como entrada os dados de treinamento e retornar como saída os parâmetros

estimados de média e matriz de covariância para cada canal polarimétrico de cada classe (o número de classes varia de acordo com a imagem). Foi criado também uma rotina *default* para as classificações, pois as diferenças entre os classificadores propostos eram as divergências das distribuições estatísticas utilizadas, que na função do *software* R refletia em alterar o núcleo do classificador.

Esta dissertação está organizada da seguinte maneira. No Capítulo 2, descrevem-se alguns modelos para dados *PolSAR multilook*. No Capítulo 3, são apresentados e discutidas as medidas de divergência usadas nesta dissertação. No Capítulo 4, discutem-se alguns classificadores utilizados em Estatística. Apresentam-se também os classificadores propostos. Resultados numéricos são apresentados e discutidos no Capítulo 5. Por fim, as conclusões estão no Capítulo 6.

Como última nota deste capítulo, convém ressaltar que grande parte das tabelas e figuras de resultados são colocadas nos apêndices para prover maior fluidez textual.

CAPÍTULO 2

Dados Polarimétricos *Multillook*

Os dados *SAR* univariados são obtidos do mapeamento de uma área (que é frequentemente chamada de cenário) através da emissão e do registro de pulsos sob uma direção, obtendo como resultado imagens `monopolarizadas` (NASCIMENTO, 2012). Recentemente, estudos têm indicado que trabalhar com dados *SAR* em múltiplas direções pode resultar em ganhos na precisão de processamento. Os dados obtidos nesta abordagem são chamados de dados *PolSAR*. As informações polarimétricas são produzidas pela emissão e obtenção de sinais multivariados e complexos. Como resultado, são registrados a amplitude e a fase dos sinais retornados para uma cena, em que são consideradas as combinações das polarizações lineares de transmissão e recepção: *HH*, *HV*, *VH* e *VV* (*H* e *V* são representações para as polarizações horizontais e verticais, respectivamente).

Devido à utilização de `iluminação coerente` [1] no seu processamento, as imagens *PolSAR* são contaminadas por um ruído multidimensional, chamado de *speckle* O aparecimento do *speckle* se deve à interferência coerente do sinal refletido a partir das muitas unidades elementares da área sob estudo. O *ruído speckle*, além de dificultar a interpretação e análise de imagens *PolSAR*, impõe um comportamento não aditivo e não gaussiano nos elementos dos dados resultantes, o que inviabiliza o uso de técnicas clássicas de processamento de imagens (FRERY *et al.*, 1997); por exemplo, procedimentos adaptados foram propostos nas áreas de classificação (FRERY *et al.*, 2007), segmentação (BEAULIEU; TOUZI, 2004) e detecção de borda (SCHOU *et al.*, 2003).

O comportamento dos dados *PolSAR* são diferentes a depender do tipo de terreno, como ilustrado na Figura 2.1. As superfícies com rugosidade forte dispersam o sinal recebido para várias direções. A medida que a rugosidade vai se tornando fraca (ou seja,

[1] A formação de uma imagem SAR requer o processamento `coerente` do sinal de retorno recebido após a emissão dos pulsos. O sinal recebido (complexo) de cada pulso emitido é demodulado em fase (I) e quadratura (Q), amostrado (dada uma função de referência) e convertido para o formato digital (para formar uma seqüência de dados) (FREITAS *et al.*, 2003).

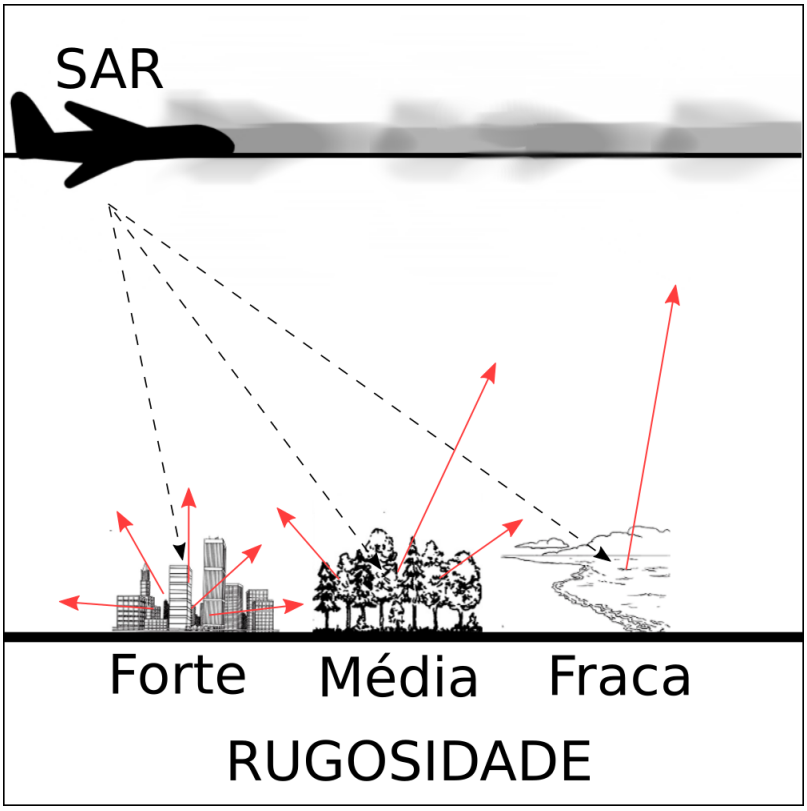


Figura 2.1: Reflexividade em diferentes rugosidades.

quando as superfícies vão ficando mais lisas), os sinais retornados ficam menos dispersos. As superfícies lisas, como pasto e oceano, por exemplo, refletem a energia incidente em direção oposta ao sistema imageador. Este último fato caracteriza regiões menos texturados nas imagens resultantes com cores escuras. Em contrapartida, quanto mais rugosa é a superfície (por exemplo, áreas urbanas), mais energia é retornada ao radar, fazendo com que a área seja mais clara. Cenários de florestas são menos texturadas do que cenários urbanos. Assim, regiões de florestas em uma imagem *PolSAR* são caracterizadas com um brilho menor do que o de áreas urbanas e maior em relação às regiões de oceanos e pastos. Existem situações em que os *pixels* exibem altos valores de variabilidade associados aos pulsos retornados, este fenômeno é denominado *"double bounce"*. Este evento geralmente é provocado por comportamentos naturais ou resultado de erros de mensuração do responsável e estão presentes em cenários de áreas urbanas e florestas densas. A análise de *pixels* afetados por *double bounce* é essencial, no sentido de que eles modificam as propriedades estatísticas de sua vizinhança, o que pode produzir estimativas e inferências mais viesadas do que o usual (NASCIMENTO, 2012).

2.1 Modelos para dados polarimétricos *multilook*

Uma imagem *PolSAR* é tal que cada entrada é associada aos elementos da seguinte matriz:

$$\boldsymbol{S} = \begin{pmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{pmatrix}, \tag{2.1}$$

em que $S_A \in \mathbb{C}$ é o valor complexo que representa o canal de polarização $A \in \{hh, hv, vh, vv\}$. Na prática, as polarizações cruzadas são iguais; isto é, que $S_{hv} = S_{vh}$ (ULABY; ELACHI, 1990). Desta forma, a matriz (2.1) pode ser representada como

$$\boldsymbol{z} = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \end{pmatrix}, \tag{2.2}$$

em que os índices 1, 2 e 3 denotam as polarizações hh, hv e vv (FREITAS *et al.*, 2004).

Dados na forma do vetor (2.2) são chamados de *PolSAR single-look*. Entretanto, modelos para estes dados não levam em consideração o controle do efeito do *ruído speckle* sobre as imagens *PolSAR*. A fim de contornar esta dificuldade, a literatura costuma propor um procedimento chamado de processo *multilook* ilustrado na figura a seguir.

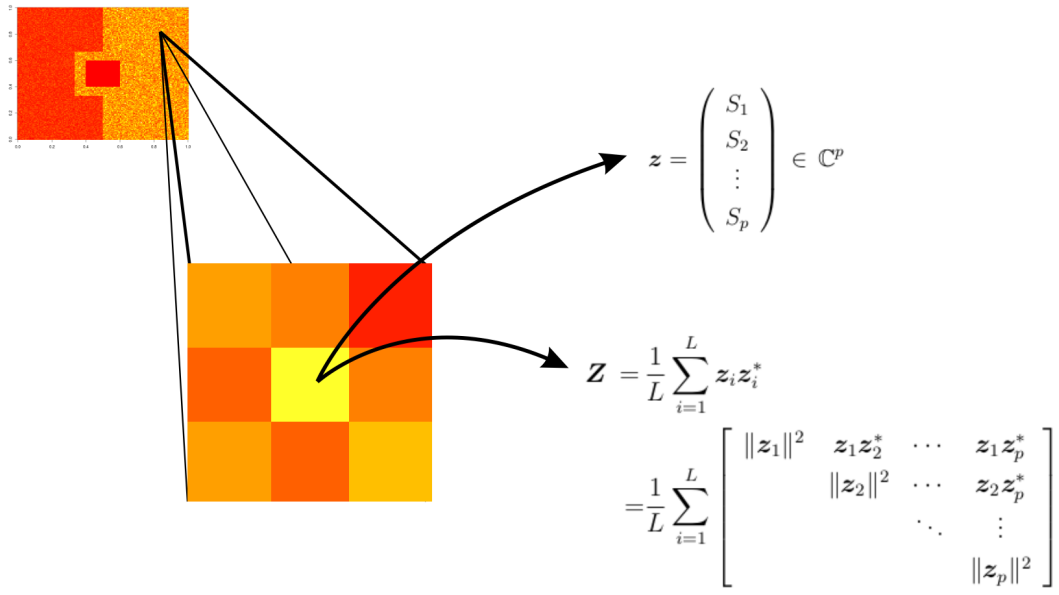


Figura 2.2: Definição de dados *single-look* e *multilook*.

De modo geral, seja $\boldsymbol{z}_i = (S_1^{(i)}\ S_2^{(i)}\ \cdots\ S_p^{(i)})^\top \in \mathbb{C}^p$ o i-ésimo vetor associado a p canais de polarização em uma amostra de L informações extraídas de uma mesma cena, para $i = 1, \ldots, L$. Assumindo que $\boldsymbol{z}_i$ segue uma distribuição normal complexa multivariada com média zero e matriz de covariância $\boldsymbol{\Sigma}$ $\left(\boldsymbol{z}_i \sim N_p^{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{\Sigma})\right)$ com função densidade de

probabilidade (fdp) (GOODMAN, 1963):

$$f(\boldsymbol{z};\boldsymbol{\Sigma}) = \frac{1}{\pi^p|\boldsymbol{\Sigma}|}\exp\left(-\boldsymbol{z}^*\boldsymbol{\Sigma}^{-1}\boldsymbol{z}\right), \tag{2.3}$$

em que $|\cdot|$ é o determinante de uma matriz ou valor absoluto de um escalar, $(\cdot)^*$ é o transposto conjugado de um vetor e $\boldsymbol{\Sigma}$ é a matriz de covariância hermitiana positiva definida, que contém todas as informações necessárias para caracterizar retorno polarimétrico dos dados em análise (FRERY *et al.*, 2014).

O processo *multilook* é definido como $\boldsymbol{Z} = L^{-1}\sum_{i=1}^{L}\boldsymbol{z}_i\boldsymbol{z}_i^*$ e, para $\boldsymbol{z}_i \sim N_p^{\mathbb{C}}(\boldsymbol{0},\boldsymbol{\Sigma})$, segue uma distribuição Wishart complexa escalonada (FRERY *et al.*, 2014) com fdp dada por:

$$f_{\boldsymbol{Z}}(\boldsymbol{Z};\boldsymbol{\Sigma},L) = \frac{L^{pL}|\boldsymbol{Z}|^{L-p}}{|\boldsymbol{\Sigma}|^L\Gamma_p(L)}\exp\left[-L\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z})\right], \tag{2.4}$$

em que $\Gamma_p(L) = \pi^{p(p-1)/2}\prod_{i=0}^{p-1}\Gamma(L-i)$, $L \geq p$, representa a função gama multivariada, $\Gamma(\cdot)$ é a função gama, e $\mathrm{tr}(\cdot)$ é o operador traço. Esta situação é denotada como $\boldsymbol{Z} \sim \mathcal{W}_p(\boldsymbol{\Sigma},L)$ e essa distribuição satisfaz $E\{\boldsymbol{Z}\} = \boldsymbol{\Sigma}$.

A distribuição WCE é considerada a distribuição básica para os dados *PolSAR*. Ela leva em conta um parâmetro de controle do efeito multidimensional do *ruído speckle* (denotado por L) e também o $\boldsymbol{\Sigma}$ que mapeia a dependência entre os canais de polarização. Embora este modelo tenha uma fdp com forma analítica tratável, ele não incorpora o efeito de rugosidade de relevo. Nesta dissertação, utiliza-se o modelo Wishart complexo escalonado para dados matriciais hermitianos e outras distribuições multivariadas vetoriais para múltiplas intensidades na elaboração de novos classificadores. A *performance* dos classificadores propostos será testada, tanto com dados de imagens sintéticas como reais. No primeiro caso, os dados gerados seguirão o modelo $\mathcal{G}_p^0$ proposto por Freitas *et al.* (2004), que estende o modelo WCE. Esta suposição é feita a fim de gerar cenários mais próximos aos dados reais, visto que o modelo $\mathcal{G}_p^0$ incorpora um parâmetro associado a textura de uma imagem *PolSAR*. Para esse fim, considerou-se a abordagem multiplicativa que será discutida a seguir.

Diferentes modelos podem ser usados para representar os dados *PolSAR*, porém uma estrutura estatística frequentemente utilizada é abordagem multiplicativa definida como (FREITAS *et al.*, 2004):

> "Um *pixel* de uma imagen *PolSAR* é descrito por uma matriz aleatória $\boldsymbol{Z} = X \cdot \boldsymbol{Y}$, em que X está associado com o *backscatter* (retroespalhamento do terreno) e $\boldsymbol{Y}$ descreve o *ruído speckle* multidimensional."

Para modelar o *backscatter* será utilizada a distribuição gama inversa denotada por $X \sim \Gamma^{-1}(-\beta,(-\beta-1)\mu))$ e com fdp dada por (NASCIMENTO, 2012)

$$f_X(x;\mu,\beta) = \frac{1}{\Gamma(-\beta)}\frac{1}{(-\beta-1)\mu}\left[\frac{(-\beta-1)\mu}{x}\right]^{\beta+1}\exp\left(-\frac{(-\beta-1)\mu}{x}\right), \tag{2.5}$$

em que $\mu > 0$ (escala), $\beta < 0$ (forma) e x é uma realização simples de uma variável aleatória X. Assumindo $\boldsymbol{Y} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, L)$ e $X \sim \Gamma^{-1}(-\beta, (-\beta-1)\mu))$, a distribuição para $\boldsymbol{Z}$ segue o modelo $\mathcal{G}_p^0$. De acordo com Freitas *et al.* (2004), a distribuição $\mathcal{G}_p^0$ é um caso especial no modelo multiplicativo. Ele pode descrever diferentes níveis de rugosidade e a sua fdp é dada por

$$f(\boldsymbol{Z}; \beta, \boldsymbol{\Sigma}, L) = \frac{L^{Lp}\Gamma(Lp-\beta)}{|\boldsymbol{\Sigma}|^L(-\beta-1)^\beta\Gamma(-\beta)\Gamma_p(L)}|\boldsymbol{Z}|^{L-p}\left[(-\beta-1)+L\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z})\right]^{\beta-Lp}, \qquad (2.6)$$

em que β é o parâmetro de rugosidade. Como pode ser visualizado na Figura 2.3, quanto menor for o valor do parâmetro responsável pela rugosidade, mais lisa será a área representada pelos dados gerados.

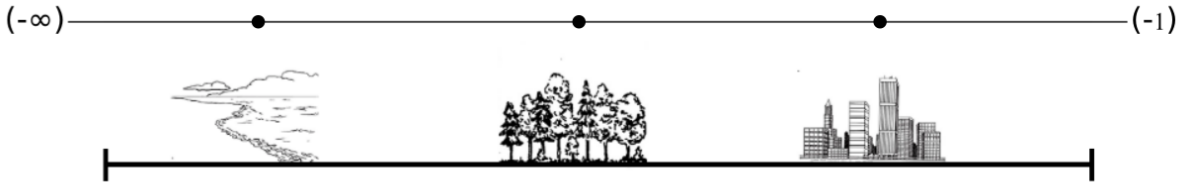


Figura 2.3: Diferentes rugosidades através do valor de β na distribuição $\mathcal{G}_p^0$.

CAPÍTULO 3

Distribuições Estocásticas em modelos para Dados de Intensidades Múltiplas e *PolSAR*

Como discutido no capítulo anterior, os dados *PolSAR multilook* são da forma:

$$\begin{bmatrix} Z_{11} & & [Z_{ij}] \\ & \ddots & \\ [Z_{ij}^*] & & Z_{pp} \end{bmatrix}_{p\times p} \in \{\boldsymbol{Z} \in \mathbb{C}^p \times \mathbb{C}^p : \boldsymbol{Z} = \boldsymbol{Z}^*, \boldsymbol{Z} \succ \boldsymbol{0}\}. \tag{3.1}$$

Nesta dissertação, objetiva-se construir classificadores munidos com medidas de divergências entre modelos de probabilidades para:

(a) intensidades múltiplas:

$$\boldsymbol{z} = (Z_{11}, \ldots, Z_{pp})^\top \in \mathbb{R}_+^p$$

e

(b) dados *full PolSAR multilook*:

$$\boldsymbol{Z} \text{ definido em (3.1).}$$

Para este último caso, utiliza-se a distribuição WCE, discutida no Capítulo 2 como modelo teórico. Para o caso (a), consideraram-se os modelos vetoriais apresentados a seguir.

3.1 Distribuições de vetores aleatórios

3.1.1 Distribuição normal multivariada

Também conhecida como `distribuição gaussiana`, ela apresenta dois parâmetros, um vetor de médias $\boldsymbol{\mu} \in \mathbb{R}^k$ e uma matriz de covariâncias $\boldsymbol{\Sigma} \succeq \mathbf{0}$. Este modelo é bastante utilizado devido a sua tratabilidade analítica e ao fato de vários resultados assintóticos (teorema central do limite, normalidade assintótica dos estimadores de máxima verossimilhança, ...) terem esta distribuição como quantidade limite. Seja $\mathbf{X}$ um vetor normal k variado, sua fdp é dada por

$$f_{\mathbf{X}}(\dot{\mathbf{X}}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} \phi_k(\dot{\mathbf{X}}; \boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

em que $\phi_k(\dot{\mathbf{X}}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-k/2} \exp\left\{-\frac{1}{2}(\dot{\mathbf{X}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\dot{\mathbf{X}} - \boldsymbol{\mu})\right\}$. Esta situação é denotado por $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Sejam $\mathbf{X}_1, \ldots, \mathbf{X}_n$ uma amostra aleatória obtida de $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, os EMVs para $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são dados por, respectivamente (JOHNSON; WICHERN, 2007):

$$\widehat{\boldsymbol{\mu}} = \overline{\mathbf{X}} = \sum_{i=1}^{n} \frac{\mathbf{X}_i}{n} \quad \text{e} \quad \widehat{\boldsymbol{\Sigma}} = \mathbf{S} = \sum_{j=1}^{n} \frac{(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})^\top}{n}.$$

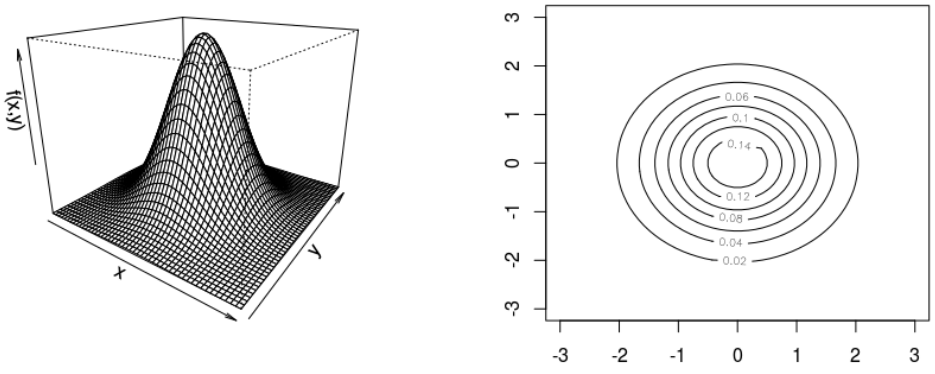


Figura 3.1: Densidade normal bivariada padrão.

Na Figura 3.1, os parâmetros utilizados foram o vetor $\boldsymbol{\mu} = \mathbf{0}$ e a matriz de covariâncias $\boldsymbol{\Sigma} = \mathbb{I}_2$, em que $\mathbb{I}_2$ é uma matriz identidade de ordem 2. Através desta figura é possível perceber a simetria desta distribuição pois em ambos os gráficos os dados são simétricos em torno do zero.

3.1.2 Distribuição skew-normal multivariada

A distribuição multivariada skew-normal foi introduzida por AZZALINI e VALLE (1996). Esta distribuição é definida com três parâmetros, um para o vetor de médias ou locação, um para a matriz de covariâncias e outro para o vetor de assimetrias. Como se

objetiva descrever intensidades múltiplas e estas são positivas, este modelo pode se tornar mais apropriado para estes tipos de dados. A distribuição skew-normal e suas extensões podem ser promissoras, uma vez que preservam as vantagens da distribuição normal com o benefício adicional de flexibilidade em relação à assimetria e à curtose. Seja $\mathbf{X}$ um vetor k dimensional que segue uma distribuição skew-normal multivariada, sua fdp é dada por

$$f_{\mathbf{X}}(\dot{\mathbf{X}};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\eta}) = 2\phi_k(\dot{\mathbf{X}};\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi[\boldsymbol{\eta}^\top\boldsymbol{\Sigma}^{-1/2}(\dot{\mathbf{X}}-\boldsymbol{\mu})], \quad \text{para } \dot{\mathbf{X}} \in \mathbb{R}^k,$$

em que $\Phi(z) = \int_{-\infty}^{z}(2\pi)^{-1/2}e^{-t^2/2}dt$ é função de distribuição acumulada (fda) da normal padrão e o vetor k dimensional $\boldsymbol{\eta}$ é o parâmetro de assimetria (CONTRERAS-REYES, 2014). Este caso é denotado por $\mathbf{X} \sim SN_k(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\eta})$.

Com respeito aos EMVs para os componentes de $\boldsymbol{\eta}$, Arellano-Valle *et al.* (2004) propuseram um processo recursivo para este fim. Em geral os EMVs, $\widehat{\eta}_j$, são definidos como:

$$\widehat{\eta}_j \;=\; \arg\max_{\eta_j\in\mathbb{R}}\left\{-\frac{n}{2}\log\frac{\pi\sigma^2}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i-\mu)^2 + \sum_{i=1}^{n}\log\Phi\left[\frac{\eta_j(z_i-\mu)}{\sqrt{\sigma^2}}\right]\right\}. \tag{3.2}$$

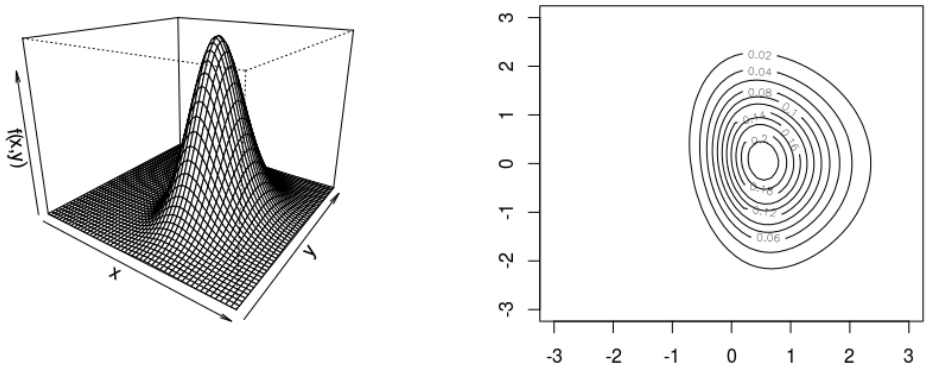


Figura 3.2: Densidade skew-normal bivariada.

