

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Claudomir Cardoso de Carvalho Junior

Classificação Automática de Modulação
Baseada em Aprendizagem Discriminativa

TD - 07/2015

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2015

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Claudomir Cardoso de Carvalho Junior

Classificação Automática de Modulação
Baseada em Aprendizagem Discriminativa

Tese submetida à Banca Examinadora do Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal do Pará, como parte dos requisitos para a obtenção do Grau de Doutor em Engenharia Elétrica, ênfase em Computação Aplicada.

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2015

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da UFPA

Carvalho Junior, Claudomir Cardoso de, 1982 -

Classificação automática de modulação baseada em aprendizagem discriminativa

Claudomir Cardoso de Carvalho Junior. - 2015.

Orientador: Aldebaro Klautau Júnior.

Tese (Doutorado) - Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2015.

1. Rádio definido por software (RDS). 2. Algoritmos de classificação. 3. Modulação Digital.

I.

Título.

CDD 22. ed. 621.384

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Classificação Automática de Modulação Baseada em Aprendizagem Discriminativa

AUTOR: CLAUDOMIR CARDOSO DE CARVALHO JUNIOR

TESE DE DOUTORADO SUBMETIDA À AVALIAÇÃO DA BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA DA UNIVERSIDADE FEDERAL DO PARÁ E JULGADA ADEQUADA PARA OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM 19/06/2015

BANCA EXAMINADORA:

.....
Prof. Dr. Aldebaro Barreto da Rocha Klautau Júnior (UFPA) - Orientador

.....
Prof. Dr. Carlos Aurélio Faria da Rocha (UFSC) - Membro externo

.....
Prof. Dr. Cláudio Alex Jorge da Rocha (IFPA) - Membro externo

.....
Prof. Dr. Glaucio Haroldo Silva de Carvalho (UFPA) - Membro

.....
Prof. Dr. João Crisóstomo Weyl Albuquerque Costa (UFPA) - Membro

VISTO:

.....
Prof. Dr. Evaldo Gonçalves Pelaes

COORDENADOR DO PPGEE/ITEC/UFPA

Agradecimentos

A caminhada foi mais longa do que pensava, mas finalmente a realização de mais essa etapa da vida acadêmica se concretizou. Defender a Tese de Doutorado não seria possível sem a ajuda e o incentivo de um grupo de pessoas que auxiliaram de forma direta ou indireta para realização desta conquista.

Agradeço primeiramente ao Meu Deus pelos desafios e problemas ocorridos nessa caminhada. Obrigado por ter me capacitado o suficiente para enfrentar essas dificuldades, as quais acabaram resultando nessa conquista pessoal. Obrigado por ter colocado amigos verdadeiros e leais em meu caminho durante este longo período de estudo, pois terminar essa etapa, sem ajuda deles, seria mais difícil.

Agradeço ao meu pai, Claudomir, e minha mãe, Carmen, pela imensa ajuda e incentivo aos estudos proporcionados desde à infância. Nesta conquista, vocês foram fundamentais a fim de que pudesse alcançá-la.

Ao meu orientador Prof. Aldebaro Barreto da Rocha Klautau Júnior pela amizade, apoio e pela maneira atenciosa como conduziu este trabalho.

A todos os amigos do Laboratório de Processamento de Sinais (LAPS) pelo ambiente de verdadeira amizade e companheirismo proporcionado ao longo dos anos, especialmente a Francisco Muller, Bruno Haick, Pedro Batista, Carolina Quintão, Marcos Takeda, Johathas Trindade, Marcus Dias, Rafael Sousa e a Kelly Souza por estar sempre disposta a ajudar.

E por fim, agradeço a todos os professores da UFPA que me ajudaram desde a graduação a atingir esse objetivo.

Muito Obrigado.

Resumo

A principal finalidade dos algoritmos para classificação automática de modulação em um receptor inteligente para rádio cognitivo ou aplicações militares é a identificação da modulação de um sinal para que o receptor possa, posteriormente, realizar o processo de demodulação e em seguida o processamento da informação recebida. A falta de conhecimento acerca de parâmetros como, por exemplo: potência do sinal, frequência e fase da portadora, sincronismo temporal e outros; somados aos efeitos dos canais como desvanecimento de multipercurso, tornam o problema de classificação automática de modulação desafiador e bastante investigado atualmente. Neste contexto, esta Tese apresenta uma investigação sobre algoritmos de classificação automática de modulação baseados na aprendizagem discriminativa e avalia o desempenho dos mesmos em distintos cenários de uso. A Tese apresenta uma nova proposta de classificação baseada no extrator de parâmetro (*front end*) chamado de HISTO. A classificação é realizada sobre os *símbolos* recebidos, onde os histogramas de magnitude e fase são calculados. Os resultados obtidos usam o extrator de parâmetro proposto e o algoritmo Máquina de Vetores de Suporte (MVS), os quais são comparados com outras técnicas. Na literatura, alguns dos algoritmos propostos são avaliados assumindo-se o conhecimento de parâmetros da modulação e em cenários distintos, dificultando a comparação entre os algoritmos. Nesta Tese, algoritmos representativos do estado da arte são comparados com os propostos de maneira sistemática e uniforme, incluindo-se cenários não-ideais como os com *phase jitter* e *offset* de frequência. Os resultados das simulações mostram que proposta de classificação baseada em histogramas é eficiente com relativo baixo custo computacional.