Gráficos sobre a forma da distribuição skew-normal bivariada podem ser visualizadas na Figura 3.2, em que os parâmetros utilizados foram $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbb{I}_2$ e $\boldsymbol{\eta}^\top = [2, 0.1]^\top$. O parâmetro $\boldsymbol{\eta}$ controla a assimétria, portanto uma distribuição skew-normal assimétrica pode ser obtida, através da manipulação desse parâmetro. A distribuição normal é um caso particular desta distribuição, sendo obtida quando $\boldsymbol{\eta} = \mathbf{0}$.

3.1.3 Distribuição t-Student multivariada

Esta distribuição é definida por três parâmetros, sendo um deles os graus de liberdade. Vale ressaltar que, quanto maior for o valor dos graus de liberdade, mais a distribuição t-Student se aproxima da distribuição normal. Este caso é denotado por $\mathbf{X} \sim \mathrm{t}_k(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$. Seja $\mathbf{X}$ um vetor que segue uma distribuição t-Student, sua fdp pode ser definida por (HOFERT, 2013)

$$f_{\mathbf{X}}(\dot{\mathbf{X}};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)=\frac{\Gamma\left(\frac{\nu+k}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{k/2}\sqrt{\det\boldsymbol{\Sigma}}}\left[1+\frac{(\dot{\mathbf{X}}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\dot{\mathbf{X}}-\boldsymbol{\mu})}{\nu}\right]^{-\frac{\nu+k}{2}}.$$

Para a distribuição t-Student, os estimadores não possuem soluções em forma fechada simples. Assim, utilizou-se o método de otimização não linear BFGS (BROYDEN, 1969; FLETCHER, 1970; GOLDFARB, 1970; SHANNO, 1970). Seja $\boldsymbol{\theta}=(\boldsymbol{\mu}^{\top},\mathrm{vec}(\boldsymbol{\Sigma})^{\top},\nu)$, tais estimadores podem ser obtidos maximizando:

$$\ell(\boldsymbol{\theta})=-\frac{1}{2}\sum_{i=1}^{N}\Bigg\{-2\delta(\nu)+k\log(\pi)+\log(\nu-2)+(k-1)\log(\nu)$$
$$+\log(det(\boldsymbol{\Sigma}))+(\nu+k)\log\left[1+\frac{(\dot{\mathbf{X}}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}_1^{-1}(\dot{\mathbf{X}}-\boldsymbol{\mu})}{\nu-2}\right]\Bigg\},$$

em que $\delta(\nu)=\log\Gamma((\nu+k)/2)/\Gamma(\nu/2)$.

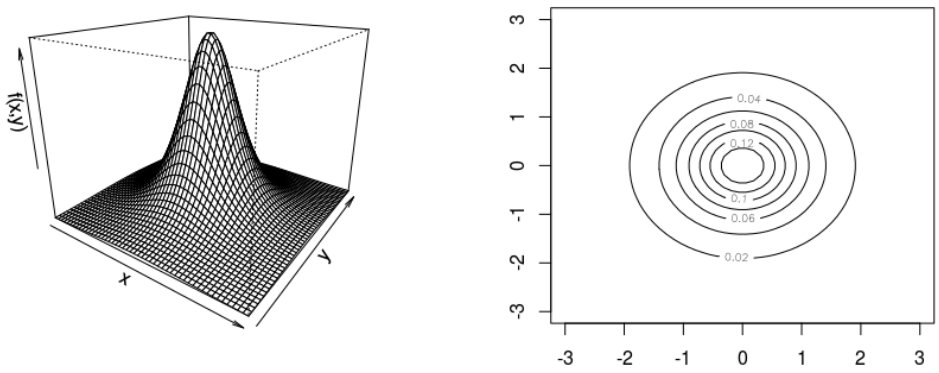


Figura 3.3: Densidade t-Student bivariada.

Na Figura 3.3, encontram-se gráficos para distribuição t-Student, na qual os parâmetros utilizados foram $\boldsymbol{\mu}=\mathbf{0}$, $\boldsymbol{\Sigma}=\mathbb{I}_2$ e $\nu=2$. Esta distribuição também é simétrica, mas relativamente à normal apresenta curvas de nível mais alongadas.

3.1.4 Distribuição skew-t multivariada

A distribuição skew-t multivariada foi proposta por Branco e Dey (2001). Ela proporciona maior flexibilidade e melhor capacidade para acomodar os dados de caudas longas quando comparada com a distribuição skew-Normal. Dizemos que um vetor aleatório $\dot{\mathbf{X}}\in\mathbb{R}^k$ tem uma distribuição skew-t com vetor de locação $\boldsymbol{\mu}\in\mathbb{R}^k$, matriz de covariância $\boldsymbol{\Sigma}\in\mathbb{R}^{k\times k}$, vetor de parâmetro de assimetria $\boldsymbol{\eta}\in\mathbb{R}^k$ e graus de liberdade $\nu>0$, se a fdp é (CONTRERAS-REYES, 2014):

$$f_{\mathbf{X}}(\dot{\mathbf{X}};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu,\boldsymbol{\eta})=2|\boldsymbol{\Sigma}|^{-1/2}t_k(\dot{\mathbf{X}}_0;\nu)T\left(\sqrt{\frac{\nu+k}{\nu+||\ \dot{\mathbf{X}}_0\ ||^2}}\overline{\boldsymbol{\eta}}^{\top}\ \dot{\mathbf{X}}_0;\nu+k\right),$$

em que $\dot{\mathbf{X}}_0 = \boldsymbol{\Sigma}^{-1/2}(\dot{\mathbf{X}} - \boldsymbol{\mu})$, $\overline{\boldsymbol{\eta}} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\eta}$, $t_k(\dot{\mathbf{X}}_0; \nu)$ é a fdp $t_k(\mathbf{0}, \boldsymbol{I}_k, \nu)$ e $T(\mathbf{X}; \nu + k)$ é a função de distribuição acumulada $T_1(0, 1, \nu + k)$. Os parâmetros $\boldsymbol{\theta} = [\boldsymbol{\mu}^\top, \text{vec}(\boldsymbol{\Sigma})^\top, \boldsymbol{\eta}^\top, \nu^\top]^\top$ foram estimados maximizando a função de verossimilhança da distribuição skew-t via algoritmo EM (*Expectation - Maximization*) (LEE; MCLACHLAN, 2011; MCLACHLAN; LEE, 2013):

$$\begin{aligned}\log L_c(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u}, w) = & K - \frac{1}{2} n \log |\boldsymbol{\Sigma}| - n \log \Gamma\left(\frac{1}{2}\nu\right) + \frac{1}{2} n\nu \log\left(\frac{1}{2}\nu\right) \\ & - \frac{1}{2} w \left(d(\mathbf{y}) + (\mathbf{u} - \mathbf{q})^\top \boldsymbol{\Lambda}^{-1} (\mathbf{u} - \mathbf{q})\right) \\ & + \left(\frac{1}{2}\nu + p - 1\right) \log(w),\end{aligned}$$

em que K não depende de $\boldsymbol{\theta}$, $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta})$, $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, $q = \boldsymbol{\Delta}^\top \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})$, $d(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})$, $\boldsymbol{\Lambda} = \mathbf{I}_p - \boldsymbol{\Delta}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Delta}$ e $\mathbf{U}|w \sim HN_p\left(\mathbf{0}, \frac{1}{w}\mathbf{I}_p\right)$ representa uma distribuição *half*-normal p dimensional e $W \sim gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ é uma distribuição gamma (α, β) com média α/β.

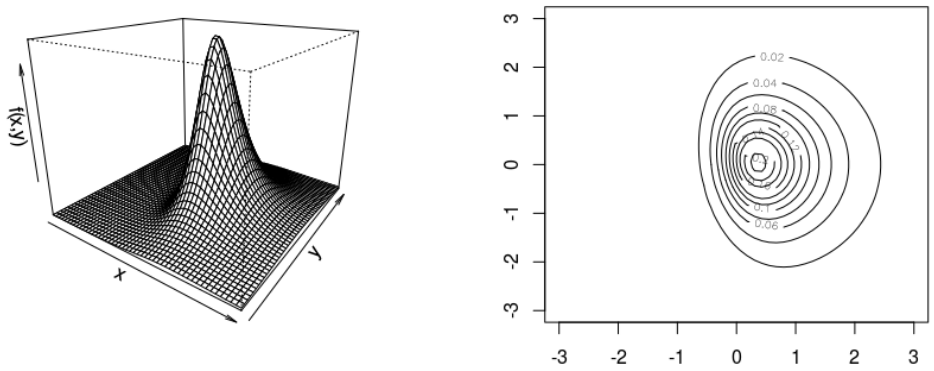


Figura 3.4: Densidade skew-t bivariada.

Assim como a distribuição skew-normal, a distribuição skew-t possui um parâmetro de assimetria. Na Figura 3.4, encontram-se gráficos com esta distribuição sendo os seguintes parâmetros utilizados: $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbb{I}_2$, $\nu = 2$ e $\boldsymbol{\eta}^\top = [2, 0.1]^\top$. Esta distribuição se torna a distribuição t-Student quando o vetor de parâmetros de assimetria for nulo. E se a situação anterior ocorrer simultaneamente quando os graus de liberdade forem para o infinito, esta distribuição se aproxima do modelo normal.

3.2 Distâncias estocásticas

Uma medida de divergência pode ser entendida como uma função não-negativa entre duas medidas de probabilidade, diga-se $D(\mathbf{X}||\mathbf{Y})$, que obedece a propriedade de definitividade:

1. $D(\mathbf{X}||\mathbf{Y}) \geq 0$ (não-negatividade),

2. $D(\mathbf{X}||\mathbf{Y}) = 0 \Leftrightarrow \mathbf{X} = \mathbf{Y}$ (definitividade).

Assim, uma classe de divergências entre os vetores aleatórios $\mathbf{X}$ e $\mathbf{Y}$ com fdp $f_{\mathbf{X}}$ e $f_{\mathbf{Y}}$, respectivamente, é definida por

$$D_{\phi}^{h}(\mathbf{X}||\mathbf{Y}) = h\left(\mathrm{E}_{\mathbf{Y}}\, \phi\left(\frac{f_{\mathbf{X}}(\mathbf{Y};\boldsymbol{\theta}_1)}{f_{\mathbf{Y}}(\mathbf{Y};\boldsymbol{\theta}_2)} \right) \right), \tag{3.3}$$

em que $\mathrm{E}_{\mathbf{Y}}(\cdot)$ é o valor esperado de um argumento com respeito a um vetor aleatório $\mathbf{Y}$, $h\colon (0,\infty) \to [0,\infty)$ é uma função estritamente crescente e $\phi\colon (0,\infty) \to [0,\infty)$ é uma função convexa tal que $0\,\phi(0/0) = 0$, $0\,\phi(x/0) = \lim_{x\to\infty} \phi(x)/x$ e $\phi(1) = h(0) = 0$ e $\phi''(1) > 0$. Adicionalmente, para $\boldsymbol{X}$ e $\boldsymbol{Y}$ sendo duas matrizes aleatórias com funções densidades $f_{\boldsymbol{X}}(\boldsymbol{\mathcal{Z}};\boldsymbol{\theta}_1)$ e $f_{\boldsymbol{Y}}(\boldsymbol{\mathcal{Z}};\boldsymbol{\theta}_2)$ tem-se que a divergência entre $\boldsymbol{X}$ e $\boldsymbol{Y}$ é dada por

$$D_{\phi}^{h}(\boldsymbol{X} \parallel \boldsymbol{Y}) = h\left(\mathrm{E}_{\boldsymbol{Y}}\, \phi\left(\frac{f_{\boldsymbol{X}}(\boldsymbol{Y};\boldsymbol{\theta}_1)}{f_{\boldsymbol{Y}}(\boldsymbol{Y};\boldsymbol{\theta}_2)} \right) \right). \tag{3.4}$$

Esta classe foi proposta por Salicrú *et al.* (1994) e é denotada por $h-\phi$ divergência. Em geral, uma função é considerada medida de distância estocástica, diga-se $d_{\phi}^{h}(\cdot,\cdot)$, entre dois vetores aleatórios se as seguintes propriedades foram satisfeitas (NASCIMENTO, 2012): seja $\boldsymbol{x}$ e $\boldsymbol{y}$ dois vetores aleatórios definidos sobre o mesmo suporte $\boldsymbol{\mathcal{Z}}$ tendo densidades $f_{\boldsymbol{x}}(\boldsymbol{z};\boldsymbol{\theta}_1)$ e $f_{\boldsymbol{y}}(\boldsymbol{z};\boldsymbol{\theta}_2)$, respectivamente, em que $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ são vetores de parâmetros, então

1. $d_{\phi}^{h}(\mathbf{X},\mathbf{Y}) \geq 0$ (não-negatividade),
2. $d_{\phi}^{h}(\mathbf{X},\mathbf{Y}) = d_{\phi}^{h}(\mathbf{Y},\mathbf{X})$ (simetria),
3. $d_{\phi}^{h}(\mathbf{X},\mathbf{Y}) = 0 \Leftrightarrow \mathbf{X} = \mathbf{Y}$ (definitividade).

Como discutido por Nascimento (2012), definir-se-ão distâncias como divergências simetrizadas. Com base neste princípio, divergências foram submetidas a um processo de simetrização,

$$d_{\phi}^{h}(\mathbf{X},\mathbf{Y}) = \frac{D_{\phi}^{h}(\mathbf{X} \parallel \mathbf{Y}) + D_{\phi}^{h}(\mathbf{Y} \parallel \mathbf{X})}{2}, \tag{3.5}$$

com o fim de serem obtidas distâncias para utilização nesta dissertação.

Na Tabela 3.1, as divergências e distâncias de Kullback-Leibler, Jeffreys e Rényi são relacionadas aos pares $[h(x), \phi(x)]$.

Tabela 3.1: Distâncias e divergências $h-\phi$ e funções ϕ e h relacionadas.

divergência $(h-\phi)$	$h(x)$	$\phi(x)$
Kullback-Leibler $\left[\int f_{\boldsymbol{X}} \log \frac{f_{\boldsymbol{X}}}{f_{\boldsymbol{Y}}}\right]$	x	$x\log(x)$
Jeffreys $\left[\int (f_{\boldsymbol{X}} - f_{\boldsymbol{Y}}) \log \frac{f_{\boldsymbol{X}}}{f_{\boldsymbol{Y}}}\right]$	x	$(x-1)\log(x)$
Rényi (ordem α) $\left[\frac{\log \int f_{\boldsymbol{X}}^{\alpha} f_{\boldsymbol{Y}}^{1-\alpha}}{(\alpha-1)}\right]$	$\frac{1}{\alpha-1}\log((\alpha-1)x+2)$, $0 \le x < \frac{1}{1-\alpha}$	$\frac{x^{\alpha}-\alpha(x-1)-2}{\alpha-1}$, $0<\alpha<1$
distância $(h-\phi)$		
Kullback-Leibler $\left[\frac{1}{2}\int (f_{\boldsymbol{X}} - f_{\boldsymbol{Y}}) \log \frac{f_{\boldsymbol{X}}}{f_{\boldsymbol{Y}}}\right]$	$\frac{x}{2}$	$(x-1)\log(x)$
Rényi (ordem α) $\left[\frac{\log \int f_{\boldsymbol{X}}^{\alpha} f_{\boldsymbol{Y}}^{1-\alpha} + \log \int f_{\boldsymbol{X}}^{1-\alpha} f_{\boldsymbol{Y}}^{\alpha}}{2(\alpha-1)}\right]$	$\frac{1}{\alpha-1}\log((\alpha-1)x+1)$, $0 \le x < \frac{1}{1-\alpha}$	$\frac{x^{1-\alpha}+x^{\alpha}-\alpha(x-1)-2}{2(\alpha-1)}$, $0<\alpha<1$

3.2.1 Divergências de KL e J para distribuição normal multivariada

Sejam $\mathbf{X} \sim N_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ e $\mathbf{Y} \sim N_k(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ dois vetores normais k variados. Então a divergência $\mathrm{D_{KL}}$ entre $\mathbf{X}$ e $\mathbf{Y}$ é dada por (CONTRERAS-REYES; ARELLANO-VALLE, 2012):

$$\mathrm{D}_{\mathrm{KL}}^{N}\left([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1]||[\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2]\right) = \frac{1}{2}\Bigg\{ \log\left(\frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|}\right) + \mathrm{tr}(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\top}\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) - k \Bigg\}. \tag{3.6}$$

A divergência de Jeffreys ou "J-Divergence", $\mathrm{J}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1], [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$, pode ser expressada em termos de $\mathrm{D_{KL}}$ como

$$\mathrm{J}^{N}\left([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1]||[\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2]\right) = \mathrm{D}_{\mathrm{KL}}^{N}\left([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1]||[\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2]\right) + \mathrm{D}_{\mathrm{KL}}^{N}\left([\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2]||[\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1]\right). \tag{3.7}$$

As quantidades $\mathrm{D}_{\mathrm{KL}}^{N}$ e J^{N} serão calculadas avaliando EMVs obtidos a partir de duas amostras em (3.6) e (3.7). Os classificadores baseados em (3.6) e (3.7) são denotados por classificador de Kullback-Leibler para distribuição normal (CKLN) e classificador de Jeffreys para distribuição

Normal (CJN).

3.2.2 Divergências KL e J para distribuição skew-normal multivariada

Contreras-Reyes e Arellano-Valle (2012) utilizaram a divergência de KL com base na distribuição skew-normal para detectar *clusters* no catálogo sísmico do serviço sismologico nacional do Chile. Resultados mostraram que esta medida é uma ferramenta útil. Sejam $\mathbf{X} \sim SN_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1)$ e $\mathbf{Y} \sim SN_k(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2)$ dois vetores k dimensionais seguindo uma distribuição skew-normal multivariada. Neste caso, a divergência de KL é dada por (CONTRERAS-REYES, 2014):

$$D_{\mathrm{KL}}^{SN}\left([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1]||[\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2]\right) = D_{\mathrm{KL}}^{N}\left([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1]||[\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2]\right) + \frac{\sqrt{\frac{2}{\pi}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\delta}_1}{\boldsymbol{\Sigma}_2^{-1}} + \mathrm{E}[\log\{2\Phi(W_{11})\}] - \mathrm{E}[\log\{2\Phi(W_{21})\}], \tag{3.8}$$

em que $W_{11} \sim SN_1\left(\mathbf{0}, \boldsymbol{\eta}_1^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_1, (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_1)^{\frac{1}{2}}\right)$ e $W_{21} \sim SN_1\left(\boldsymbol{\eta}_2^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \boldsymbol{\eta}_2^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_2, \boldsymbol{\Xi}\right)$ com $\boldsymbol{\Xi} = \frac{\boldsymbol{\eta}_2^\top \boldsymbol{\delta}_1}{\sqrt{\boldsymbol{\eta}_2^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_2 - (\boldsymbol{\eta}_2^\top \boldsymbol{\delta}_1)^2}}$ e $\boldsymbol{\delta}_1 = \frac{\boldsymbol{\Sigma}_1 \boldsymbol{\eta}_1}{\sqrt{1 + \boldsymbol{\eta}_1^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_1}}$. Pode-se notar que a Equação (3.8) apresenta duas quantidades $\mathrm{E}[\log\{2\Phi(W_{11})\}]$ e $\mathrm{E}[\log\{2\Phi(W_{21})\}]$ que não possui expressões de forma fechada. Para superar essa dificuldade, utilizamos a integração de Monte Carlo. Para este fim, foi utilizado o algoritmo proposto por Contreras-Reyes e Arellano-Valle (2012) para gerar W_{11} e W_{21}:

- $W_{11} \equiv 0 + \boldsymbol{\delta}|\mathrm{U}_0| + \mathbf{U}$ com $\mathrm{U}_0 \sim N(0,1)$, $\mathbf{U} \sim N_1\left(\mathbf{0}, (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_1) - \boldsymbol{\delta}^2\right)$ e $\boldsymbol{\delta} = \dfrac{(\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_1)^{\frac{3}{2}}}{\sqrt{1 + (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_1)^2}}$;
- $W_{21} \equiv \boldsymbol{\eta}_2^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \boldsymbol{\delta}|\mathrm{U}_0| + \mathbf{U}$ com $\mathrm{U}_0 \sim N(0,1)$, $\mathbf{U} \sim N_1\left(\mathbf{0}, (\boldsymbol{\eta}_2^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_2) - \boldsymbol{\delta}^2\right)$, $\boldsymbol{\delta} = \frac{(\boldsymbol{\eta}_2^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_2)\left(\frac{\boldsymbol{\eta}_2^\top \boldsymbol{\delta}_1}{\sqrt{\boldsymbol{\eta}_2^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_2 - (\boldsymbol{\eta}_2^\top \boldsymbol{\delta}_1)^2}}\right)}{\sqrt{1 + \boldsymbol{\eta}^\top (\boldsymbol{\eta}_2^\top \Omega_1 \boldsymbol{\eta}_2)\boldsymbol{\eta}}}$ e $\boldsymbol{\eta} = \frac{\boldsymbol{\eta}_2^\top \boldsymbol{\delta}_1}{\sqrt{\boldsymbol{\eta}_2^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_2 - (\boldsymbol{\eta}_2^\top \boldsymbol{\delta}_1)^2}}$.

A assimetria $\boldsymbol{\eta}$ foi estimada maximizando iterativamente a função de verossimilhança (ARELLANO-VALLE *et al.*, 2004). Como este problema não apresenta expressão de forma fechada, foi considerado o método de Newton-Raphson: Para $j = 1, \ldots, k$,

$$\widehat{\eta}_j \;=\; \arg\max_{\eta_j \in \mathbb{R}} \left\{ -\frac{n}{2}\log\frac{\pi\sigma^2}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i - \mu)^2 + \sum_{i=1}^{n}\log\Phi\left[\frac{\eta_j(z_i - \mu)}{\sqrt{\sigma^2}}\right] \right\}. \tag{3.9}$$

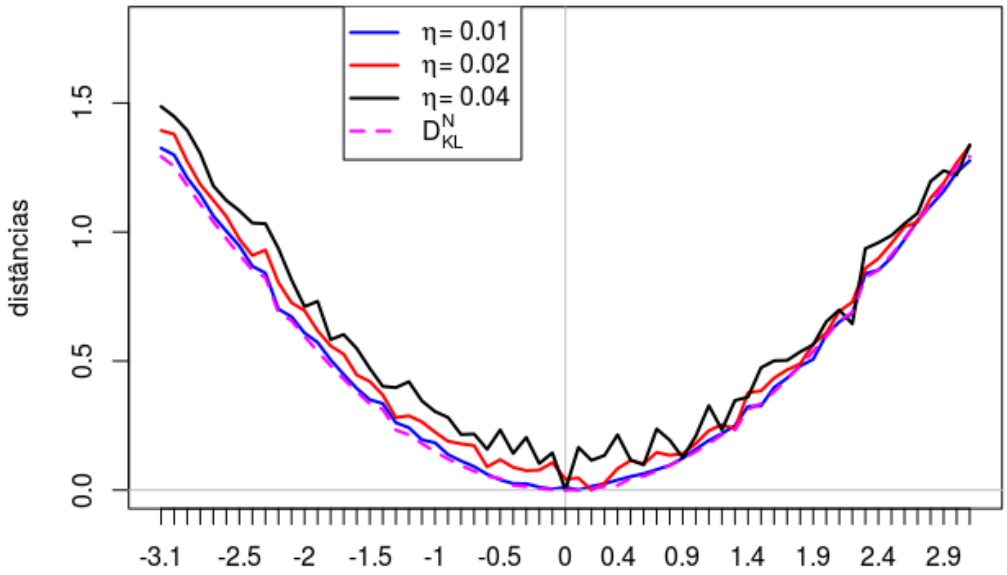


Figura 3.5: Comportamento da divergência KL para a distribuição skew-normal para diferentes coeficientes de assimetria.

A Figura 3.5 mostra o comportamento da classificação utilizando a divergência KL baseada na distribuição skew-normal, flexibilizando a assimetria. Nota-se que quando a assimetria vai para zero (a curva com menor assimetria cor azul) as distâncias se aproxima da curva da divergência de KL baseado na distribuição normal (curva tracejada cor rosa). O classificador com base em (3.8) é chamado de classificador de Kulback-Leibler para a distribuição skew-normal (CKLSN). Aquele com base em uma simetrização de (3.8) é chamado de classificador de Jeffreys para skew-normal (CJSN).

3.2.3 Divergências de KL e J para distribuição t-Student multivariada

Seja $\boldsymbol{X} \in \mathbb{R}^k$ uma variável aleatória contínua com fdp sendo $f_{\boldsymbol{X}}$. A entropia para $\boldsymbol{X}$ é dada por $H(\boldsymbol{X}) = -\int f_{\boldsymbol{X}}(\boldsymbol{x}) \log f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}$ e sejam $\boldsymbol{X}$ e $\boldsymbol{Y} \in \mathbb{R}^k$ duas variáveis aleatórias contínuas com fdp $f_{\boldsymbol{X}}$ e $f_{\boldsymbol{Y}}$, a cross-entropia é definida como $CH(\boldsymbol{X}, \boldsymbol{Y}) = -\int f_{\boldsymbol{X}}(\boldsymbol{x}) \log f_{\boldsymbol{Y}}(\boldsymbol{x}) d\boldsymbol{x}$. A divergência de KL pode ser definida em termos de $H(\boldsymbol{X})$ e $CH(\boldsymbol{X}, \boldsymbol{Y})$, como

$$KL(\boldsymbol{X}, \boldsymbol{Y}) = \int f_{\boldsymbol{X}}(\boldsymbol{x}) \log \left\{ \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{\boldsymbol{Y}}(\boldsymbol{x})} \right\} d\boldsymbol{x} = CH(\boldsymbol{X}, \boldsymbol{Y}) - H(\boldsymbol{X}),$$

com $CH(\boldsymbol{X}, \boldsymbol{Y}) \geq H(\boldsymbol{X})$, com a igualdade se $\boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{Y}$. Portanto, sejam $\mathbf{X} \sim \mathrm{t}_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1)$ e

$\mathbf{Y} \sim \mathrm{t}_k(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2)$. A Entropia de $\mathbf{X}$ é dada por (CONTRERAS-REYES, 2014):

$$H([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1]) = H(\mathbf{X}_0) + \frac{1}{2}\log|\boldsymbol{\Sigma}_1|, \tag{3.10}$$

em que $H(\mathbf{X}_0) = -\log\{B_k(\nu_1)\} + \left(\frac{\nu_1+k}{2}\right)\left[\psi\left(\frac{\nu_1+k}{2}\right) - \psi\left(\frac{\nu_1}{2}\right)\right]$ e

$$B_k(\nu_1) = \frac{\Gamma\left(\frac{\nu_1+k}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)(\nu_1\pi)^{\frac{k}{2}}}, \quad \nu_1 > 0.$$

A cross-entropia para t-Student é dada por:

$$CH([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1]||[\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2]) \approx \frac{1}{2}\log\left\{(2\pi)^k|\boldsymbol{\Sigma}_2|\right\} + \frac{1}{2}\left(\frac{\nu_2+k}{\nu_2}\right)\mathbb{E}[\mathbf{Q}_2], \tag{3.11}$$

em que $\mathbb{E}[\mathbf{Q}_2] = \kappa_2(X)\mathrm{tr}(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$ e $\kappa_2(X) = \frac{\nu_1}{(\nu_1-2)}$. Assim, a divergência de Kullback-Leibler para t-Student é dada por:

$$\begin{aligned} D_{\mathrm{KL}}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1]||[\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2]) &= CH([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1]||[\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2]) \\ &\quad - H([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1]). \end{aligned} \tag{3.12}$$

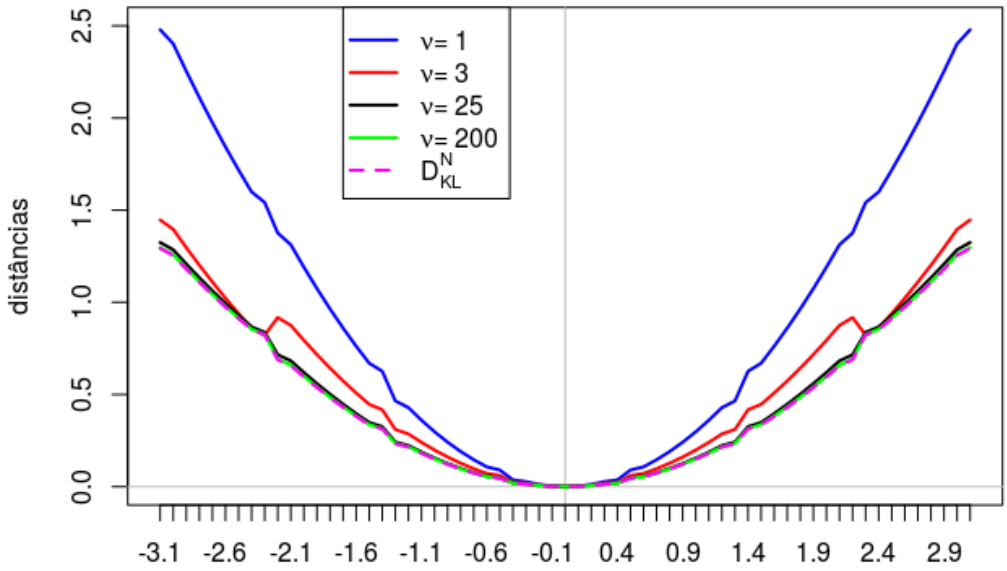


Figura 3.6: Comportamento da divergência KL para a distribuição t-Student para diferentes graus de liberdade.

A distribuição t-Student se aproxima da distribuição normal quando os graus de liberdade tendem ao infinito (parâmetro $\nu \to \infty$), o comportamento da divergência KL baseado na distribuição t-Student, manteve o mesmo comportamento. Como pode ser observado na Figura

3.6, a curva verde quando os graus de liberdade são $\nu = 200$ se sobrepõe a curva da divergência KL para o modelo Normal. Os classificadores com base em (3.12) e em sua simetrização são denotados como classificador KL para distribuição t-Student (CKLt) e classificador J para a distribuição t-Student (CJt).

3.2.4 Divergências de KL e J para distribuição skew-t multivariada

Sejam $\mathbf{X} \sim \mathrm{St}_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1, \nu_1)$ e $\mathbf{Y} \sim \mathrm{St}_k(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2, \nu_2)$ então a entropia é obtida por:

$$H([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1, \nu_1]) = H(\mathbf{X}_0) - I\!E\left[\log\left\{2T\left(\frac{\sqrt{\nu_1+k}\parallel \overline{\boldsymbol{\eta}_1} \parallel W}{\sqrt{\nu_1+k-1+W^2}}; \nu_1+k\right)\right\}\right],$$

em que $W \sim ST_1(0, 1, \parallel \overline{\boldsymbol{\eta}_1} \parallel, \nu_1+k-1)$, $\parallel \overline{\boldsymbol{\eta}_1} \parallel = \overline{\boldsymbol{\eta}_1}^\top \overline{\boldsymbol{\eta}_1} = (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma}_1 \boldsymbol{\eta}_1)^{\frac{1}{2}}$ e $H(\mathbf{X}_0)$ é a entropia de $\mathbf{X}_0 \sim t_k(\mathbf{0}, \mathbf{I}_k, \nu_1)$ dada por

$$H(\mathbf{X}_0) = -\log\{B_k(\nu_1)\} + \left(\frac{\nu_1+k}{2}\right)\left[\psi\left(\frac{\nu_1+k}{2}\right) - \psi\left(\frac{\nu_1}{2}\right)\right] \quad \text{e}$$

$$B_k(\nu_1) = \frac{\Gamma\left(\frac{\nu_1+k}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)(\nu_1\pi)^{\frac{k}{2}}}, \text{ para } \nu_1 > 0.$$

A cross-entropia pode ser aproximado por (CONTRERAS-REYES, 2014):

$$\begin{aligned} CH([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1, \nu_1] \parallel [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2, \nu_2]) \approx \; & \frac{1}{2}\log\left\{(2\pi)^k|\boldsymbol{\Sigma}_2|\right\} \\ & + \frac{1}{2}\left(\frac{\nu_2+k}{\nu_2}\right) I\!E[\mathbf{Q}_2] \\ & - I\!E\left[\log\left\{2\mathrm{T}\left(\overline{\boldsymbol{\eta}}_2^\top(\mathbf{X}-\boldsymbol{\mu}_2); \nu_2+k\right)\right\}\right], \end{aligned} \tag{3.13}$$

quando $\nu_2 \to \infty$ e $\overline{\boldsymbol{\eta}}_2^\top = \boldsymbol{\eta}_2^\top\sqrt{\frac{\nu_2+k}{\nu_2+\mathbf{Q}_2}}$, $\mathbb{E}[\mathbf{Q}_2] = \mathrm{K}_2(\mathbf{X})\mathrm{tr}(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) + 2K_1(\mathbf{X})(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}_1$, em que $K_1(\mathbf{X}) = \frac{\sqrt{\frac{\nu_1}{2}}\Gamma\left(\frac{\nu_1-1}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)}$, $K_2(\mathbf{X}) = \frac{\nu_1}{(\nu_1-2)}$ e $\boldsymbol{\delta}_1 = \frac{\boldsymbol{\Sigma}_1\boldsymbol{\eta}_1}{\sqrt{1+\boldsymbol{\eta}_1{}^\top\boldsymbol{\Sigma}_1\boldsymbol{\eta}_1}}$.

A divergência de Kullback-Leibler para skew-t, é obtida da seguinte maneira:

$$\begin{aligned} D_{KL}^{St}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1, \nu_1] \parallel [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2, \nu_2]) = \; & CH([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1, \nu_1] \parallel [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2, \nu_2]) \\ & - H([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1, \nu_1]). \end{aligned} \tag{3.14}$$

O classificador em função de (3.14) ou na simetrização desta são denotados por classificador de KL para o modelo skew-t multivariado (CKLSt) e classificador de J para o modelo skew-t (CJSt).

3.2.5 Divergência de Rényi para as distribuições normal multivariada e WCE

Para intensidades multivariadas, considere $\boldsymbol{\theta}_1 = (\boldsymbol{\mu}_1^\top, \text{vec}(\boldsymbol{\Sigma}_1)^\top)^\top$, $\boldsymbol{\theta}_2 = (\boldsymbol{\mu}_2^\top, \text{vec}(\boldsymbol{\Sigma}_2)^\top)^\top$ e $K(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2;\alpha) = \int_{\mathbb{R}^P} f(\mathbf{z};\boldsymbol{\theta}_1)^\alpha f(\mathbf{z};\boldsymbol{\theta}_2)^{1-\alpha} \text{d}\mathbf{z}$, em que $\mathbf{z} \in \mathbb{R}^P$, $\text{d}\mathbf{z} = \prod_{i=1}^P \text{d}z_i$, $\boldsymbol{\mu}_i \in \mathbb{R}^P$, $\boldsymbol{\Sigma}_i \in \{\mathbf{Z} \in \mathbb{R}^P \times \mathbb{R}^P : \mathbf{Z} \succeq \mathbf{0}\}$ e vec$(\cdot)$ é o operador de vetorização de um argumento matricial.

Pode-se mostrar que para a normalidade multivariada,

$$\begin{aligned} K(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2;\alpha) = \; & \frac{|\alpha\boldsymbol{\Sigma}_2 + (1-\alpha)\boldsymbol{\Sigma}_1|^{-\frac{1}{2}}}{|\boldsymbol{\Sigma}_1|^{\frac{\alpha-1}{2}}|\boldsymbol{\Sigma}_2|^{-\frac{\alpha}{2}}} \\ & \times \exp\left\{\tfrac{\alpha(\alpha-1)}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top[\alpha\boldsymbol{\Sigma}_2 + (1-\alpha)\boldsymbol{\Sigma}_1]^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right\}. \end{aligned} \tag{3.15}$$

Deste resultado, a divergência de Rényi definida em $\alpha > 0$ e $\alpha \neq 1$ é dada por

$$\begin{aligned} \text{D}_\text{R}^\alpha(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) &= \frac{1}{\alpha(\alpha-1)}\log\left[\text{K}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2;\alpha)\right] \\ &= (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top[\alpha\boldsymbol{\Sigma}_2 + (1-\alpha)\boldsymbol{\Sigma}_1]^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \\ &\quad -\frac{1}{2\alpha(\alpha-1)}\log\left[\frac{|\alpha\boldsymbol{\Sigma}_2 + (1-\alpha)\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_1|^{(1-\alpha)}|\boldsymbol{\Sigma}_2|^\alpha}\right]. \end{aligned} \tag{3.16}$$

Neste estudo, utilizam-se quatro medidas de distâncias baseadas na Rényi, sendo duas divergências e duas distâncias:

$$\begin{cases} 1. \quad \text{D}_\text{R}^\alpha(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2); \\ 2. \quad \text{D}_{\text{R}2}^\alpha(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1); \\ 3. \quad d_R^\alpha(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) = \dfrac{D_R^\alpha(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) + D_R^\alpha(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1)}{2}; \\ 4. \quad d_{R2}^\alpha(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) = \dfrac{1}{\alpha(\alpha-1)}\log\left[\frac{\text{K}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2;\alpha)+\text{K}(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1;\alpha)}{2}\right]. \end{cases} \tag{3.17}$$

Para as entradas de uma imagem *PolSAR* dadas por (3.1), considere $\boldsymbol{\theta}_1 = (L_1, \text{vec}(\boldsymbol{\Sigma}_1)^\top)^\top$ e $\boldsymbol{\theta}_2 = (L_2, \text{vec}(\boldsymbol{\Sigma}_2)^\top)^\top$, em que $\boldsymbol{\Sigma}_1$ e $\boldsymbol{\Sigma}_2$ são matrizes hermitianas positivas definidas e L_1 e L_2 são os números de *looks*. As divergências e distâncias de Rényi nesses casos podem ser obtidas como segue: (Assuma $L_1 = L_2 = L$, isto é, que os números de *looks* são equivalentes)

$$
\begin{cases}
1. \quad D^{\alpha}_{R_{\mathbb{C}}}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) = & \dfrac{n}{\alpha-1}\log\left[\dfrac{|(\alpha\boldsymbol{\Sigma}_1^{-1}+(1-\alpha)\boldsymbol{\Sigma}_2^{-1})^{-1}|}{|\boldsymbol{\Sigma}_1|^{\alpha}|\boldsymbol{\Sigma}_2|^{1-\alpha}}\right]; \\
2. \quad D^{\alpha}_{R2_{\mathbb{C}}}(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1) = & \dfrac{n}{\alpha-1}\log\left[\dfrac{|(\alpha\boldsymbol{\Sigma}_2^{-1}+(1-\alpha)\boldsymbol{\Sigma}_1^{-1})^{-1}|}{|\boldsymbol{\Sigma}_2|^{\alpha}|\boldsymbol{\Sigma}_1|^{1-\alpha}}\right]; \\
3. \quad d^{\alpha}_{R_{\mathbb{C}}}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) = & \dfrac{D^{\alpha}_{R}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2)+D^{\alpha}_{R}(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1)}{2}; \\
4. \quad d^{\alpha}_{R2_{\mathbb{C}}}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) = & \dfrac{\log 2}{1-\alpha}+\dfrac{1}{\alpha-1}\times\log\left\{\left[\dfrac{|(\alpha\boldsymbol{\Sigma}_1^{-1}+(1-\alpha)\boldsymbol{\Sigma}_2^{-1})^{-1}|}{|\boldsymbol{\Sigma}_1|^{\alpha}|\boldsymbol{\Sigma}_2|^{(1-\alpha)}}\right]^n\right. \\
& \left.+\left[\dfrac{|(\alpha\boldsymbol{\Sigma}_2^{-1}+(1-\alpha)\boldsymbol{\Sigma}_1^{-1})^{-1}|}{|\boldsymbol{\Sigma}_2|^{\alpha}|\boldsymbol{\Sigma}_1|^{(1-\alpha)}}\right]^n\right\}.
\end{cases}
\tag{3.18}
$$

Note que todas as medidas desta seção são classes de medidas. Para cada valor de α (ordem da Rényi), tem-se uma divergência/distância. Em particular, valem-se:$D^{\alpha}_{R}(\cdot||\cdot) \longrightarrow D_{KL}(\cdot||\cdot)$,

$$-\log D^{\alpha}_{R}(\cdot||\cdot) \overset{\alpha=\frac{1}{2}}{=} -\log\int_{\mathbb{R}^p}\sqrt{f(\boldsymbol{z},\boldsymbol{\theta}_1)f(\boldsymbol{z},\boldsymbol{\theta}_2)}d\boldsymbol{z}$$

e

$$1-D^{\alpha}_{R}(\cdot||\cdot) \overset{\alpha=\frac{1}{2}}{=} 1-\int_{\mathbb{R}^p}\sqrt{f(\boldsymbol{z},\boldsymbol{\theta}_1)f(\boldsymbol{z},\boldsymbol{\theta}_2)}d\boldsymbol{z}.$$

As duas últimas divergências são chamadas de divergências de Bhattacharyya e de Hellinger. Os classificadores como base em (3.17) serão denotados por:

CRN1:α: Classificador da divergência de Rényi para distribuição normal,

CRN2:α: Classificador da divergência de Rényi simétrico para distribuição normal,

CRND1:α: Classificador da distância de Rényi para distribuição normal e

CRND2:α: Classificador da distância de Rényi 2 para distribuição normal.

Por outro lado, os classificadores com base em (3.18) serão denotados por:

CRWC1:α: Classificador da divergência de Rényi para Wishart complexa escalonada,

CRWC2:α: Classificador da diver. de Rényi simétrico para Wishart complexa escalonada,

CRWCD1:α: Classificador da distância de Rényi para Wishart complexa escalonada e

CRWCD2:α: Classificador da distância de Rényi 2 para Wishart complexa escalonada.

CAPÍTULO 4

Classificadores Clássicos e Baseados em TEI para Dados *PolSAR*

Inicialmente, assume-se que o suporte da uma imagem *PolSAR* é particionado em r segmentos, diga-se, $C_1, \ldots, C_r$. Os atributos dos dados *PolSAR* para cada segmento são definidos como

$$\begin{cases} [\,\boldsymbol{z}_i|C_k\,] = \left[\begin{array}{c} Z_i^{\text{Channel HH}} \\ Z_i^{\text{Channel HV}} \\ Z_i^{\text{Channel VV}} \end{array}\middle| C_k \right], \text{ para intensidades multivariadas,} \\ [\,\boldsymbol{Z}_i|C_k\,] = \left[\begin{array}{ccc} ||Z_i^{\text{Channel HH}}||^2 & Z_i^{\text{Channel HH HV}} & Z_i^{\text{Channel HH VV}} \\ & ||Z_i^{\text{Channel HV}}||^2 & Z_i^{\text{Channel HV VV}} \\ & & ||Z_i^{\text{Channel VV}}||^2 \end{array}\middle| C_k \right], \\ \text{para dados } \textit{full PolSAR}, \end{cases}$$

em que $i = 1, \ldots, N_k$, $k = 1, \ldots, r$ e N_k é o número de *pixels* no k-ésimo segmento, tal que $N = \sum_{k=1}^{r} N_k$. Em problemas de classificação tradicionais, esses r conjuntos são chamados *amostras de treinamento*. Em outras palavras,

> "$[\boldsymbol{z}_1|C_k], \ldots, [\boldsymbol{z}_{n_k}|C_k]$" representa uma *amostra aleatória* (independente e identicamente distribuída) obtida de
>
> $$[\boldsymbol{z}|C_k] \sim \left\{ N_3(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),\ SN_3(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{\eta}_k),\ \mathrm{t}_3(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \nu_k),\ \mathrm{St}_3(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{\eta}_k, \nu_k) \right\},$$
>
> para todo $k = 1, \ldots, r$, ou sob a hipótese de que "$[\boldsymbol{Z}_1|C_k], \ldots, [\boldsymbol{Z}_{n_k}|C_k]$" indica uma amostra aleatória obtida de
>
> $$[\boldsymbol{Z}|C_k] \sim \mathcal{W}_3^{\mathbb{C}}(\boldsymbol{\Sigma}_k, L_k), \text{ para todo } k = 1, \ldots, r.$$

Subsequentemente, ajustam-se os respectivos modelos vetoriais e matriciais via estimação por máxima verossimilhança como discutida previamente.

4.1 *K* vizinhos mais próximos - KNN

Existem situações em que as suposições probabilísticas não são satisfeitas. A utilização de métodos não paramétricos é uma solução plausível para tais situações. Dentre estes métodos, destaca-se o KNN proposto por Cover e Hart (1967). Este método é baseado no cálculo da probabilidade de que uma observação pertença a cada grupo sem suposto probabilístico. Dado um inteiro positivo K e uma observação de teste x, o KNN trabalha como segue (JAMES *et al.*, 2013):

1. identifica os k pontos mais próximos de x baseado nos N_k elementos dos r conjuntos da amostra de treinamento;

2. estima-se a probabilidade condicional para a classe r como a fração de pontos em N cujos valores de resposta são iguais a r;

3. finalmente, KNN aplica a regra de Bayes e classifica a observação de teste x para a classe com maior probabilidade.

$$I\!P(Y = r|X = x) = \frac{1}{K} \sum_{i \in N} I\!I(y_i = r). \tag{4.1}$$

Tal algoritmo pode ser definido como:

$$\widehat{i}_{\text{KNN}} = \underset{\text{i}\in\{1,\ldots,\text{r}\}}{\arg\max}\, I\!P(Y = r|X = x).$$

4.2 Análise discriminante linear - LDA

A LDA é uma técnica estatística supervisionada usada para encontrar uma combinação linear de características que separa duas ou mais classes de objetos. Esta técnica está relacionada com a análise de variância e análise de regressão linear, que também tenta expressar uma variável dependente como uma combinação linear de outras características (HAIR *et al.*, 2009). A LDA consiste em classificar uma observação x no grupo cuja probabilidade de x pertencer ao mesmo seja máxima. A probabilidade é calculada sob o suposto de normalidade por meio do teorema de Bayes (JAMES *et al.*, 2013):

$$I\!P(Y = r|\boldsymbol{X} = \boldsymbol{x}) = \frac{\pi_k f_k(\boldsymbol{x})}{\sum_{l=1}^{k} \pi_l f_l(\boldsymbol{x})}, \tag{4.2}$$

em que $f_k(\boldsymbol{x})$ é a fdp da distribuição normal multivariada que foi definida na Seção 3.1.1 e π_k representa a `probabilidade a priori` de que uma observação escolhida aleatoriamente pertença a k-ésima classe.

A LDA supõe que as diferenças entre os grupos podem ser observada levando em conta as diferenças das médias; isto é, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_j$ para todo $i, j = 1, \ldots, r$. Portanto, a regra de classificação do LDA é definida por

$$\widehat{i}_{\text{LDA}} = \underset{\text{i}\in\{1,\ldots,\text{r}\}}{\arg\max}\, I\!P(Y = r|\boldsymbol{X} = \boldsymbol{x}) = \underset{\text{i}\in\{1,\ldots,\text{r}\}}{\arg\min}\, d_M^2(\boldsymbol{x}, [\boldsymbol{\mu}_k, \boldsymbol{\Sigma}]) + \log(\pi_k)$$

em que a última quantidade pode ser obtida maximizando a expressão (4.2) e para $d_M^2(\boldsymbol{x}, [\boldsymbol{\mu}_k, \boldsymbol{\Sigma}]) = \boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k + \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{x}$ (RODRÍGUEZ, 2015).

4.3 Análise discriminante quadrática - QDA

A QDA não impõe nenhuma restrição sobre $\boldsymbol{\Sigma}_k$; isto é, as diferenças podem ser percebidas nos vetores de médias e nas estruturas de covariâncias (JAMES *et al.*, 2013). Portanto a regra de classificação para o QDA é dada por

$$\widehat{i}_{\text{QDA}} = \underset{\text{i}\in\{1,\ldots,\text{r}\}}{\arg\max}\, I\!P(Y = r|\boldsymbol{X} = \boldsymbol{x}) = \underset{\text{i}\in\{1,\ldots,\text{r}\}}{\arg\min}\, d_M^2(\boldsymbol{x}, [\boldsymbol{\mu}_k, \boldsymbol{\Sigma}]) + \log|\boldsymbol{\Sigma}_k| + \log(\pi_k),$$

em que $d_M^2(\boldsymbol{x}, [\boldsymbol{\mu}_k, \boldsymbol{\Sigma}]) = \boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k + \boldsymbol{x}^\top \boldsymbol{\Sigma}_k^{-1}\boldsymbol{x}$ (RODRÍGUEZ, 2015). A LDA é um caso particular da QDA quando $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$ para todo $k = 1, \ldots, r$.

4.4 Máquina de vetores de suporte - SVM

O SVM se baseia em idéias simples oriundas da Teoria da Aprendizagem Estatística (VAPNIK, 1998). No que segue, apresentam-se prévias discursões sobre SVM nas versões linear e não linear.

4.4.1 SVM - Caso linear

Será considerado um problema de classificação com duas classes e P amostras de treinamento. Cada amostra é descrito por um vetor k dimensional $V_j^{(k)} = V_j(j \geq k)$, pertencente a um espaço de entrada E (com a mesma dimensão k). As k componentes de vetores são também chamadas de primitivas ou, neste caso, de indicadores polarimétricos. O rótulo da j-ésima amostra V_j é Y_j. Portanto, para um problema com duas classes, os rótulos das classes são -1 ou $+1$(isto é, $Y_j \in \{-1; +1\}$).

O problema de classificação consiste em encontrar um hiperplano ótimo (um sub-espaço de dimensão $k-1$) para separar as duas classes. Tal hiperplano é definido pelo seu vetor normal ω e um deslocamento b. Uma função de classificação f definindo o rótulo Y_j da amostra V_j é

definido por

$$f(V_j) = Sgn((\omega.V_j) + b) = Y_j \tag{4.3}$$

em que Sgn é a função sinal, e $(\omega.V_j)$ representa o produto escalar entre ω e V_j. Um subconjunto de vetor correspondente às amostras de treinamento que estão mais próximas ao hiperplano considerado é chamado de vetor de suporte. A distância entre o vetor de suporte e o hiperplano considerado é chamado de margem e é igual a $1/\parallel \omega \parallel$, em que $\parallel \omega \parallel$ é referente a norma do vetor ω. O objetivo do SVM é calcular o hiperplano ótimo separando as duas classes ou, de forma equivalente, maximizar a margem encontrando o $\min(\parallel \omega \parallel)$. Essa otimização é realizada através do formalismo Lagrangiano (LARDEUX *et al.*, 2009)

$$f(V_j) = Sgn\left(\sum_{i=1}^{k} Y_i.\alpha_i.(\omega.V_j) + b\right) \tag{4.4}$$

onde α_i denota o multiplicador de Lagrange.

4.4.2 SVM - Caso não linear

Quando nenhum hiperplano pode ser encontrado no espaço de entrada E entre as duas classes, os vetores de treinamento são projetados em um espaço H de dimensão M ($M > k$), resultando com o retorno para um problema linear. A projeção é realizada através de uma função de projeção Φ ($\Phi : \mathcal{R}^k \mapsto H$).

É possível definir uma medida de similaridade a partir do produto escalar em H,

$$K(x, x^\top) := \langle x, x^\top \rangle = \langle \Phi(x), \Phi(x^\top) \rangle, \tag{4.5}$$

esta substituição, que é referido por truque do kernel (*kernel trick*), foi utilizado por Boser *et al.* (1992) para estender o classificador hiperplano *Generalized Portrait* para o SVM não linear. O truque do kernel pode ser aplicado, pois todos os vetores de características ocorreram em produtos escalares. Portanto, $\langle \Phi(V_j), \Phi(V_i) \rangle$ é chamado função Kernel K

$$K(V_j, V_i) = \langle \Phi(V_j), \Phi(V_i) \rangle = (\Phi(V_j) \cdot \Phi(V_i)). \tag{4.6}$$

A função de classificação (4.4) para este caso é definida por

$$f(V_j) = Sgn\left(\sum_{i=1}^{k} Y_i.\alpha_i.(\Phi(V_j) \cdot \Phi(V_i)) + b\right), \tag{4.7}$$

em que, através de (4.6) e (4.7), é possível obter

$$f(V_j) = Sgn\left(\sum_{i=1}^{k} Y_i.\alpha_i.K(V_j, V_i) + b\right). \tag{4.8}$$

Os três tipos de *kernels* a seguir são geralmente utilizados (KARATZOGLOU *et al.*, 2006):

1) o *kernel* polinomial:

$$K(V, V_i) = ((V \cdot V_i) + 1)^p;$$

2) o *kernel* tangente hiperbólico:

$$K(V, V_i) = \tanh(k(V \cdot V_i) - \delta);$$

3) o *kernel* função base radial:

$$K(V, V_i) = \exp\left\{-\frac{\| V - V_i \|^2}{2\sigma^2}\right\}.$$

em que $\| \cdot \|$ denota a norma L_2 ou distância euclidiana.

O princípio da SVM foi desenvolvido para um problema de duas classes, mas pode ser facilmente estendido a um problema multiclasse (HSU; LIN, 2002). Mais detalhes sobre SVM podem ser encontrados em Burges (1998).

4.5 Classificadores baseados em divergências

Classificadores baseados em modelos de probabilidade para dados *PolSAR* são utilizados frequentemente (GAO *et al.*, 2014; MORAES; MACHADO, 2014; CINTRA *et al.*, 2013). Classificadores baseados em divergências podem ser entendidos como uma alternativa nesta área porque são fundamentados em um suposto probabilístico. Nesta seção, objetiva-se desenvolver classificadores mais flexíveis baseado em distâncias estocásticas. Silva *et al.* (2013) desenvolveram e avaliaram classificadores para segmentos *PolSAR* por distâncias estocásticas e testes de hipóteses. Um desses classificadores objetivou classificar imagens *SAR* de intensidades multivariadas e imagens de sensores ópticos quando seus dados associados seguem o modelo normal multivariado. Nesta dissertação, os métodos propostos por Silva *et al.* (2013) foram estendidos. Para este fim, utilizou-se a seguinte estratégia para construção de classificadores: Novas propostas

(a) em termos das divergências de Kullback-Leibler e Jeffreys para modelos não normal multivariados,

(b) baseados nas divergências e distâncias de Rényi para distribuição normal multivariada e

(c) com respeito as divergências e distâncias de Rényi para distribuição Wishart complexa escalonada.

Para os itens (b) e (c), consideram-se classificadores ótimos com respeito as ordens das divergências de Rényi. No que se segue, descrevem-se os classificadores propostos. Eles serão comparados com os classificadores clássicos *LDA*, *QDA*, *KNN* e *SVM* bem como outros baseados nas divergências de Kullback-Leible e Jeffreys entre modelos normal multivariados (SILVA *et al.*, 2013) que são casos especiais de nossas propostas.

Sejam $\widehat{\boldsymbol{\theta}}_k$ estimativas de máxima verossimilhança de $\boldsymbol{\theta}_k$,

$$\widehat{\boldsymbol{\theta}}_k = \begin{cases} \left[\widehat{\boldsymbol{\mu}}_k^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_k)^\top\right]^\top & \text{- } \textit{distribuição normal multivariada,} \\ \left[\widehat{\boldsymbol{\mu}}_k^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_k)^\top, \widehat{\boldsymbol{\eta}}_k^\top\right]^\top & \text{- } \textit{distribuição skew-normal multivariada,} \\ \left[\widehat{\boldsymbol{\mu}}_k^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_k)^\top, \widehat{\nu}_k\right]^\top & \text{- } \textit{distribuição t-Student multivariada,} \\ \left[\widehat{\boldsymbol{\mu}}_k^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_k)^\top, \widehat{\boldsymbol{\eta}}_k^\top, \widehat{\nu}_k\right]^\top & \text{- } \textit{distribuição skew-t multivariada,} \\ \left[\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_k)^\top, L\right]^\top & \text{- } \textit{distribuição WCE com número de looks } L, \end{cases}$$

em que $\mathrm{vec}(\cdot)$ é o operador de vetorização.

Considere que se deseja testar se um retorno *PolSAR* pertence a uma entre r classes. Trabalhos sobre a classificação de características *PolSAR* avaliam esta questão sob duas perspectivas: (a) extrair *amostras de teste*, sobre as quais se conhece a física de formação da imagem (isto é, sua textura), a fim de quantificar a precisão global e comparar o desempenho de diferentes classificadores e (b) testar cada entrada de uma imagem *PolSAR* para traçar um mapa de classificação que é usado como medida de comparação qualitativa. Para este fim, dado um retorno *PolSAR* na entrada (h_1, h_2), diga-se $\boldsymbol{Z}_h(h_1, h_2)$, defina a sua vizinhança para um janelamento $k \times k$ com $k = 2n + 1$, $n \in \mathbb{Z}_+$ por

$$\left\{ \boldsymbol{Z}_h(h_1 - k_1, h_2 - k_2), \quad \textit{para } k_1, k_2 = \frac{-(k-1)}{2}, \ldots, 0, \ldots, \frac{(k-1)}{2} \right\}$$

e, sobre ela, obtem-se EMVs, $\widehat{\boldsymbol{\theta}}_h$. O classificador proposto é dado por

$$\widehat{i}_{\mathcal{M}} = \underset{\mathrm{i}\in\{1,\ldots,\mathrm{r}\}}{\arg\min} \left\{ d_h(i) \right\} \quad \text{e}$$

$$d_h(i) = \begin{cases} D_{\mathrm{KL}}^{N}\left([\widehat{\boldsymbol{\mu}}_h^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_h)^\top] || [\widehat{\boldsymbol{\mu}}_i^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_i)^\top]\right), D_{\mathrm{J}}^{N}(\cdot,\cdot), & \text{para CKLN e CJN,} \\ D_{\mathrm{KL}}^{SN}\left([\widehat{\boldsymbol{\mu}}_h^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_h)^\top, \widehat{\boldsymbol{\eta}}_h^\top] || [\widehat{\boldsymbol{\mu}}_i^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_i)^\top, \widehat{\boldsymbol{\eta}}_i^\top]\right), D_{\mathrm{J}}^{SN}(\cdot,\cdot), & \text{para CKLSN e CJSN,} \\ D_{\mathrm{KL}}^{t}\left([\widehat{\boldsymbol{\mu}}_h^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_h)^\top, \widehat{\nu}_h] || [\widehat{\boldsymbol{\mu}}_i^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_i)^\top, \widehat{\nu}_i]\right), D_{\mathrm{J}}^{t}(\cdot,\cdot), & \text{para CKLt e CJt,} \\ D_{\mathrm{KL}}^{St}\left([\widehat{\boldsymbol{\mu}}_h^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_h)^\top, \widehat{\boldsymbol{\eta}}_h^\top, \widehat{\nu}_h] || [\widehat{\boldsymbol{\mu}}_i^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_i)^\top, \widehat{\boldsymbol{\eta}}_i^\top, \widehat{\nu}_i]\right), D_{\mathrm{J}}^{St}(\cdot,\cdot), & \text{para CKLSt e CJSt,} \\ D_{\mathrm{R}}^{\alpha}\left([\widehat{\boldsymbol{\mu}}_h^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_h)^\top] || [\widehat{\boldsymbol{\mu}}_i^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_i)^\top]\right), D_{\mathrm{R2}}^{\alpha}\left([\widehat{\boldsymbol{\mu}}_i^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_i)^\top] || [\widehat{\boldsymbol{\mu}}_h^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_h)^\top]\right), & \text{para CRN1:}\alpha \text{ e CRN2:}\alpha \\ d_{\mathrm{R}}^{\alpha}\left([\widehat{\boldsymbol{\mu}}_h^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_h)^\top] || [\widehat{\boldsymbol{\mu}}_i^\top, \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_i)^\top]\right), d_{\mathrm{R2}}^{\alpha}(\cdot,\cdot), & \text{para CRND1:}\alpha \text{ e CRND2:}\alpha, \\ D_{\mathrm{R}_{\mathbb{C}}}^{\alpha}\left([\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_h)^\top, L] || [\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_i)^\top, L]\right), D_{\mathrm{R}_{\mathbb{C}}}^{\alpha}\left([\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_i)^\top, L] || [\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_h)^\top, L]\right), & \text{para CRWC1:}\alpha \text{ e CRWC2:}\alpha \\ d_{\mathrm{R}_{\mathbb{C}}}^{\alpha}\left([\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_h)^\top, L] || [\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_i)^\top, L]\right), d_{\mathrm{R2}_{\mathbb{C}}}^{\alpha}(\cdot,\cdot), & \text{para CRWCD1:}\alpha \text{ e CRWCD2:}\alpha, \end{cases}$$

$$\mathcal{M} \in \{CKLN, CJN, CKLSN, CJSN, CKLt, CJt, CKLSt, CJSt, CRN1:\alpha, CRN2:\alpha$$
$$CRND1:\alpha, CRND2:\alpha, CRWC1:\alpha, CRWC2:\alpha, CRWCD1:\alpha, CRWCD2:\alpha\}.$$

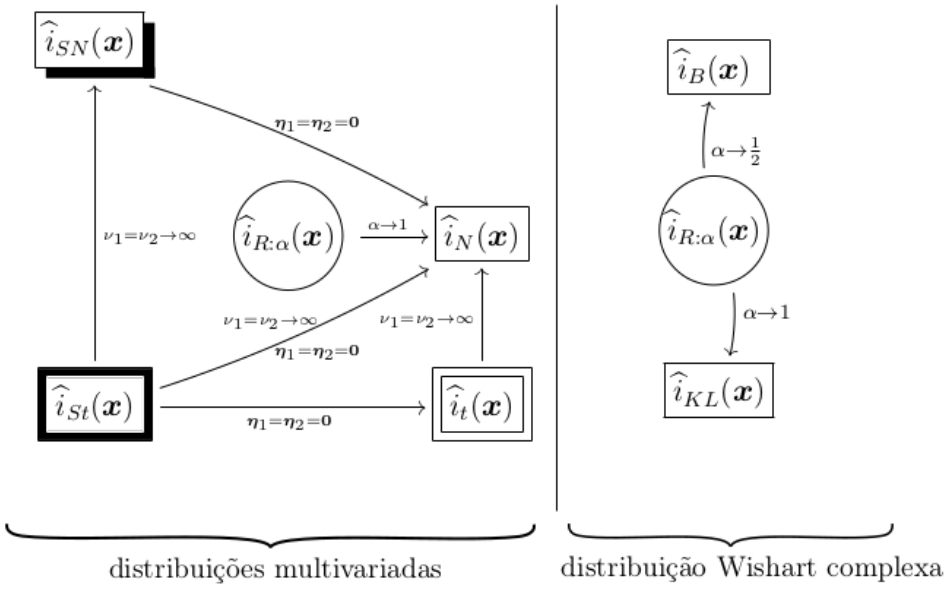


Diagrama 4.5: Relações entre os classificadores propostos.

O Diagrama 4.5 mostra relações entre os classificadores utilizados. Pode-se perceber na parte relacionada às distribuições vetoriais multivariadas que todos classificadores se degeneram sob algumas condições no classificador com base na normal. O classificador KL baseado na distribuição skew-normal, representa o classificador KL com base na normal quando os vetores de assimetria são iguais a zero. O classificador de Rényi na normal tende ao classificador KL com base na normal quando o parâmetro α tende a 1. O classificador induzido da KL para skew-t é o mais flexível, pois é possível obter três classificadores a partir dele: (i) Quando os graus de liberdade são suficientementes grande, o classificador caminha para o da distribuição skew-normal. (ii) Quando os vetores de assimetria são iguais a zero, ela se torna igual ao classificador da t-Student. (iii) Quando as duas coisas acontecem simultaneamente o classificador fica igual ao KL da distribuição normal. Nos casos matriciais do classificador de Rényi com base na WCE, se aproxima do classificador de KL para a mesma distribuição quando o α vai para 1. Quando o α vai para $\frac{1}{2}$, o classificador CRWC1:α tende para o classificador de Bhattacharyya (FRERY *et al.*, 2014). O processo de classificação é detalhado no Algoritmo 1.

Algoritmo 1 Processo de Classificação

1: Obtenha $\boldsymbol{X}_r = [\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}]^\top$ uma amostra aleatória de treinamento para as r classes;
2: Derivar $\boldsymbol{\mu}_r$, $\boldsymbol{\Sigma}_r$ (ν_r e $\boldsymbol{\eta}_r$, quando necessário), utilizando $\boldsymbol{X}$ a amostra de treinamento;
3: Seja M a quantidade de elementos mais próximos de $\boldsymbol{x^*}$ (*pixel* a ser classificado), calcular as estimativas $\boldsymbol{\mu}_*$, $\boldsymbol{\Sigma}_*$ (ν_* e $\boldsymbol{\eta}_*$, quando necessário) para $\boldsymbol{x^*}$, através dele e de seus vizinhos;
4: Calcular $d_h(i)$, sendo $h = 1, \ldots, r$, em que $d_h(i)$ representa uma das medidas de distância exibidas na seção 4.5;
5: O *pixel* $\boldsymbol{x^*}$ é classificado para a classe em que d_h é mínima.

CAPÍTULO 5

Resultados Numéricos

5.1 Análise com dados sintéticos

Neste capítulo, apresentam-se estudos de *performance* dos classificadores sobre dados sintéticos e reais. Para o último caso, consideram-se duas imagens obtidas pelos sensores AIRSAR e EMISAR. Seguem breves discussões sobre eles.

- **AIRSAR (Radar de Abertura Sintética Airborne)**: Este sistema foi projetado e construído pelo Laboratório de propulsão (*Jet Propulsion Laboratory*- NASA-JPL). Sua imagem é apresentada na Figura 5.1 (Fonte: https://adayinthelifeofageek.files.wordpress.com/2012/05/airsar.gif). AIRSAR é uma ferramenta de geoprocessamento para todos os estados atmosféricos capaz de penetrar através das nuvens e recolher dados durante a noite. O AIRSAR serviu como um teste da NASA para demonstrar a nova tecnologia de radar e aquisição de dados em várias aplicações. Como parte da *NASA Earth Science Enterprise* o AIRSAR voou pela primeira vez em 1988 e sua última missão foi em 2004;

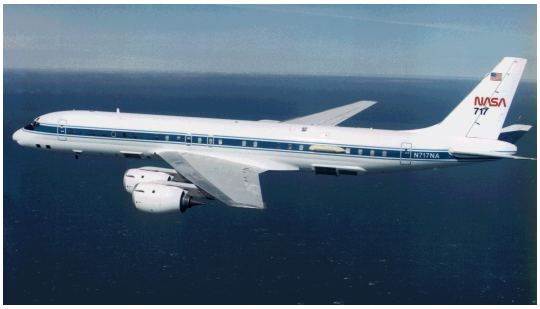


Figura 5.1: Imagem do sistema AIRSAR.

- **EMISAR (*Electromagnetics Institute SAR*)**: Desenvolvido pelo Instituto de Eletromagnética da Universidade Técnica da Dinamarca (Lyngby), o EMISAR é um radar

com dupla frequência (L- e C-bandas) totalmente polarimetrico (4 canais complexos por frequência) para sistema *SAR*. Criado para aplicações de sensoriamento remoto, tal sistema é recentemente utilizado para coletar dados *SAR* de alta qualidade para programas científicos internacionais e para pesquisas em sensoriamento remoto no `Centro Dinamarquês para Sensoriamento Remoto`. O sistema é apresentado na Figura 5.2 (Fonte: https://sites.google.com/site/bancodedadospolsar/_/rsrc/1468876762375/missoes/EMISAR.png).

Figura 5.2: Imagem do sistema EMISAR.

Para o primeiro estudo, consideram-se imagens sintéticas geradas da distribuição $\mathcal{G}_p^0$ discutidas no Capítulo 2. A escolha por este modelo se deve ao mesmo levar em conta tanto o efeito do ruído *speckle* multidimensional (na forma dos parâmetros L para o número de *looks* e $\boldsymbol{\Sigma}$ como a matriz de covariâncias dos canais complexos) como o nível de rugosidade do relevo (na forma do parâmetro β ilustrado na Figura 2.3). Quanto ao parâmetro β, vale-se que áreas mais homogêneas (cenas de oceanos, rios, pastagens) são representadas por menores valores de β. Neste estudo de simulação, adotam-se

(i) $\beta = -1.5$ para cenários extremamente heterogêneos;

(ii) $\beta = -6.0$ para áreas heterogêneas;

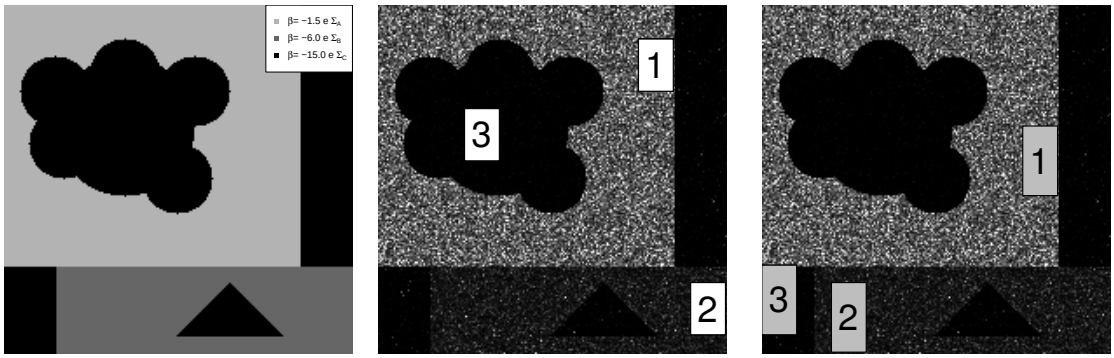
(iii) $\beta = -15.0$ para regiões homogêneas.

Para a definição de $\boldsymbol{\Sigma}$, são utilizadas três matrizes uma para cada textura, e obtidas a partir da imagem real de São Francisco. $L \in \{3$ (menor valor em polarimetria), 4 (*looks* da imagem de São Francisco), 8 (*looks* da imagem Foulum)$\}$. A Figura 5.3 representa a imagem sintética gerada para análise. Esta imagem contém três áreas associadas à pares (β, $\boldsymbol{\Sigma}$) diferentes. Na Figura 5.3(a), podem-se observar as formas geradas para a imagem sintética com as seguintes configurações de parâmetros:

1. $\beta = -1.5$ e $\boldsymbol{\Sigma}_A = \begin{bmatrix} 0.3848 & 0.118 + 0.008i & -0.098 - 0.009i \\ & 0.0770 & -0.050 + 0.015i \\ & & 0.3028 \end{bmatrix} \left(|\boldsymbol{\Sigma}_A| = 4.1 \times 10^{-3},\ \mathrm{tr}(\boldsymbol{\Sigma}_A) = 0.765\right);$

2. $\beta = -6.0$ e $\boldsymbol{\Sigma}_B = \begin{bmatrix} 0.0988 & 0.002 - 0.008i & -0.008 + 0.020i \\ & 0.0439 & 0.001 + 0.002i \\ & & 0.0957 \end{bmatrix} \left(|\boldsymbol{\Sigma}_B| = 3.8 \times 10^{-4},\ \mathrm{tr}(\boldsymbol{\Sigma}_B) = 0.238\right);$

3. $\beta = -15.0$ e $\boldsymbol{\Sigma}_C = \begin{bmatrix} 0.0084 & 0.001 - 0.001i & 0.011 + 0.002i \\ & 0.001 & -0.000 + 0.002i \\ & & 0.0247 \end{bmatrix} \left(|\boldsymbol{\Sigma}_C| = 5.22 \times 10^{-8},\ \mathrm{tr}(\boldsymbol{\Sigma}_C) = 0.034\right).$



(a) áreas da imagem sintética. (b) amostras de treinamento. (c) amostras de teste.

Figura 5.3: Imagem sintética.

Os parâmetros utilizados na área 1 são selecionados tal que a área gerada tenha um comportamento heterogêneo, que implica na área mais clara dentre as geradas. Levando em consideração a mesma ideia, as outras áreas são geradas. Assim, tem-se que a região 2 possui uma heterogeneidade inferior a área 1 e maior com relação a região 3, que é a mais homogênea.

As amostras de treinamento selecionadas para estimação dos parâmetros são destacadas na Figura 5.3(b) e as áreas de amostras de teste utilizadas para verificar a eficiência dos classificadores são exibidas na Figura 5.3(c). Para cada resultado da simulação, 100 réplicas via Monte Carlo são usadas. Além disso, os resultados são obtidos variando o número de *looks* em 3, 4 e 8 e o janelamento em 3×3, 5×5 e 7×7. Como critérios de comparação, utilizou-se o `percentual de acerto` definido como segue.

O cálculo do acerto percentual é dependente da matriz de classificação. A matriz de classificação é um quadro-resumo do fim da execução do método. Sua estrutura pode ser observada na Tabela 5.1, em que as linhas correspondem as classes reais e as colunas correspondem as classes estimadas pelo método. Quando o método classifica corretamente a classe os resultados são armazenados na diagonal principal da matriz e os demais termos correspondem à classificação errada do método.

Tabela 5.1: Matriz de Classificação.

		classes obtidas ($\hat{c}$)			
		1	2	3	4
classes reais (c)	1	n_{11}	n_{12}	n_{13}	n_{14}
	2	n_{21}	n_{22}	n_{23}	n_{24}
	3	n_{31}	n_{32}	n_{33}	n_{34}
	4	n_{41}	n_{42}	n_{43}	n_{44}

A medida denominada de `acerto percentual` é definida como (LANDIS; KOCH, 1977):

$$AC = \frac{\sum_{i=1}^{M} n_{ii}}{\sum_{i=1}^{M}\sum_{j=1}^{M} n_{ij}} \times 100. \tag{5.1}$$

em que $\sum_{i=1}^{M} n_{ii}$ é a soma da diagonal principal, $\sum_{i=1}^{M}\sum_{j=1}^{M} n_{ij}$ é a soma de todos os outros elementos da matriz e o M é o número total de classes.

Outra medida de avaliação é o `coeficiente kappa` que foi proposto Cohen (COHEN, 1960). Esta medida tem se mostrado robusta e leva em conta acertos e erros entre duas fontes de informação, a partir de uma matriz de confusão. O coeficiente kappa é expresso por:

$$K = \frac{(P_0 - P_c)}{(1 - P_c)}, \tag{5.2}$$

em que $P_0 = \frac{\sum_{i=1}^{M} n_{ii}}{N}$, $P_c = \frac{\sum_{i=1}^{M} n_{i+}n_{+i}}{N^2}$, n_{ii} é o total da diagonal principal da matriz de classificação, n_{i+} é o total da linha i na matriz, n_{+i} é o total da coluna i na matriz, M é o número total de classes e N é o número total de possíveis decisões presentes na matriz de confusão.

De acordo com Landis e Koch (1977), o coeficiente kappa pode ser interpretado em termos de percentuais de concordância como dado na Tabela 5.2.

Tabela 5.2: Interpretação do coeficiente kappa.

Estatística Kappa	Grau de Concordância
<0.00	Concordância Pobre
0.00-0.20	Concordância Pequena
0.21-0.40	Concordância Regular
0.42-0.60	Concordância Moderada
0.61-0.80	Concordância Considerável
0.81-1.00	Concordância Quase Perfeita

Antes de discutir os resultados de cada classificador, necessita-se comentar sobre duas etapas na estrutura da classificação:

(i) otimalidade na divergência de Rényi;

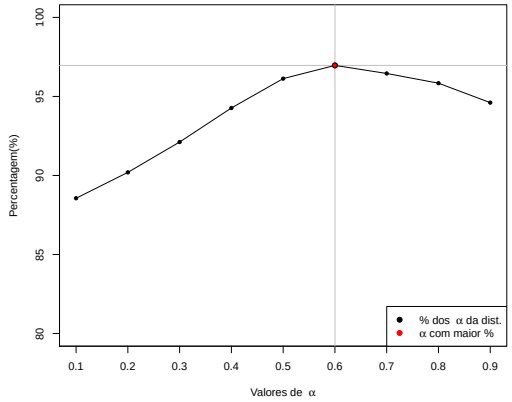
(ii) transformação Box-Cox para intensidades SAR multivariadas.

Quanto ao item (i), note que os classificadores propostos CRN1:α, CRN2:α, CRND1:α, CRND2:α, CRWC1:α, CRWC2:α, CRWCD1:α e CRWCD2:α dependem do parâmetro α adicionalmente aos outros parâmetros induzidos do suposto probabilístico adotado. Para definição deste parâmetro e dada as amostras de treinamento de partições de um problema sob estudo, utilizou-se um processo de otimização discutido como segue:

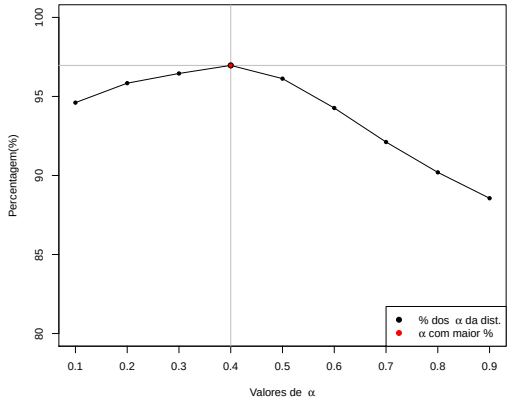
Como um estimador para α, utilizou-se a seguinte expressão:

$$\hat{\alpha} = \underset{\alpha \in \mathbb{R}_{(0,1)}}{\arg\min} \left[\sum_{i=1}^{g} \sum_{k_i=1}^{n_i} \mathcal{D}_{\mathrm{R}}^{\alpha}([\boldsymbol{\mu}_{k_i}, \boldsymbol{\Sigma}_{k_i}], [\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i]) \right], \tag{5.3}$$

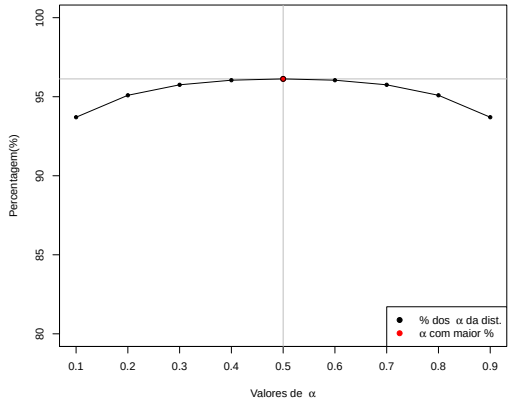
em que g é o número de grupos, n_i é o número de elementos (ou *pixels*) no i-ésimo grupo, $(\widehat{\boldsymbol{\mu}_i}, \widehat{\boldsymbol{\Sigma}}_i)$ são os EMVs para $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ do i-ésimo grupo com base na amostra de treinamento, $(\widehat{\boldsymbol{\mu}}_{k_i}, \widehat{\boldsymbol{\Sigma}}_{k_i})$ são os EMVs para $[\boldsymbol{\mu}_{k_i}, \boldsymbol{\Sigma}_{k_i}]$ com base na vizinhança do k_i-ésimo elemento do grupo i e $\mathcal{D}_{\mathrm{R}}^{\alpha}(\cdot,\cdot)$ representa as divergências discutidas na Seção 3.2.5. A Figura 5.4 apresenta as *performances* de 36 classificadores com base em $\mathcal{D}_{\mathrm{R}}^{\alpha}$ para $\alpha \in \{0.1, 0.2, \ldots, 0.9\}$. Pode-se notar que em todos os casos existe pelo menos um α que otimiza o desempenho dos classificadores. Nesta dissertação, propõe-se um estudo sobre 5.3 em comparação com o valor do α ótimo observado empiricamente. Os resultados podem ser visualizados no Apêndice C. Pode-se perceber que a inclusão de um processo de otimização em classificadores tipo Rényi é interessante. Embora este estudo tenha sido feito, optamos por determinar um valor ótimo dado um "grid" de valores. Em estudos futuros, adicionaremos a determinação do α iterativamente.



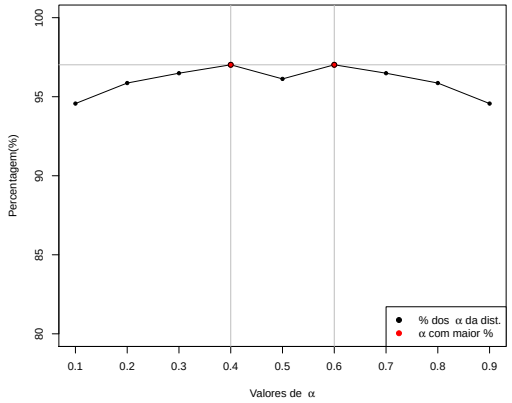
(a) classificações para $D_R^{\alpha}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2)$.



(b) classificações para $D_R^{\alpha}(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1)$.



(c) classificações para $d_R^{\alpha}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2)$.



(d) classificações para $d_{R2}^{\alpha}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2)$.

Figura 5.4: Classificação com base nas divergências/distâncias de Rényi para escolha do α ótimo.

Por outro lado quando ao item (ii), os dados também foram classificados utilizando a transformação do Box-Cox. Como descrito no Capítulo 1, "as técnicas de processamento de imagens geralmente impõe a suposição de normalidade, entretanto as diagonais principais dos dados *Pol-*

SAR não são descritos pela distribuição normal multivariada". A fim de obter dados que se aproximem da distribuição normal e satisfazer a suposição dos classificadores foi utilizada a transformação de Box-Cox (SAKIA, 1992):

"Seja $\boldsymbol{y} = (Y_1, Y_2, \ldots, Y_n)^\top$ uma amostra aleatória que segue uma distribuição com fdp $f_{\boldsymbol{y}}(\boldsymbol{y})$, tal transformação é definida por:

$$\boldsymbol{y}^{(\lambda)} = \begin{cases} \dfrac{\boldsymbol{y}^\lambda - 1}{\lambda} & (\lambda \neq 0), \\ \log \boldsymbol{y} & (\lambda = 0), \end{cases}$$

em que λ é estimado através da maximização da função de verossimilhança da distribuição normal pelo Jacobiano da transformação (SAKIA, 1992),$\log \boldsymbol{y} = (\log Y_1, \ldots, \log Y_n)^\top$ e $\dfrac{\boldsymbol{y}^\lambda - 1}{\lambda} = \left(\dfrac{Y_1^\lambda - 1}{\lambda}, \ldots, \dfrac{Y_n^\lambda - 1}{\lambda}\right)$."

As tabelas com todas as classificações encontram-se no Apêndice B. Os parâmetros utilizados no KNN e SVM encontram-se também no Apêndice B nas tabelas B.1 e B.2, respectivamente. Para o número de *looks* igual a 3 e para o janelamento 3×3 na Tabela B.3, os melhores classificadores são baseados nas divergências de Rényi para distribuição Wishart complexa escalonada, CRWC1:α e CRWC2:α. Neste caso, as duas divergências de Rényi apresentaram o melhor desempenho com valor 98,30%. A primeira com ordem 0.1, enquanto a segunda ordem 0.9. Para o janelamento 5×5 e 7×7 as duas divergências anteriores também obtiveram a maior percentagem de acerto. A pior classificação foi obtida pela divergência de Jeffreys com base na distribuição multivariada Skew-t (CJSt), com precisão de 31,16% na janela 7×7. A divergência KL (CKLN) apresentou uma acurácia de 95,19% com janelamento 3×3. O KNN, LDA, QDA e SVM obtiveram precisões de 87,38%, 76,30%, 86,49 % e 90,45%, respectivamente. Ao realizar a transformação Box-Cox nos dados, os classificadores CRN1:α, CRN1:α e CRND2:α obtiveram a melhor classificação com uma percentagem de acerto de 99.35%. É possível observar que houve um ganho de desempenho dos dados transformados em relação à utilização dos dados sem transformação. Os dados matriciais não foram submetidos a transformação.

Os resultados para o número de *looks* igual a 4 são apresentadas na Tabela B.4. A melhor classificação para o janelamento 3×3 foi obtida pelo classificador CRWC1:α com $\alpha = 0.1$ e acurácia de 98.30%. A divergência simétrica (CRWC2:α) obteve a mesma percentagem de acerto com α de 0.9. O pior classificador foi CKLSt com o janelamento de 5×5, que obteve uma percentagem de classificação de 28.13%. Com exceção do KNN, LDA, QDA e também CKLt, CJt, CKLSt e CJSt (com o janelamento 7×7), treze classificadores obtiveram acurácia superior a 90% de acerto.

Com o aumento do número de *looks*, as áreas tendem a se tornar mais homogêneas. As melhores percentagens de classificação foram obtidas no janelamento 3×3 com o CRWC1:α e CRWC2:α para os $\alpha = (0, 1, 0, 9)$ obtendo um acerto percentual de 98,26%.

As divergências e distâncias de Rényi baseadas na distribuição normal foram melhores que as divergências e distâncias de Rényi baseados na distribuição Wishart para o janelamento 5×5 e 7×7. O classificador baseado na medida de Rényi obteve uma *performance* melhor do que os devidos as medidas de Kullback-Leibler e Jeffreys. Os piores classificadores foram os de

Kullback-Leiler e Jeffreys baseados nas distribuições t-Student multivariada (CKLt e CJt) e Skew-t multivariada (CKLSt e CJSt), com o classificador CKLSt com uma percentagem de 31,54% como o pior de todos.

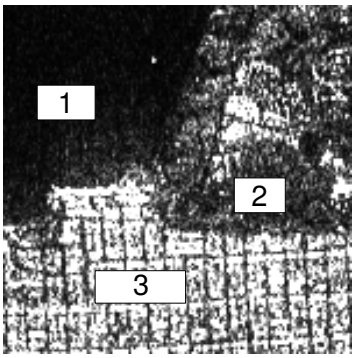
Quando os dados são transformados por Box-Cox, os melhores classificadores continuam sendo os da Rényi, baseados na distribuição normal, obtendo a melhor clasificação no janelamento 7×7 com percentagem de acerto de 99,69%. A pior classificação neste caso foi com o classificador CJSt e janelamento 7×7 com percentagem de 63,96%.

A Figura B.1(a) apresenta um mapa qualitativo de comparações entre a estrutura física da simulação e os resultados dos classificadores. As melhores classificações se encontram na primeira linha. A Figura B.1(b) representa a melhor classificação para o número de *looks* 3, a Figura B.1(c) para o número de *looks* 4 e a Figura B.1(d) para o número de *looks* igual a 8. A segunda linha representa as piores classificações de acordo com as Tabelas B.3, B.4 e B.5 para os números de *looks* iguais 3, 4 e 8, respectivamente. As Figuras da divergência de KL baseadas na distribuição normal foram colocados na quarta linha e nas linhas subsequente são apresentadas as classificações baseados no LDA, QDA e KNN, respectivamente.

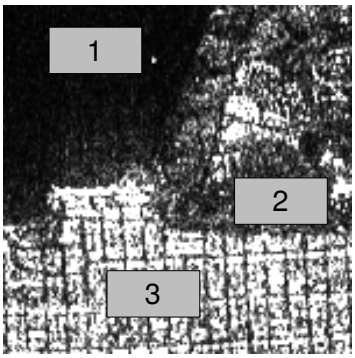
5.2 Estudo com dados reais

Os métodos propostos são usados para classificação em imagens *PolSAR*. O primeiro experimento foi feito em uma região da baía de São Francisco (*San Francisco - EUA*). Essa imagem foi capturada pelo sensor *AIRSAR* com o número de *looks* igual a quatro. Além disso, o segundo experimento foi realizado em uma região de *Foullum (DK)*. A imagem resultante foi obtida pelo sensor *EMISAR* com número de *looks* igual a oito.

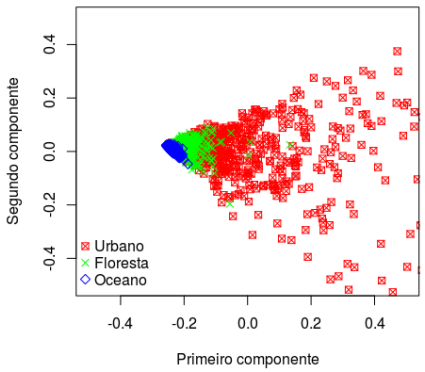
A Figura 5.5 mostra a imagem *AIRSAR* com as áreas destacadas das amostras de treinamento e testes. Neste caso, selecionamos três regiões para as etapas de treinamento e teste. Eles são escolhidos de tal forma que cada um representa apenas uma textura; isto é, mar (caso menos texturizado), floresta (meio textura) e urbano (extremamente texturizado) cenários. Por outro lado, a Figura 5.6 apresenta a imagem obtida pelo EMISAR, que possui seis regiões bem definidas (1- Trigo, 2- Colza, 3- Centeio, 4- Conífera, 5- Aveia, 6- *background*). As amostras de treino e de teste são destacadas.



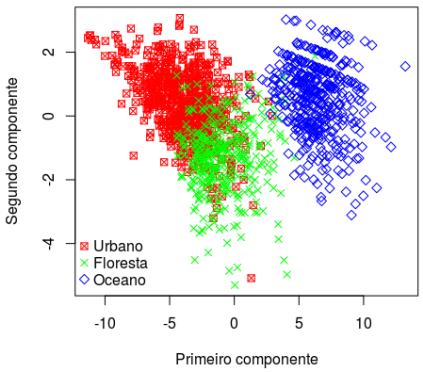
(a) Amostras de treinamento.



(b) Amostras de teste.



(c) Análise de componentes principais.

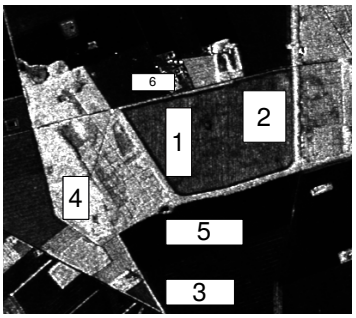


(d) Análise de componentes principais para os dados transformados.

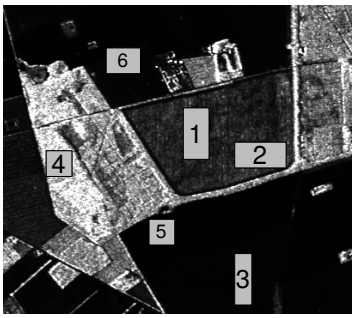
Figura 5.5: Imagem de São Francisco.

Antes de discutir sobre os resultados da classificação, foi realizado uma breve análise descritiva multivariada dos dados obtidos a partir das duas imagens reais. Para tanto, considerou-se uma redução da dimensionalidade, representando as imagens consideradas por suas duas principais componentes da amostra. As Figuras 5.5(c)- 5.5(d) e 5.6(c) - 5.6(d) mostram gráficos de dispersão de tais componentes. Para a imagem de São Francisco observe que os dados das três regiões estão sobrepostos, que impõem cenários duros de classificação. Ao aplicar a transformação de Box-Cox nos dados, pode-se notar que as regiões de oceano se tornam mais distintas das regiões de florestas e urbanas, mas as duas últimas apresentam algum grau de sobreposição. Para a imagem de Foulum, embora se possa segmentar algumas texturas distintas (tais como regiões de *background* e coníferas), existe sobreposição expressiva entre os dados originais e nos dados transformados. Além disso, esta última imagem envolve mais segmentos a serem classificados e, portanto, sua classificação parece impor mais dificuldade em relação a primeira imagem.

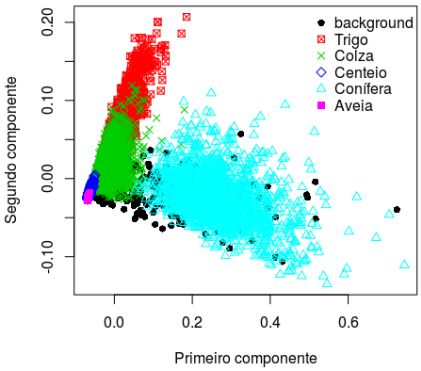
Resultados da classificação de imagens de São Francisco podem ser verificados na Tabela B.6. Conforme discutido nas Figuras 5.5(c), 5.5(d), 5.6(c) e 5.6(d), podem-se ver que as taxas de classificação melhoraram para dados transformados. Este fato, pode ser explicado pelo fato de todas as distribuições multivariadas (com exceção dos modelos WCE) consideradas para obtenção das distâncias utilizadas apresentam suporte em todo espaço euclidiano e multidimensional.



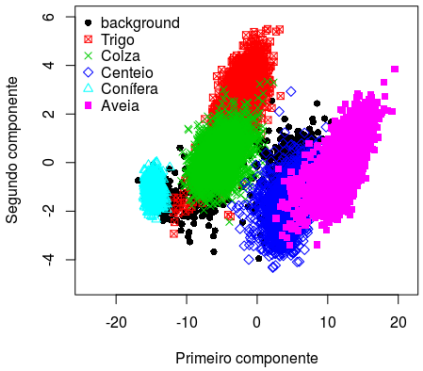
(a) amostras de treinamento.



(b) amostras de teste.



(c) Análise de componentes principais.



(d) Análise de componentes principais para os dados transformados.

Figura 5.6: Imagem de Foulum.

O melhor classificador para a imagem de São Francisco utilizando os dados reais foi obtido com a percentagem de acerto de 99.69% pela divergência de Rényi (CRN1:α e CRN2:α) com α de 0.9 para CRN1:α e 0.1 para CRN2:α, os dois com janelamento de 7×7. Sua imagem correspondente pode ser encontrada na Figura B.2(d) e comparado com a Figura B.2(b) que é a imagem *ground-truth* da área de São Francisco. Com os dados transformados, o melhor classificador foi com a percentagem de 99,73% e obtido através de três classificadores propostos, o CRN1:α com parâmetro 0.9, CRN2:α com parâmetro 0.1 e CRND2:α com parâmetros α podendo ser 0.1 ou 0.9, a classificação obtida com esses dois parâmetros foram iguais. A pior classificação para os dados reais foi obtida pelo CJSt, com percentagem de 38,06% e janelamento de 3×3, e para os dados transformados, a pior classificação foi com o mesmo classificador porém no janelamento de 5×5.

Para a imagem de Foulum, a melhor classificação foi apresentada pelo CKLt com o janelamento 5×5 e com a acurácia de 93,18%. CKLt também apresentou a melhor classificação para os dados transformados e ocorreu no janelamento 7×7 com acerto percentual de 96.77%. As imagens classificadas pelo CKLt nos dois casos, com dados originais e transformados, são exibidas nas Figuras B.3(b) e B.3(c), respectivamente. As piores classificações para dados não transformados foram obtidas com CKLSt com janelamento 3×3 e com acerto percentual de 27,58%. Com respeito aos dados transformados, o pior classificador foi CJt com janelamento de

7×7 e acurácia de 68,19%. As imagens relacionadas são encontradas na terceira linha da Figura B.3.

A divergência KL com base na distribuição Normal (CKLN) para os dados reais obteve *performances* inferiores a 90% nos três janelamentos, quando os dados foram transformados, a *performance* do classificador foi 94,90% no janelamento 7×7. As *performances* LDA, QDA, KNN e SVM foram também melhoradas quando os dados foram transformados.

CAPÍTULO 6

Considerações Finais

Nesta dissertação foram propostos doze classificadores baseados em medidas de divergências e distâncias entre dois vetores aleatórios e quatro classificadores para duas matrizes aleatórias hermitianas. Especificamente, foram consideradas as divergências de Kullback-Leibler, Jeffreys e Rényi induzidas dos modelos normal, skew-normal, t-Student, skew-t e wishart complexa escalonada. Essas medidas foram compreendidas por meio de esquema tradicional Bayesiano de classificadores para intensidades *SAR* trivariadas (resultantes de modelos vetoriais) e para dados *full PolSAR* (a partir do modelo de matricial considerado). Em particular, para a divergência de Rényi (que possui um parâmetro de flexibilidade, chamado *ordem*) foi proposto um procedimento de otimização para obter o melhor classificador de acordo com um critério de acurácia.

As propostas foram avaliadas tanto em dados sintéticos como em dados reais. Para o primeiro caso, realizaram-se experimentos Monte Carlo nos quais os dados gerados seguiram o modelo $\mathcal{G}_p^0$, que emula cenários reais relacionados ao terreno e ruído *speckle* multidimensional. Os resultados têm mostrado evidências em favor da utilização da divergência de Rényi otimizada entre modelos wishart complexo escalonada. As *performances* também foram verificadas para dois experimentos com dados reais que, em geral, para os classificadores considerados para dados *SAR* de intensidades vetoriais, a utilização da transformação Box-Cox melhoraram as taxas de classificação em todos os casos. Pode-se destacar que considerando as imagens de São Francisco e Foulum, a transformação nos dados resultou em alguns casos uma melhora de 100% das taxas de acerto dos piores classificadores, entre eles, CJSt e CKLSt.

O primeiro experimento foi com a imagem de São Francisco, que é formada por três regiões oceano, floresta e urbano, as quais são áreas consideradas homogêneas, heterogênea e extremamente heterogênea, respectivamente. Através da análise dos componentes principais foi possível identificar que as áreas encontravam-se sobrepostas e ao realizar a transformação Box-Cox as regiões tornaram-se mais distintas. Para os dois tipos de dados utilizados, as divergências otimizada de Rényi com base na distribuição normal obtiveram os melhores desempenhos, sendo os

classificadores CRN1:α e CRN2:α, com janelamento de 7×7 superior aos demais classificadores.

O segundo experimento desta dissertação foi realizado com a área da imagem de Foulum, que é composta por seis áreas distintas e é considerada uma imagem que impõe maior dificuldade para classificação em relação a imagem de São Francisco, pois existe uma sobreposição expressiva entre as regiões. Para este caso o classificador com base na divergência de Kullback-Leibler para distribuição t-Student (CKLt) foi superior aos demais classificadores, superando os clássicos LDA, QDA, KNN e SVM, e também superando o classificador KL entre distribuições normais.

A distribuição Skew-t foi o modelo mais flexível utilizado, porém os classificadores obtiveram os piores resultados em todos os estudos realizados. A sua melhor classificação foi obtida com a imagem Foulum, com os dados transformados, porém não superou 80%. Outra observação interessante é que tanto nas simulações quanto nas aplicações com dados reais as divergências de Jeffreys (KL simetrizada) sempre obtiveram acurácia pior do que aquelas induzidas das divergências KL com a distribuição relacionada.

Referências

ARELLANO-VALLE, R. B.; GÓMEZ, H. W.; QUINTANA, F. A. A new class of skew-normal distributions. *Communications in Statistics–Theory and Methods*, v. 33, p. 1465–1480, 2004.

AZZALINI, A.; VALLE, A. D. The multivariate skew-normal distribution. *Biometrika*, v. 83, p. 715–726, 1996.

BEAULIEU, J. M.; TOUZI, R. Segmentation of textured polarimetric SAR scenes by likelihood approximation. *IEEE Transactions on Geoscience and Remote Sensing*, v. 42, p. 2063–2072, 2004.

BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. [S.l.]: ACM, 1992. p. 144–152.

BRANCO, M. D.; DEY, D. K. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, v. 79, p. 99 – 113, 2001.

BROYDEN, C. A new double-rank minimization algorithm. *Notices American Mathematical Society*, 1969.

BURGES, C. J. *A Tutorial on Support Vector Machines for Pattern Recognition*. [S.l.: s.n.], 1998. v. 2. 121-167 p.

CINTRA, R. J.; FRERY, A. C.; NASCIMENTO, A. D. C. Parametric and nonparametric tests for speckled imagery. *Pattern Analysis and Applications*, v. 16, p. 141–161, 2013.

COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, v. 20, p. 37–46, 1960.

COMMENGES, D. Information Theory and Statistics: an overview. *ArXiv e-prints*, 2015. Disponível em: <http://adsabs.harvard.edu/abs/2015arXiv151100860C>.

CONTRERAS-REYES, J. E. Asymptotic form of the Kullback–Leibler divergence for multivariate asymmetric heavy-tailed distributions. *Physica A: Statistical Mechanics and its Applications*, v. 395, p. 200–208, 2014.

CONTRERAS-REYES, J. E.; ARELLANO-VALLE, R. B. Kullback-Leibler divergence measure for multivariate skew-normal distributions. *Entropy*, v. 14, p. 1606–1626, 2012.

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

CSISZÁR, I. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, v. 2, p. 299–318, 1967.

DOULGERIS, A.; ANFINSEN, S. N.; ELTOFT, T. Analysis of non-Gaussian POLSAR data. In: IEEE. *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*. [S.l.], 2007. p. 160–163.

FLETCHER, R. A new approach to variable metric algorithms. *The Computer Journal*, 1970.

FREITAS, C. C.; FRERY, A. C.; CORREIA, A. H. The polarimetric G distribution for SAR data analysis. *Environmetrics*, v. 16, p. 13–31, 2004.

FREITAS, C. C.; SANT'ANNA, S. J. S.; RENNO, C. D.; CORREIA, A. H. Utilização de imagens de radar de abertura sintética na classificação do uso e ocupação do solo. *EMBRAPA. Geomática: Técnicas e Aplicações Avançadas*, 2003.

FRERY, A. C.; CORREIA, A. H.; FREITAS, C. d. C. Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information. *IEEE Transactions on Geoscience and Remote Sensing*, v. 45, p. 3098–3109, 2007.

FRERY, A. C.; MÃijLLER, H.-J.; YANASSE, C. da C. F.; SANT'ANNA, S. J. S. A model for extremely heterogeneous clutter. *IEEE Transactions on Geoscience and Remote Sensing*, v. 35, 1997.

FRERY, A. C.; NASCIMENTO, A. D. C.; CINTRA, R. J. Analytic expressions for stochastic distances between relaxed complex Wishart distributions. *IEEE Trasactions on Geoscience and Remote and Sensing*, v. 52, p. 1213–1226, 2014.

GAO, W.; YANG, J.; MA, W. Land cover classification for Polarimetric SAR images based on mixture models. *Remote Sensing*, v. 6, p. 3770, 2014.

GOLDFARB, D. A family of variable metric updates derived by variational means. *Mathematics of Computation*, 1970.

GOODMAN, N. R. Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 34, p. 152–177, 1963.

HAIR, J.; BLACK, W.; BABIN, B.; ANDERSON, R.; TATHAM, R. *Análise multivariada de dados - 6ed:*. [S.l.]: Bookman, 2009.

HOEKMAN, D. H.; VISSERS, M. A. A new polarimetric classification approach evaluated for agricultural crops. *IEEE Transactions on Geoscience and Remote Sensing*, v. 41, p. 2881–2889, 2003.

HOFERT, M. On sampling from the multivariate t distribution. *The R Journal*, v. 5, p. 129–136, 2013.

HSU, C.-W.; LIN, C.-J. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, v. 13, p. 415–425, 2002.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. *An Introduction to Statistical Learning: with Applications in R*. [S.l.]: Springer, 2013. (Springer Texts in Statistics).

JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. [S.l.]: Pearson Education, Inc., 2007.

KARATZOGLOU, A.; MEYER, D.; HORNIK, K. Support vector machines in R. *Journal of Statistical Software*, v. 15, p. 1–28, 2006.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, n. 1, p. 159–174, 1977.

LARDEUX, C.; FRISON, P. L.; TISON, C.; SOUYRIS, J. C.; STOLL, B.; FRUNEAU, B.; RUDANT, J. P. Support vector machine for multifrequency SAR Polarimetric data classification. *IEEE Transactions on Geoscience and Remote Sensing*, v. 47, p. 4143–4152, 2009.

LEE, S.; MCLACHLAN, G. J. On the fitting of mixtures of multivariate skew t-distributions via the EM algorithm. *ArXiv e-prints*, 2011. Provided by the SAO/NASA Astrophysics Data System.

LOVE, A. In memory of Carl A. Wiley. *IEEE Antennas and Propagation Society Newsletter*, v. 27, p. 17–18, 1985.

MAITRE, H. *Processing of Synthetic Aperture Radar Images (Digital Signal and Image Processing Series)*. [S.l.]: ISTE, 2006.

MCLACHLAN, G.; LEE, S. EMMIXuskew: An R package for fitting mixtures of Multivariate Skew t distributions via the EM Algorithm. *Journal of Statistical Software*, v. 55, p. 1–22, 2013.

MISHRA, A. K. Validation of pca and lda for sar atr. In: *TENCON 2008 - 2008 IEEE Region 10 Conference*. [S.l.: s.n.], 2008.

MORAES, R. M.; MACHADO, L. S. Psychomotor skills assessment in medical training based on virtual reality using a weighted possibilistic approach. *Knowledge-Based Systems*, v. 70, p. 97 – 102, 2014.

NASCIMENTO, A. D. C. *Teoria Estatística da Informação para Dados de Radar de Abertura Sintética Univariados e Polarimétricos*. Tese (Doutorado) — Pós-graduação em Estatística, 2012.

NASCIMENTO, A. D. C.; CINTRA, R. J.; FRERY, A. C. Hypothesis testing in speckled data with stochastic distances. *IEEE Transactions on Geoscience and Remote Sensing*, v. 48, p. 373?385, 2010.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2015.

RODRÍGUEZ, W. D. A. *Classificação Estatística nos Espaços Euclideanos e Não-Euclideanos com aplicação em dados de forma*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2015.

SAKIA, R. M. The box-cox transformation technique: A review. *Journal of the Royal Statistical Society. Series D (The Statistician)*, [Royal Statistical Society, Wiley], v. 41, n. 2, p. 169–178, 1992. ISSN 00390526, 14679884.

SALICRÚ, M.; MORALES, D.; MENÉNDEZ, M. L.; PARDO, L. On the applications of divergence type measures in testing statistical hypotheses. *Journal of Multivariate Analysis*, Academic Press, Inc., Orlando, FL, USA, v. 51, p. 372–391, 1994. ISSN 0047-259X.

SCHOU, J.; SKRIVER, H.; NIELSEN, A. A.; CONRADSEN, K. CFAR Edge Detector for Polarimetric SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, v. 41, p. 20–32, 2003.

SHANNO, D. Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation*, 1970.

SHANNON, C. E. A mathematical theory of communication. *Bell System Tech*, v. 27, 1948.

SILVA, W. B.; FREITAS, C. C.; Sant'Anna, S. J. S.; FRERY, A. C. Classification of segments in PolSAR imagery by minimum stochastic distances between Wishart distributions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 6, p. 1263–1273, 2013.

TAO, M.; ZHOU, F.; LIU, Y.; ZHANG, Z. Tensorial independent component analysis-based feature extraction for polarimetric SAR data classification. *IEEE Transactions on Geoscience and Remote Sensing*, v. 53, p. 2481–2495, 2015.

ULABY, F. T.; ELACHI, C. *Radar polarimetry for geoscience applications*. [S.l.]: Norwood, MA : Artech House, 1990.

VAPNIK, V. N. *Statistical learning theory*. New York: Wiley, 1998. (Adaptive and learning systems for signal processing, communications, and control).

APÊNDICE A

Faixas de Comprimento e Frequência de Sistemas de Radar

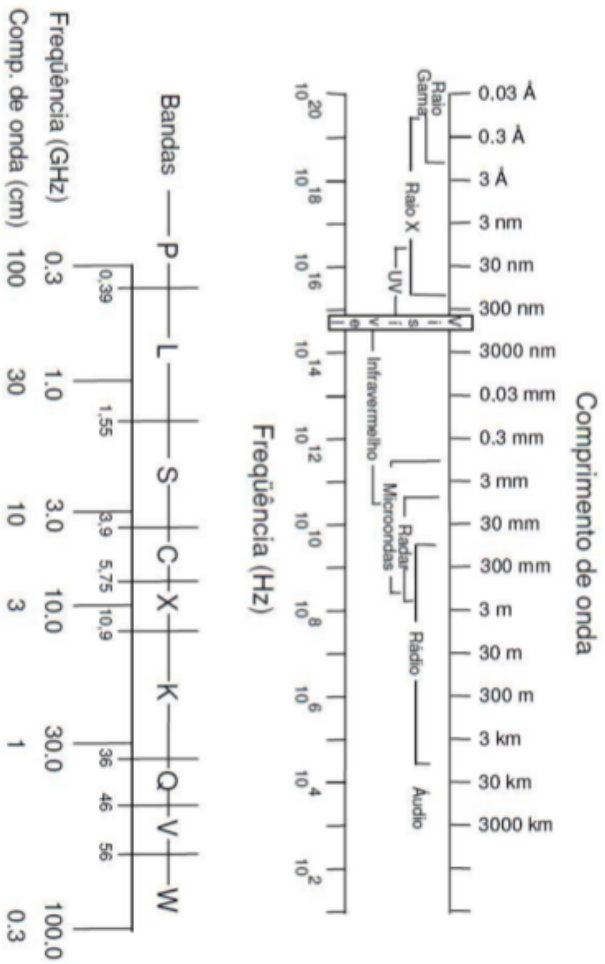


Figura A.1: Faixas do comprimento de onda e da frequência para um espectro eletromagnético (Fonte: Freitas et al. 2003, p. 2).

APÊNDICE B

Resultados das Classificações

Tabela B.1: *Grid* utilizado para KNN.

área	*grid*	máximo
Simulação L3	$1, \cdots, 10$	6
Simulação L3 - BC	$1, \cdots, 10$	5
Simulação L4	$1, \cdots, 10$	8
Simulação L4 - BC	$1, \cdots, 10$	8
Simulação L8	$1, \cdots, 10$	7
Simulação L8 - BC	$1, \cdots, 10$	7
São Francisco	$1, \cdots, 10$	3
São Francisco - BC	$1, \cdots, 10$	2
Foulum	$1, \cdots, 10$	9
Foulum - BC	$1, \cdots, 10$	9

Tabela B.2: Parâmetros utilizados no SVM.

área	*kernel*	$1/2\sigma^2$	'C'-constant
Simulação L3	radial	.1	10
Simulação L3 - BC	radial	.1	10
Simulação L4	radial	.1	10
Simulação L4 - BC	radial	.1	10
Simulação L8	radial	.1	10
Simulação L8 - BC	radial	.1	10
São Francisco	radial	3000	10
São Francisco - BC	radial	10	10
Foulum	radial	3	1000
Foulum - BC	radial	10	1

B.1 Resultados para a imagem sintética

Tabela B.3: Classificação para imagem sintética com número de *looks* 3.

L3						
	janelamento 3×3		janelamento 5×5		janelamento 7×7	
	α	% accu.	α	% accu.	α	% accu.
$D_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.7	96.70	0.6	96.96	0.6	97.08
$D_R^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.3	96.70	0.4	96.96	0.4	97.08
$d_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	95.10	0.5	96.13	0.5	96.43
$d_{R2}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.3,0.7)	96.67	(0.4,0.6)	97.02	(0.4,0.6)	97.20
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.1	98.30	0.1	97.81	0.1	97.73
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.9	98.30	0.9	97.81	0.9	97.73
$d_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	98.11	0.5	97.37	0.5	96.57
$d_{R2_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	98.11	0.5	97.37	0.5	96.57
$\mathrm{D}_{\mathrm{KL}}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	95.19	-	95.01	-	94.20
$\mathrm{J}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	93.75	-	93.77	-	93.11
$\mathrm{D}_{\mathrm{KL}}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	95.19	-	95.01	-	94.20
$\mathrm{J}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	93.75	-	93.77	-	93.11
$\mathrm{D}_{\mathrm{KL}}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	73.95	-	74.85	-	75.27
$\mathrm{J}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	63.99	-	64.78	-	65.08
$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	32.05	-	31.78	-	31.38
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	31.87	-	31.61	-	31.16
KNN	87.38					
LDA	76.30					
QDA	86.49					
SVM	90.45					
L3 - transformação Box-Cox						
$D_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	98.49	0.6	98.95	0.7	99.35
$D_R^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.5	98.49	0.4	98.95	0.3	99.35
$d_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.5,0.6,0.7)	98.46	0.5	98.94	0.5	99.29
$d_{R2}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	98.49	(0.4,0.6)	98.95	(0.3,0.7)	99.35
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	-	-	-	-	-	-
$d_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-
$d_{R2_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-
$\mathrm{D}_{\mathrm{KL}}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	98.25	-	98.26	-	98.80
$\mathrm{J}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	98.05	-	98.40	-	98.87
$\mathrm{D}_{\mathrm{KL}}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	98.25	-	98.26	-	98.80
$\mathrm{J}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	98.05	-	98.40	-	98.87
$\mathrm{D}_{\mathrm{KL}}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	98.24	-	97.66	-	97.60
$\mathrm{J}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	52.54	-	51.11	-	50.79
$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	54.74	-	54.63	-	55.30
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	56.79	-	56.79	-	57.03
KNN	88.67					
LDA	90.31					
QDA	90.70					
SVM	90.30					

Tabela B.4: Classificação para imagem sintética com número de *looks* 4.

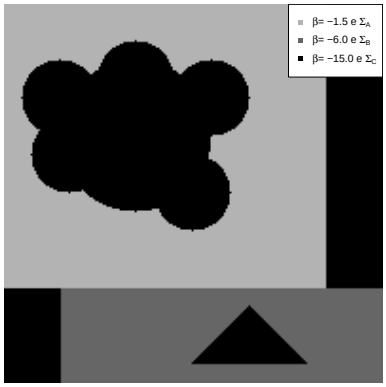
L4						
	janelamento 3×3		janelamento 5×5		janelamento 7×7	
	α	% accu.	α	% accu.	α	% accu.
$D_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.6	97.77	0.5	97.69	0.5	97.87
$D_R^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.4	97.77	0.5	97.69	0.5	97.87
$d_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	97.15	0.5	97.69	0.5	97.87
$d_{R2}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.4,0.6)	97.77	(0.4,0.6)	97.84	0.5	97.87
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.1	98.30	0.1	97.75	0.1	97.67
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.9	98.30	0.9	97.75	0.9	97.67
$d_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	98.12	(0.1,0.9)	97.42	(0.1,0.9)	96.62
$d_{R2_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.2,0.8)	98.13	(0.1,0.9)	97.42	(0.1,0.9)	96.62
$\mathrm{D}_{\mathrm{KL}}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	96.10	-	95.62	-	94.77
$\mathrm{J}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	93.81	-	93.69	-	93.04
$\mathrm{D}_{\mathrm{KL}}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	96.10	-	95.62	-	94.77
$\mathrm{J}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	93.81	-	93.69	-	93.04
$\mathrm{D}_{\mathrm{KL}}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	76.24	-	77.00	-	77.25
$\mathrm{J}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	65.64	-	66.27	-	66.44
$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	28.30	-	28.13	-	28.24
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	28.44	-	28.34	-	28.49
KNN	89.57					
LDA	78.07					
QDA	88.75					
SVM	91.89					
L4 - transformação Box-Cox						
$D_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	98.65	0.6	99.04	0.7	99.34
$D_R^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.5	98.65	0.4	99.04	0.3	99.34
$d_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	98.65	(0.4,0.6)	99.04	0.5	99.32
$d_{R2}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	98.65	0.5	99.04	(0.3,0.7)	99.34
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	-	-	-	-	-	-
$d_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-
$d_{R2_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-
$\mathrm{D}_{\mathrm{KL}}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	98.27	-	98.24	-	98.76
$\mathrm{J}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	98.13	-	98.37	-	98.85
$\mathrm{D}_{\mathrm{KL}}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	98.27	-	98.24	-	98.76
$\mathrm{J}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	98.13	-	98.37	-	98.85
$\mathrm{D}_{\mathrm{KL}}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	98.27	-	97.87	-	98.30
$\mathrm{J}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	79.88	-	79.95	-	80.27
$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	54.54	-	54.74	-	54.77
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	63.22	-	63.68	-	63.61
KNN	90.85					
LDA	91.70					
QDA	92.12					
SVM	91.89					

$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2) = \mathrm{D}_{KL}^{St}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1, \nu_1] \parallel [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2, \nu_2])$
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2) \quad = \mathrm{J}^{St}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1, \nu_1] \parallel [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2, \nu_2])$

Tabela B.5: Classificação para imagem sintética com número de *looks* 8.

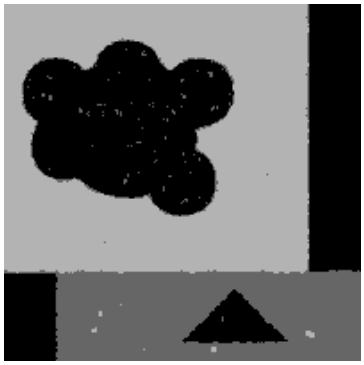
L8						
	janelamento 3×3		janelamento 5×5		janelamento 7×7	
	α	% accu.	α	% accu.	α	% accu.
$D_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	97.69	0.5	97.69	0.5	97.74
$D_R^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.5	97.69	0.5	97.69	0.5	97.74
$d_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	97.69	(0.4,0.6)	97.69	(0.3,0.7)	97.76
$d_{R2}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	97.69	0.5	97.69	0.5	97.74
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.1	98.26	0.1	97.65	0.1	97.65
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.9	98.26	0.9	97.65	0.9	97.65
$d_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.3,0.7)	98.12	(0.2,0.8)	97.41	(0.2,0.8)	97.41
$d_{R2_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.3,0.7)	98.13	(0.2,0.8)	97.41	(0.2,0.8)	97.41
$\mathrm{D}_{\mathrm{KL}}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	96.92	-	96.09	-	95.22
$\mathrm{J}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	96.12	-	95.51	-	94.81
$\mathrm{D}_{\mathrm{KL}}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	96.91	-	96.12	-	95.24
$\mathrm{J}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	95.45	-	94.90	-	94.16
$\mathrm{D}_{\mathrm{KL}}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	75.40	-	75.95	-	76.06
$\mathrm{J}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	68.97	-	69.43	-	69.48
$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	31.54	-	31.85	-	32.27
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	32.37	-	32.75	-	33.14
KNN	93.72					
LDA	82.42					
QDA	93.04					
QDA	95.00					
L8 - transformação Box-Cox						
$D_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.7	98.89	0.8	99.44	0.8	99.69
$D_R^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.3	98.89	0.2	99.44	0.2	99.69
$d_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	98.79	0.5	99.12	0.5	99.28
$d_{R2}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.3,0.7)	98.89	(0.2,0.8)	99.44	(0.2,0.8)	99.69
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-
$D_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	-	-	-	-	-	-
$d_{R_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-
$d_{R2_\mathbb{C}}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-
$\mathrm{D}_{\mathrm{KL}}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	98.47	-	98.85	-	99.20
$\mathrm{J}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	98.39	-	98.88	-	99.17
$\mathrm{D}_{\mathrm{KL}}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	98.47	-	98.85	-	99.21
$\mathrm{J}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	98.39	-	98.88	-	99.17
$\mathrm{D}_{\mathrm{KL}}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	98.34	-	97.97	-	98.32
$\mathrm{J}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	72.51	-	71.82	-	71.64
$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	65.33	-	65.15	-	65.05
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	65.74	-	64.50	-	63.96
KNN	94.37					
LDA	94.12					
QDA	95.27					
SVM	95.03					

$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2) = \mathrm{D}_{KL}^{St}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1, \nu_1] \parallel [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2, \nu_2])$
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2) \quad = \mathrm{J}^{St}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1, \nu_1] \parallel [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2, \nu_2])$

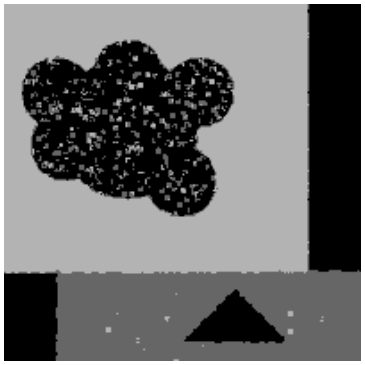


(a) áreas da imagem sintética

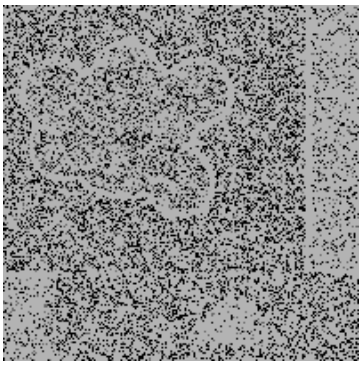
(b) L3 - $D^{0.1}_{R_{\mathbb{C}}}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 3 \times 3$



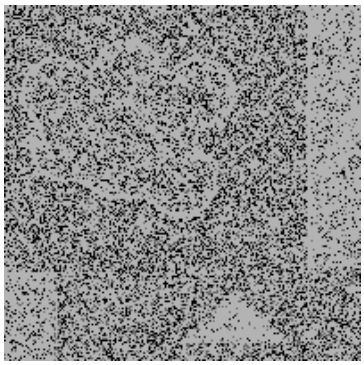
(c) L4 - $D^{0.1}_{R_{\mathbb{C}}}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 3 \times 3$



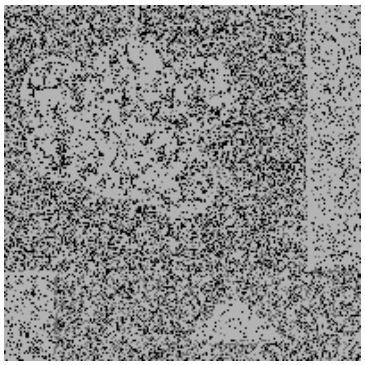
(d) L8 - $D^{0.1}_{R_{\mathbb{C}}}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 3 \times 3$



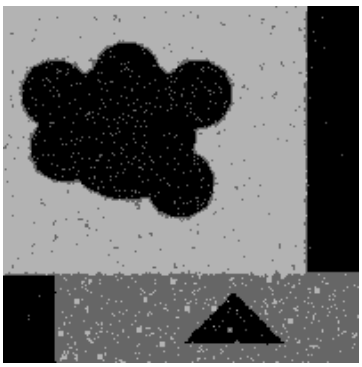
(e) L3 - $\mathrm{J}^{St}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 7 \times 7$



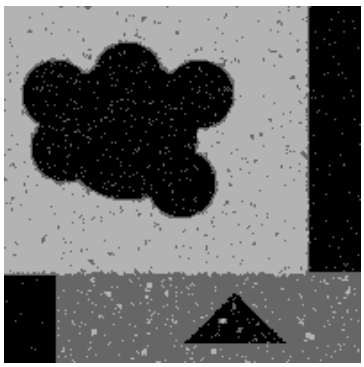
(f) L4 - $\mathrm{D}^{St}_{KL}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 5 \times 5$



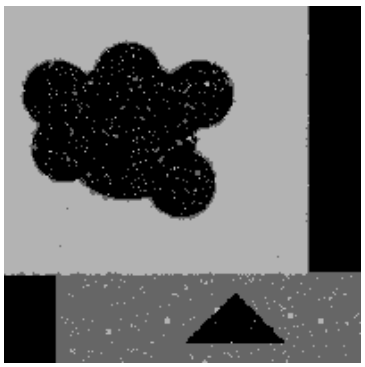
(g) L8 - $\mathrm{D}^{St}_{KL}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 3 \times 3$



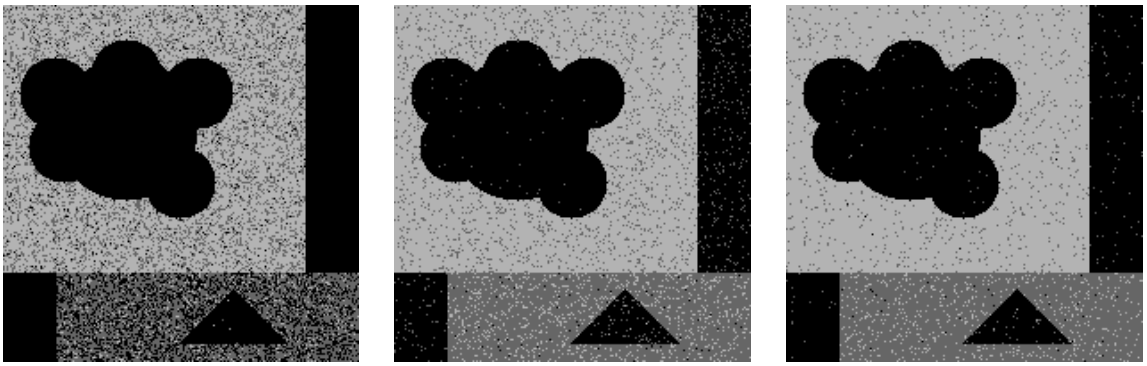
(h) L3 - $\mathrm{D}^{N}_{\mathrm{KL}}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 3 \times 3$



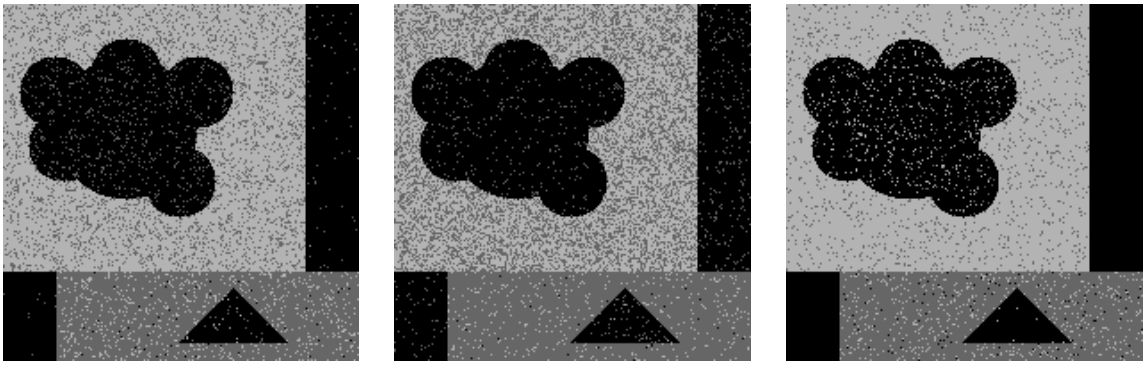
(i) L4 - $\mathrm{D}^{N}_{\mathrm{KL}}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 3 \times 3$



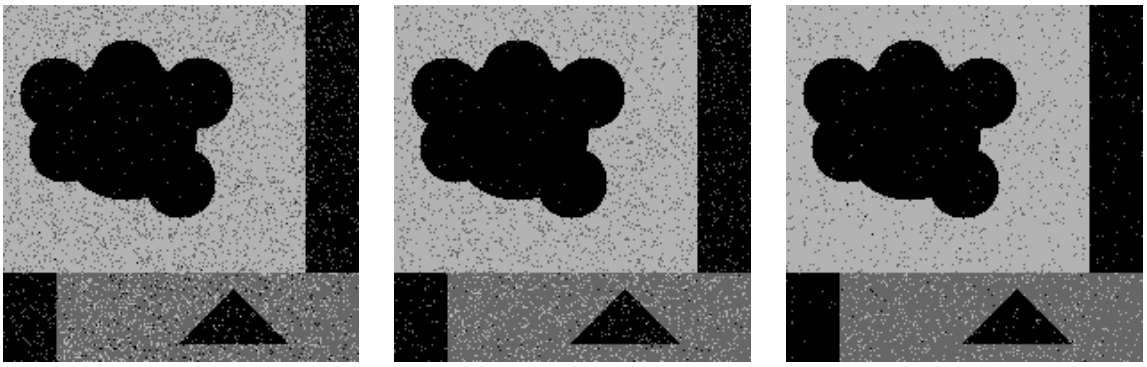
(j) L8 - $\mathrm{D}^{N}_{\mathrm{KL}}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 3 \times 3$



(k) L3 - LDA (l) L4 - LDA (m) L8 - LDA



(n) L3 - QDA (o) L4 - QDA (p) L8 - QDA



(q) L3 - KNN (r) L4 - KNN (s) L8 - KNN

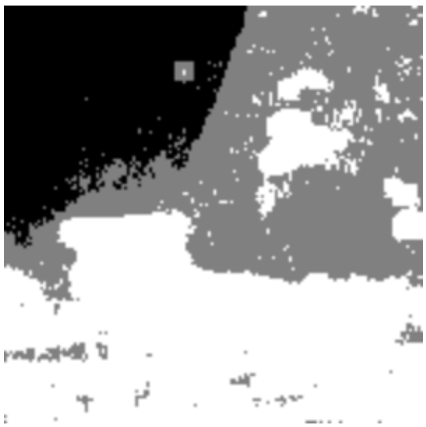
Figura B.1: Classificação da imagem sintética para *looks* = 3, 4 e 8.

Tabela B.6: Classificação para a imagem San Francisco (USA).

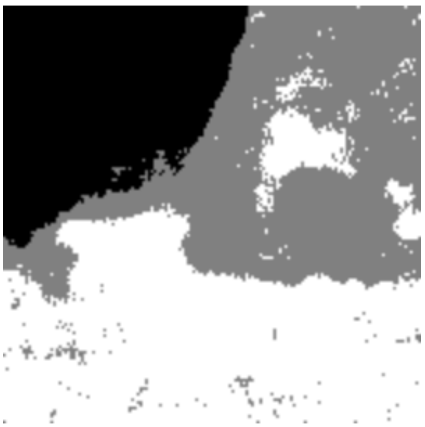
San Francisco												
	janelamento 3×3				janelamento 5×5				janelamento 7×7			
	α	% accu.	kappa	var. kappa	α	% accu.	kappa	var. kappa	α	% accu.	kappa	var. kappa
$D_R^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.9	96.24	94.37	3.14×10^{-5}	0.9	98.92	98.37	9.34×10^{-6}	0.9	99.69	99.54	2.69×10^{-6}
$D_R^{\alpha}(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.1	96.24	94.37	3.14×10^{-5}	0.1	98.92	98.37	9.34×10^{-6}	0.1	99.69	99.54	2.69×10^{-6}
$d_R^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	83.46	75.20	1.14×10^{-4}	0.5	85.91	78.86	1.01×10^{-4}	0.5	87.14	80.72	9.41×10^{-5}
$d_{R2}^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.1,0.9)	94.89	92.33	4.21×10^{-5}	(0.1,0.9)	97.60	96.40	2.04×10^{-5}	(0.1,0.9)	98.06	97.10	1.65×10^{-5}
$D_{R_c}^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.9	91.09	86.64	6.96×10^{-5}	0.9	96.48	94.71	2.95×10^{-5}	0.9	98.72	98.08	1.10×10^{-5}
$D_{R_c}^{\alpha}(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.1	91.09	86.64	6.96×10^{-5}	0.1	96.48	94.71	2.95×10^{-5}	0.1	98.72	98.08	1.10×10^{-5}
$d_{R_c}^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.4,0.5,0.6)	87.81	81.71	9.08×10^{-5}	...[1]	94.54	91.81	4.47×10^{-5}	...[2]	97.44	96.17	2.16×10^{-5}
$d_{R2_c}^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	...[3]	87.84	81.76	9.05×10^{-5}	...[4]	94.54	91.81	4.47×10^{-5}	...[5]	97.40	96.11	2.19×10^{-5}
$\mathrm{D}_{\mathrm{KL}}^{N}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	96.86	95.30	2.64×10^{-5}	-	98.64	97.97	1.16×10^{-5}	-	98.95	98.43	9.01×10^{-6}
$\mathrm{J}^{N}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	87.46	81.18	9.36×10^{-5}	-	90.59	85.89	7.34×10^{-5}	-	93.19	89.78	5.51×10^{-5}
$\mathrm{D}_{\mathrm{KL}}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	96.86	95.30	2.64×10^{-5}	-	98.64	97.97	1.16×10^{-5}	-	98.95	98.43	9.01×10^{-6}
$\mathrm{J}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	87.46	81.18	9.36×10^{-5}	-	90.59	85.89	7.34×10^{-5}	-	93.19	89.78	5.51×10^{-5}
$\mathrm{D}_{\mathrm{KL}}^{t}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	91.99	87.98	6.38×10^{-5}	-	94.46	91.69	4.54×10^{-5}	-	94.19	91.29	4.75×10^{-5}
$\mathrm{J}^{t}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	58.08	31.13	2.71×10^{-4}	-	55.21	32.81	2.70×10^{-4}	-	55.48	33.22	2.63×10^{-4}
$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	38.37	7.55	3.51×10^{-4}	-	40.03	10.05	3.27×10^{-4}	-	39.61	9.41	3.41×10^{-4}
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	38.06	7.08	3.56×10^{-4}	-	39.45	9.17	3.33×10^{-4}	-	38.83	8.25	3.49×10^{-4}
KNN	% accu.→		89.78		kappa →		84.67		var.kappa →		7.97×10^{-5}	
LDA	% accu.→		57.07		kappa →		35.60		var.kappa →		2.30×10^{-4}	
QDA	% accu.→		85.33		kappa →		77.99		var.kappa →		1.07×10^{-4}	
SVM	% accu.→		90.71		kappa →		83.86		var.kappa →		8.34×10^{-5}	
San Francisco - transformação Box-Cox												
$D_R^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.8	99.07	98.60	8.01×10^{-6}	0.9	99.57	99.36	3.69×10^{-6}	0.9	99.73	99.59	2.35×10^{-6}
$D_R^{\alpha}(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.2	99.07	98.60	8.01×10^{-6}	0.1	99.57	99.36	3.69×10^{-6}	0.1	99.73	99.59	2.35×10^{-6}
$d_R^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	98.64	97.96	1.16×10^{-5}	0.5	99.10	98.66	7.68×10^{-6}	0.5	99.38	99.07	5.36×10^{-6}
$d_{R2}^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.2,0.8)	99.07	98.60	8.01×10^{-6}	...[6]	99.54	98.66	7.68×10^{-6}	(0.1,0.9)	99.73	99.59	2.35×10^{-6}
$D_{R_c}^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-	-	-	-	-	-	-
$D_{R_c}^{\alpha}(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	-	-	-	-	-	-	-	-	-	-	-	-
$d_{R_c}^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-	-	-	-	-	-	-
$d_{R2_c}^{\alpha}(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-	-	-	-	-	-	-
$\mathrm{D}_{\mathrm{KL}}^{N}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	98.53	97.79	1.26×10^{-5}	-	99.19	98.78	7.02×10^{-6}	-	99.26	98.89	6.36×10^{-6}
$\mathrm{J}^{N}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	97.60	96.39	2.04×10^{-5}	-	98.72	98.08	1.09×10^{-5}	-	99.07	98.60	8.01×10^{-6}
$\mathrm{D}_{\mathrm{KL}}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	98.53	97.79	1.26×10^{-5}	-	99.19	98.78	7.02×10^{-6}	-	99.26	98.89	6.36×10^{-6}
$\mathrm{J}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	97.60	96.39	2.04×10^{-5}	-	98.72	98.08	1.09×10^{-5}	-	99.07	98.60	8.01×10^{-6}
$\mathrm{D}_{\mathrm{KL}}^{t}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	98.14	97.21	1.58×10^{-5}	-	99.15	98.72	7.35×10^{-6}	-	99.15	98.72	7.35×10^{-6}
$\mathrm{J}^{t}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	98.22	97.32	1.52×10^{-5}	-	98.80	98.19	1.03×10^{-5}	-	98.99	98.49	8.67×10^{-6}
$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	71.89	57.83	1.67×10^{-4}	-	70.11	55.16	1.74×10^{-4}	-	70.65	55.98	1.73×10^{-4}
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	68.99	53.48	1.72×10^{-4}	-	67.87	51.80	1.77×10^{-4}	-	68.18	52.26	1.75×10^{-4}
KNN	% accu.→		90.44		kappa →		85.65		var.kappa →		7.51×10^{-5}	
LDA	% accu.→		88.73		kappa →		83.10		var.kappa →		8.65×10^{-5}	
QDA	% accu.→		87.77		kappa →		81.64		var.kappa →		9.31×10^{-5}	
SVM	% accu.→		90.71		kappa →		86.06		var.kappa →		7.33×10^{-5}	

[1] $\alpha = (0.1, 0.4, 0.5, 0.6, 0.9)$, [2] $\alpha = (0.1, 0.3, 0.7, 0.9)$, [3] $\alpha = (0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.9)$,
[4] $\alpha = (0.3, 0.4, 0.5, 0.6, 0.7)$, [5] $\alpha = (0.1, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9)$, [6] $\alpha = (0.1, 0.2, 0.8, 0.9)$

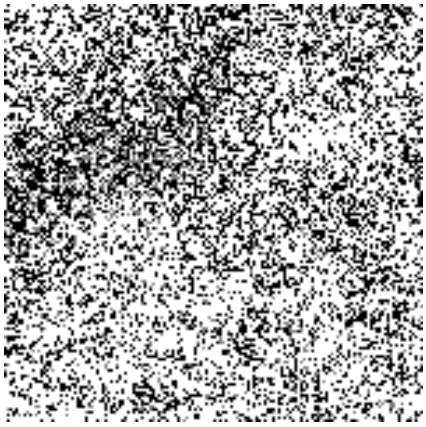
(c) *Ground-truth data* de São Francisco;



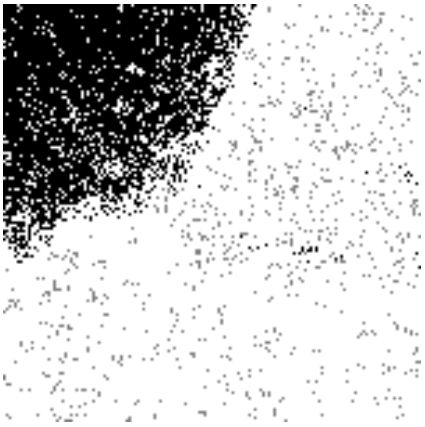
(d) $D_R^{0.9}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 7 \times 7$



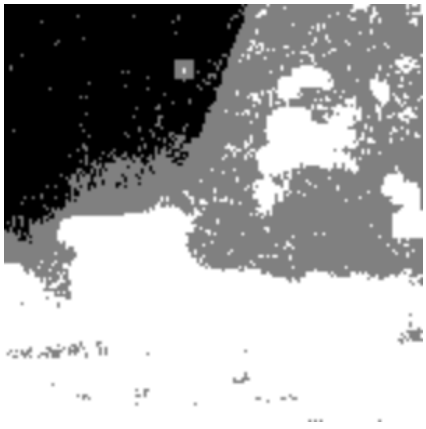
(e) $D_R^{0.9}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 7 \times 7$ - BC



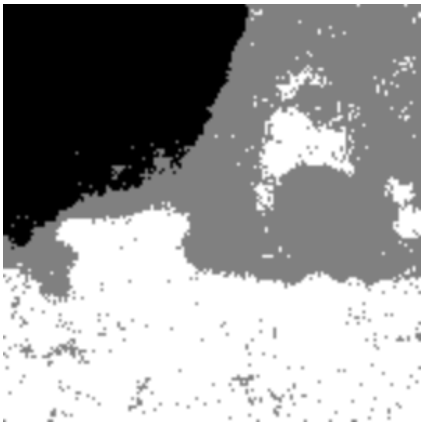
(f) $J^{St}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 3 \times 3$



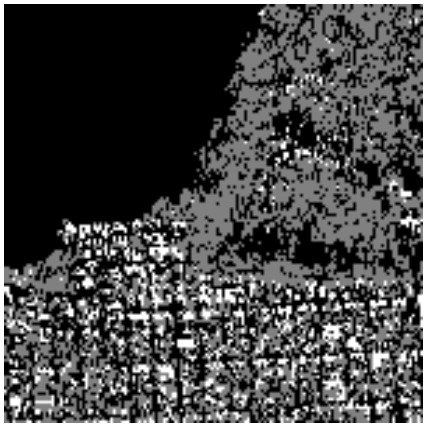
(g) $J^{St}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 5 \times 5$ - BC



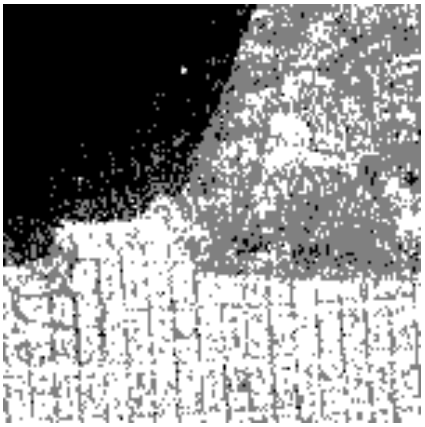
(h) $D_{KL}^{N}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 7 \times 7$



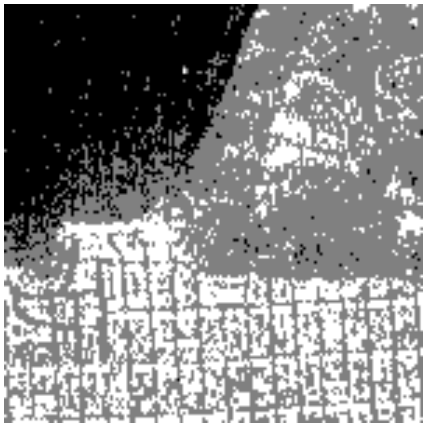
(i) $D_{KL}^{N}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 7 \times 7$ - BC



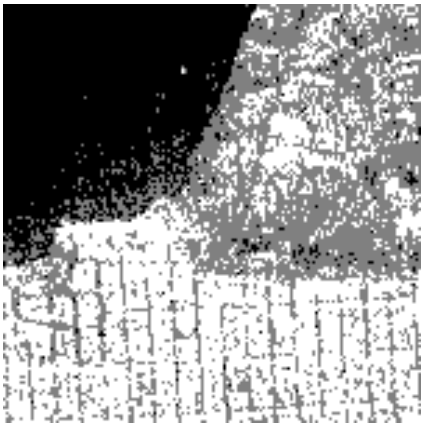
(j) LDA



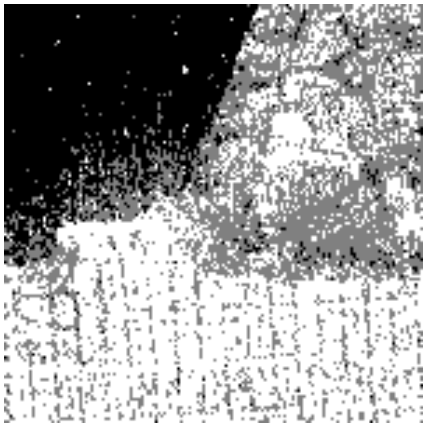
(k) LDA - BC



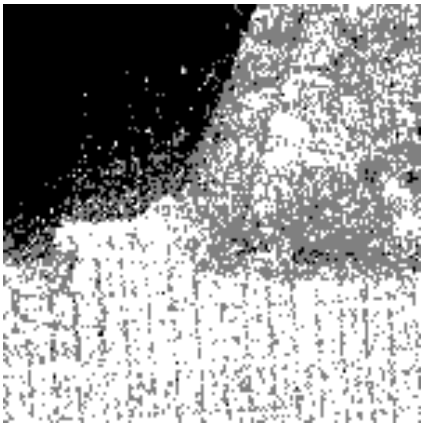
(l) QDA



(m) QDA - BC



(n) KNN



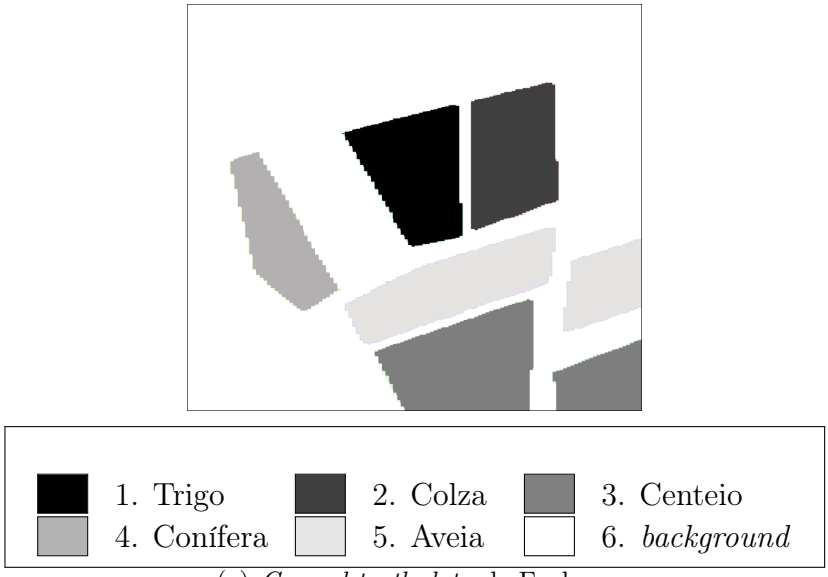
(o) KNN - BC

Figura B.2: Classificação de São Francisco

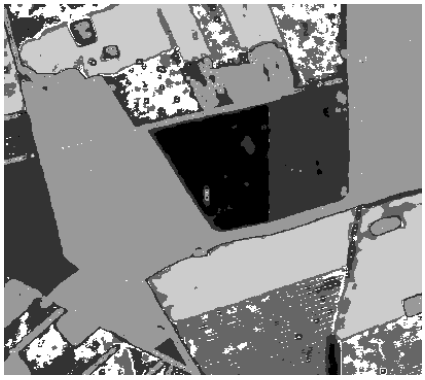
Tabela B.7: Classificação para a imagem Foulum (DK).

Foulum												
	janelamento 3×3				janelamento 5×5				janelamento 7×7			
	α	% accu.	kappa	var. kappa	α	% accu.	kappa	var. kappa	α	% accu.	kappa	var. kappa
$D_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.7	81.67	77.65	2.50×10^{-5}	0.8	83.47	79.83	2.31×10^{-5}	0.9	84.42	80.98	2.22×10^{-5}
$D_R^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.3	81.67	77.65	2.50×10^{-5}	0.2	83.47	79.83	2.31×10^{-5}	0.1	84.42	80.98	2.22×10^{-5}
$d_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	81.43	77.38	2.51×10^{-5}	(0.3,0.7)	82.61	78.81	2.37×10^{-5}	(0.4,0.6)	83.30	79.65	2.30×10^{-5}
$d_{R2}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	81.43	77.38	2.51×10^{-5}	(0.2,0.8)	82.66	78.87	2.38×10^{-5}	...[1]	83.31	79.67	2.30×10^{-5}
$D_{R_C}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.8	70.22	63.91	3.59×10^{-5}	0.8	71.72	65.70	3.47×10^{-5}	0.8	71.51	65.43	3.46×10^{-5}
$D_{R_C}^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.2	70.22	63.91	3.59×10^{-5}	0.2	71.72	65.70	3.47×10^{-5}	0.2	71.51	65.43	3.46×10^{-5}
$d_{R_C}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.2,0.8)	69.94	63.40	3.60×10^{-5}	(0.2,0.8)	71.00	64.61	3.55×10^{-5}	(0.2,0.8)	70.36	63.78	3.58×10^{-5}
$d_{R2_C}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.2,0.8)	67.31	60.26	3.72×10^{-5}	(0.2,0.8)	67.58	60.45	3.70×10^{-5}	(0.2,0.8)	67.20	59.96	3.68×10^{-5}
$\mathrm{D}_{\mathrm{KL}}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	80.23	75.86	2.69×10^{-5}	-	82.69	78.86	2.43×10^{-5}	-	83.52	79.86	2.34×10^{-5}
$\mathrm{J}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	75.92	70.64	3.11×10^{-5}	-	79.30	74.75	2.79×10^{-5}	-	80.23	75.88	2.69×10^{-5}
$\mathrm{D}_{\mathrm{KL}}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	80.23	75.86	2.69×10^{-5}	-	82.69	78.86	2.43×10^{-5}	-	83.52	79.86	2.34×10^{-5}
$\mathrm{J}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	75.92	70.64	3.11×10^{-5}	-	79.30	74.75	2.79×10^{-5}	-	80.23	75.88	2.69×10^{-5}
$\mathrm{D}_{\mathrm{KL}}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	90.67	88.60	1.51×10^{-5}	-	93.18	91.67	1.14×10^{-5}	-	93.05	91.50	1.16×10^{-5}
$\mathrm{J}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	79.34	74.87	2.89×10^{-5}	-	82.13	78.24	2.60×10^{-5}	-	82.99	79.30	2.50×10^{-5}
$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	27.58	12.36	3.60×10^{-5}	-	28.78	13.78	3.67×10^{-5}	-	29.05	14.16	3.69×10^{-5}
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	30.66	16.43	3.81×10^{-5}	-	30.86	16.60	3.81×10^{-5}	-	31.73	17.68	3.85×10^{-5}
KNN	% accu.→		85.04		kappa →		81.77		var.kappa →		2.26×10^{-5}	
LDA	% accu.→		70.22		kappa →		62.2		var.kappa →		3.68×10^{-5}	
QDA	% accu.→		74.54		kappa →		71.01		var.kappa →		3.09×10^{-5}	
SVM	% accu.→		86.53		kappa →		83.59		var.kappa →		2.06×10^{-5}	
Foulum - transformação Box-Cox												
$D_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.9	92.10	90.38	1.29×10^{-5}	0.9	93.67	92.28	1.05×10^{-5}	0.9	94.79	93.65	8.78×10^{-6}
$D_R^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	0.1	92.10	90.38	1.29×10^{-5}	0.1	93.67	92.28	1.05×10^{-5}	0.1	94.79	93.65	8.78×10^{-6}
$d_R^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	0.5	90.41	88.31	1.52×10^{-5}	0.5	91.25	89.35	1.39×10^{-5}	0.5	91.97	90.22	1.29×10^{-5}
$d_{R2}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	(0.1,0.9)	92.07	90.34	1.29×10^{-5}	(0.1,0.9)	93.64	92.25	1.05×10^{-5}	(0.1,0.9)	94.69	93.54	8.93×10^{-6}
$D_{R_C}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-	-	-	-	-	-	-
$D_{R_C}^\alpha(\boldsymbol{\theta}_2 \| \boldsymbol{\theta}_1)$	-	-	-	-	-	-	-	-	-	-	-	-
$d_{R_C}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-	-	-	-	-	-	-
$d_{R2_C}^\alpha(\boldsymbol{\theta}_1 \| \boldsymbol{\theta}_2)$	-	-	-	-	-	-	-	-	-	-	-	-
$\mathrm{D}_{\mathrm{KL}}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	92.08	90.35	1.29×10^{-5}	-	94.33	93.09	9.53×10^{-6}	-	94.90	93.79	8.63×10^{-6}
$\mathrm{J}^N([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2])$	-	86.46	83.50	2.01×10^{-5}	-	88.29	85.73	1.78×10^{-5}	-	88.75	86.29	1.72×10^{-5}
$\mathrm{D}_{\mathrm{KL}}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	92.08	90.35	1.29×10^{-5}	-	94.33	93.09	9.53×10^{-6}	-	94.90	93.79	8.63×10^{-6}
$\mathrm{J}^{SN}([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\eta}_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\eta}_2])$	-	86.46	83.50	2.01×10^{-5}	-	88.29	85.73	1.78×10^{-5}	-	88.75	86.29	1.72×10^{-5}
$\mathrm{D}_{\mathrm{KL}}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	93.99	92.68	1.01×10^{-5}	-	95.95	95.07	6.96×10^{-6}	-	96.77	96.06	5.61×10^{-6}
$\mathrm{J}^t([\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1] \| [\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2])$	-	68.98	62.16	3.65×10^{-5}	-	68.68	61.81	3.64×10^{-5}	-	68.19	61.23	3.67×10^{-5}
$\mathrm{D}_{KL}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	78.14	73.29	2.90×10^{-5}	-	78.38	73.58	2.88×10^{-5}	-	78.32	73.52	2.88×10^{-5}
$\mathrm{J}^{St}(\boldsymbol{\theta}_1 \parallel \boldsymbol{\theta}_2)$	-	75.64	70.21	3.12×10^{-5}	-	76.13	70.81	3.08×10^{-5}	-	76.76	71.59	3.02×10^{-5}
KNN	% accu.→		86.37		kappa →		83.40		var.kappa →		2.08×10^{-5}	
LDA	% accu.→		86.26		kappa →		83.29		var.kappa →		2.09×10^{-5}	
QDA	% accu.→		85.53		kappa →		82.37		var.kappa →		2.20×10^{-5}	
SVM	% accu.→		87.22		kappa →		84.43		var.kappa →		1.97×10^{-5}	

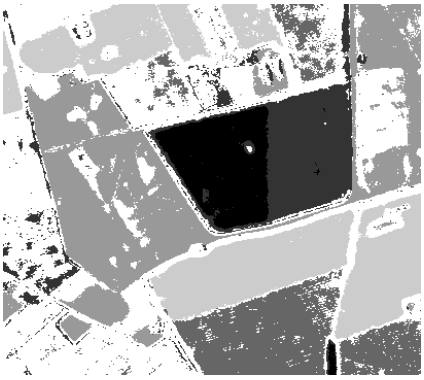
[1] $\alpha = (0.1, 0.4, 0.6, 0.9)$



(a) *Ground-truth data* de Foulum;

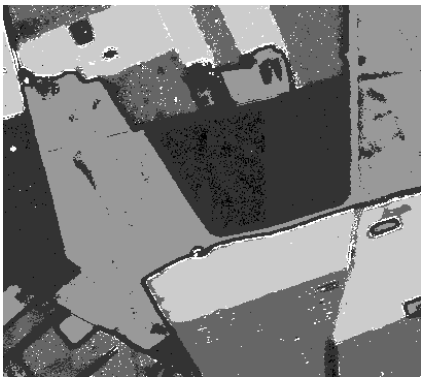


(b) $D_{KL}^{t}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 5 \times 5$

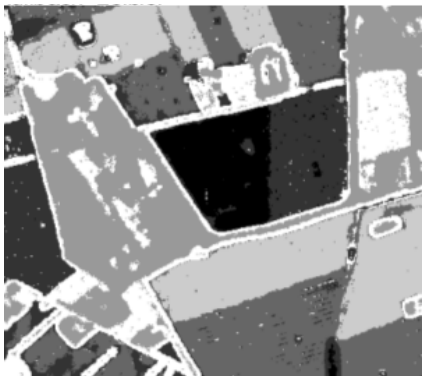


(c) $D_{KL}^{t}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 7 \times 7$ - BC

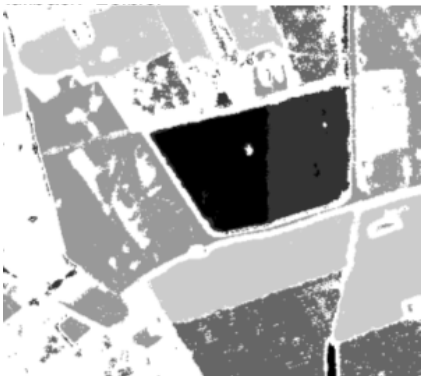
(d) $D_{KL}^{St}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 3 \times 3$



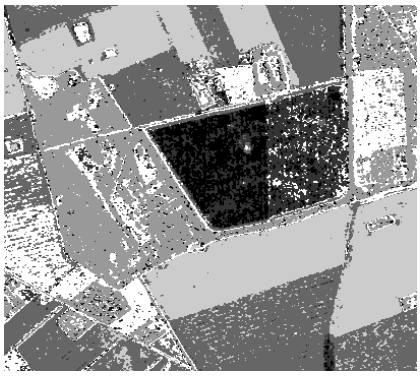
(e) $J^{t}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 7 \times 7$ - BC



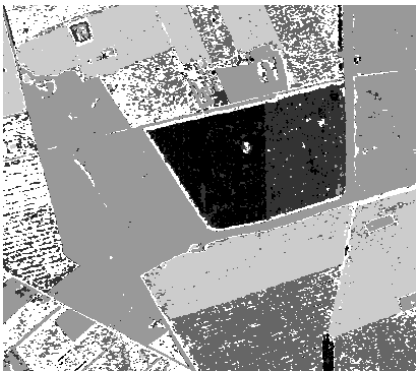
(f) $D_{KL}^{N}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 7 \times 7$



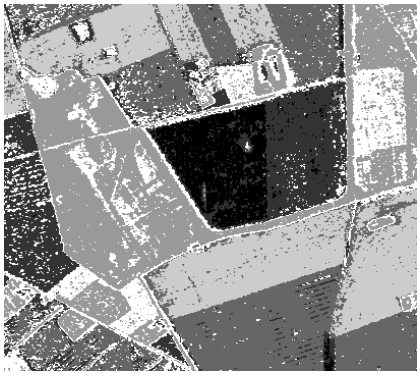
(g) $D_{KL}^{N}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) - 7 \times 7$ - BC



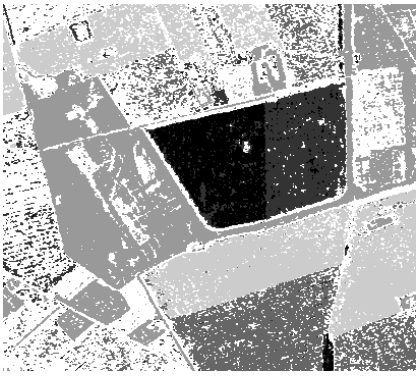
(h) LDA



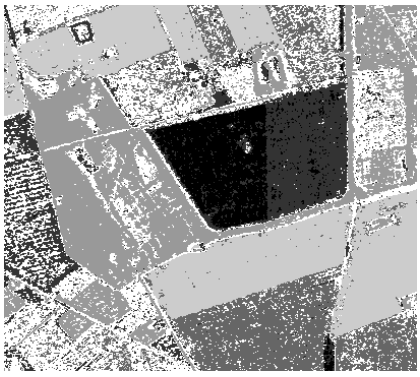
(i) LDA - BC



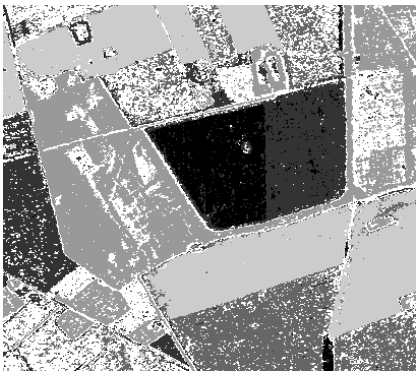
(j) QDA



(k) QDA - BC



(l) KNN



(m) KNN - BC

Figura B.3: Classificação de Foulum

APÊNDICE C

Estimação do Parâmetro α para as Imagens de São Francisco e Foulum

Tabela C.1: Resultados das estimações para α ótimo.

Expressão 1						
	Janela 3×3		Janela 5×5		Janela 7×7	
IMAGEM UTILIZADA	α	$\hat{\alpha}$ por BFGS	α	$\hat{\alpha}$ por BFGS	α	$\hat{\alpha}$ por BFGS
Sem Transformação	-	-	-	-	-	-
São Francisco	0.9	0.9	0.9	0.9	0.9	0.9
Foulum	0.6	0.9	0.7	0.9	0.7	0.9
Transformação Box-Cox	-	-	-	-	-	-
São Francisco	0.7	0.9	0.9	0.9	0.9	0.9
Foulum	0.9	0.9	0.9	0.888	0.9	0.883
Expressão 2						
Sem Transformação	-	-	-	-	-	-
São Francisco	0.1	0.1	0.1	0.1	0.1	0.1
Foulum	0.4	0.1	0.3	0.1	0.3	0.1
Transformação Box-Cox	-	-	-	-	-	-
São Francisco	0.3	0.1	0.1	0.1	0.1	0.1
Foulum	0.1	0.1	0.1	0.112	0.1	0.117
Expressão 3						
Sem Transformação	-	-	-	-	-	-
São Francisco	0.5	0.499	0.5	0.499	0.5	0.500
Foulum	0.4	0.500	0.5	0.499	0.5	0.5
Transformação Box-Cox	-	-	-	-	-	-
São Francisco	0.5	0.500	0.5	0.499	0.5	0.499
Foulum	0.5	0.499	0.5	0.500	0.5	0.499
Expressão 4						
Sem Transformação	-	-	-	-	-	-
São Francisco	(0.1,0.9)	0.9	(0.1,0.9)	0.9	(0.1,0.9)	0.695
Foulum	(0.4,0.6)	0.610	(0.3,0.7)	0.616	(0.4,0.6)	0.616
Transformação Box-Cox	-	-	-	-	-	-
São Francisco	(0.2,0.8)	0.9	(0.1,0.9)	0.9	(0.1,0.9)	0.9
Foulum	(0.1,0.9)	0.9	(0.1,0.9)	0.9	(0.1,0.9)	0.9

As expressões citadas nas tabelas estão definidas em (3.17).

Tabela C.2: Resultados para Viés e EQM.

Expressão 1						
	Janela 3×3		Janela 5×5		Janela 7×7	
	$\hat{\alpha}$ por BFGS		$\hat{\alpha}$ por BFGS		$\hat{\alpha}$ por BFGS	
IMAGEM UTILIZADA	Viés	EQM	Viés	EQM	Viés	EQM
Sem Transformação						
São Francisco	0	-	0	-	0	-
Foulum	-0.3	-	-0.2	-	-0.2	-
Transformação Box-Cox						
São Francisco	-0.2	-	0	-	0	-
Foulum	0	-	0.012	-	0.017	-
Expressão 2						
Sem Transformação						
São Francisco	0	-	0	-	0	-
Foulum	0.3	-	0.2	-	0.2	-
Transformação Box-Cox						
São Francisco	0.2	-	0	-	0	-
Foulum	0	-	-0.012	-	-0.017	-
Expressão 3						
Sem Transformação						
São Francisco	0.001	-	0.001	-	0	-
Foulum	-0.1	-	0.001	-	0	-
Transformação Box-Cox						
São Francisco	0	-	0.001	-	0.001	-
Foulum	0.001	-	0	-	0.001	-
Expressão 4						
Sem Transformação						
São Francisco	0	-	0	-	0.205	-
Foulum	-0.01	-	0.084	-	-0.016	-
Transformação Box-Cox						
São Francisco	-0.1	-	0	-	0	-
Foulum	0	-	0	-	0	-

Tabela C.3: Resultados das estimações utilizando Bootstrap.

Expressão 1						
	Janela 3×3		Janela 5×5		Janela 7×7	
IMAGEM UTILIZADA	α	$\hat{\alpha}$ por BFGS	α	$\hat{\alpha}$ por BFGS	α	$\hat{\alpha}$ por BFGS
Sem Transformação	-	-	-	-	-	-
São Francisco	0.9	0.900	0.9	0.900	0.9	0.900
Foulum	0.6	0.900	0.7	0.900	0.7	0.900
Transformação Box-Cox	-	-	-	-	-	-
São Francisco	0.7	0.898	0.9	0.894	0.9	0.892
Foulum	0.9	0.866	0.9	0.865	0.9	0.864
Expressão 2						
Sem Transformação	-	-	-	-	-	-
São Francisco	0.1	0.112	0.1	0.104	0.1	0.107
Foulum	0.4	0.104	0.3	0.113	0.3	0.106
Transformação Box-Cox	-	-	-	-	-	-
São Francisco	0.3	0.101	0.1	0.106	0.1	0.108
Foulum	0.1	0.134	0.1	0.135	0.1	0.136
Expressão 3						
Sem Transformação	-	-	-	-	-	-
São Francisco	0.5	0.506	0.5	0.502	0.5	0.505
Foulum	0.4	0.502	0.5	0.506	0.5	0.503
Transformação Box-Cox	-	-	-	-	-	-
São Francisco	0.5	0.500	0.5	0.500	0.5	0.500
Foulum	0.5	0.500	0.5	0.500	0.5	0.500
Expressão 4						
Sem Transformação	-	-	-	-	-	-
São Francisco	(0.1,0.9)	0.870	(0.1,0.9)	0.830	(0.1,0.9)	0.801
Foulum	(0.4,0.6)	0.543	(0.3,0.7)	0.547	(0.4,0.6)	0.547
Transformação Box-Cox	-	-	-	-	-	-
São Francisco	(0.2,0.8)	0.900	(0.1,0.9)	0.900	(0.1,0.9)	0.900
Foulum	(0.1,0.9)	0.899	(0.1,0.9)	0.899	(0.1,0.9)	0.899

Tabela C.4: Resultados Viés e EQM utilizando Bootstrap.

Expressão 1						
	Janela 3×3		Janela 5×5		Janela 7×7	
	$\hat{\alpha}$ por BFGS		$\hat{\alpha}$ por BFGS		$\hat{\alpha}$ por BFGS	
IMAGEM UTILIZADA	Viés	EQM	Viés	EQM	Viés	EQM
Sem Transformação						
São Francisco	0	0	0	0	0	0
Foulum	0.300	0.09	0.200	0.04	0.200	0.04
Transformação Box-Cox						
São Francisco	0.198	0.039	-0.006	1.1×10^{-4}	-0.008	1.6×10^{-4}
Foulum	-0.034	0.001	-0.035	0.001	-0.036	0.001
Expressão 2						
Sem Transformação						
São Francisco	0.012	0.010	0.004	0.003	0.007	0.006
Foulum	-0.296	0.090	-0.187	0.045	-0.194	0.042
Transformação Box-Cox						
São Francisco	-0.198	0.039	0.006	1.1×10^{-4}	0.008	1.6×10^{-4}
Foulum	0.034	0.001	0.035	0.001	0.0358	0.001
Expressão 3						
Sem Transformação						
São Francisco	0.006	0.003	0.002	0.001	0.005	0.002
Foulum	0.102	0.011	6.4×10^{-3}	0.003	0.003	0.001
Transformação Box-Cox						
São Francisco	0	0	0	0	0	0
Foulum	7.033×10^{-11}	4.623×10^{-20}	7.391×10^{-11}	5.014×10^{-20}	7.829×10^{-11}	5.431×10^{-20}
Expressão 4						
Sem Transformação						
São Francisco	-0.030	0.011	-0.070	0.024	-0.099	0.034
Foulum	-0.056	0.007	-0.153	0.028	-0.053	0.007
Transformação Box-Cox						
São Francisco	0.100	0.001	0	0	0	0
Foulum	-3.373×10^{-5}	3.921×10^{-7}	-1.700×10^{-5}	1.449×10^{-7}	-1.517×10^{-5}	1.132×10^{-7}