Abstract

The main purpose of the algorithms for automatic modulation classification in an intelligent receiver for cognitive radio or military applications is to identify the modulation of a signal so that the receiver can subsequently carry out the process of demodulation and then processing the received information. The lack of knowledge of parameters such as, for example, signal power, frequency and phase of carrier and timing; added to the channel effects such as multipath fading, make the automatic modulation classification problem challenging and fairly investigated currently. In this context, this thesis presents an investigation of automatic modulation classification algorithms based on discriminative learning and evaluates their performance in different usage scenarios. The thesis presents a new classification proposal based on front end called HISTO. Classification is performed on the received symbols, where the magnitude and phase histogram's are computed. The results obtained using the proposed front end and the Support Vector Machine algorithm (SVM) are compared with other techniques. In the literature, some of the proposed algorithms are evaluated assuming knowledge of some modulation parameters and using different scenarios, which makes difficult to compare the algorithms. In this thesis, algorithms representative of the prior art are compared with the proposed ones in a systematic and uniform manner, including non-ideal scenarios such as with phase jitter and frequency offset. The simulation results show that the proposed classification based on histograms is effective with a relatively low computational cost.

Sumário

Lista de Figuras	iv
Lista de Tabelas	vii
Lista de Abreviaturas	viii
1 Introdução	2
1.1 Elementos básicos de um sistema de comunicação digital	5
1.2 Motivações	7
1.3 Objetivos da Tese	7
1.4 Contribuições da Tese	8
1.5 Estrutura da Tese	9
2 Abordagens para Classificação Automática de Modulação	10
2.1 Pré-processamento do sinal RF	10
2.2 Modelo do sinal	13
2.3 Abordagem baseada no reconhecimento de padrão	15
2.3.1 Características baseadas no domínio de tempo instantâneo (DTI)	17
2.3.2 Características baseadas em transformações	20
2.3.3 Informações baseadas em estatísticas do sinal	21
2.3.3.1 Características baseadas em momentos de alta ordem (MAO)	21
2.3.3.2 Características baseadas em cumulantes de alta ordem (CAO)	22
2.3.3.3 Características baseadas na cicloestacionariedade	23
2.3.4 Características baseadas no formato da constelação	25
2.3.5 Informações baseadas na passagem por-zero (<i>Zero Crossing</i>)	25
2.3.6 Classificadores	26

2.3.6.1	Rede Neurais Artificiais	27
2.3.6.2	Máquina de Vetores de Suporte	27
2.3.6.3	Árvore de Decisão	28
2.4	Abordagem baseada na Teoria de Decisão	28
2.4.1	Um Limite Superior para a Probabilidade de Classificação Correta . . .	31
2.4.1.1	Teste Discreto da Razão entre Verossimilhanças	32
2.4.2	Teste de Kolmogorov-Smirnov	33
2.5	Extrator de parâmetros CSS (<i>Concatenated Sorted Symbols</i>)	35
2.6	Extrator de Parâmetros HISTO2D	37
2.7	Considerações Finais	42
3	Máquina de Vetores de Suporte - MVS	44
3.1	Introdução	44
3.2	Teoria do Aprendizado Estatístico	45
3.3	Máquina de Vetores de Suporte Linear com Margens Rígidas	48
3.4	Máquina de Vetores de Suporte Linear com Margens Suaves	53
3.5	Máquina de Vetores de Suporte Não Linear	55
3.6	Classificação Multi-classes	58
3.7	Custo Computacional	61
3.8	Considerações Finais	62
4	Sistema de CAM proposto e Avaliação de Desempenho	64
4.1	Proposta da Tese: Extrator de Parâmetros HISTO	64
4.2	Probabilidade de classificação correta	66
4.3	Descrição dos experimentos	67
4.4	Treinamento da MVS	68
4.5	Cenários	70
4.5.1	Classificação de modulação em canais AWGN	70
4.5.2	Classificação de modulação em canais com desvanecimento plano (<i>flat-fading</i>) e com <i>offset</i> de frequência	72
4.5.3	Classificação de modulação em canais com desvanecimento plano (<i>flat-fading</i>) com <i>jitter</i> de fase	75

4.5.4	Avaliação do custo computacional	76
4.6	Considerações Finais	78
5	Conclusões	80
5.1	Trabalhos Futuros	82
	Referências Bibliográficas	83

Lista de Figuras

1.1	Módulos de um sistema de Rádio Cognitivo.	4
1.2	Elementos básicos de um sistema de comunicação digital.	5
2.1	Sistema baseado em heterodina que converte o sinal em RF para FI para uma possível classificação de modulação.	11
2.2	Diagrama das Etapas de classificação de modulação usando reconhecimento de padrão.	16
2.3	Exemplos de espectros cíclicos. (a) Espectro cíclico do sinal BPSK. (b) Espectro cíclico do sinal QPSK.	25
2.4	Exemplos de diagramas de constelações para modulação 16-QAM e 8-PSK. (a) Diagrama de constelação para 16-QAM. (b) Diagrama de constelação para 8-PSK.	26
2.5	Exemplos de vetores com $D = 2N = 500$ elementos (250 magnitudes e 250 fases) para modulação 16-QAM e 8-PSK, sem ruído. (a) Diagrama de constelação para 16-QAM. (b) Diagrama de constelação para 8-PSK. (c) Amostras ordenadas e não ordenadas 16-QAM. (d) Amostras ordenadas e não ordenadas 8-PSK.	36
2.6	Comparação entre exemplos de dois vetores com $D = 2N = 500$ amostras (250 magnitudes e 250 fases) para comparação entre 16-QAM e 8-PSK, sem ruído.	37
2.7	Exemplos de vetores com $D = 2N = 500$ elementos (250 magnitudes e 250 fases) para modulação 16-QAM e 8-PSK, $RSR = 15$ dB. (a) Diagrama de constelação para 16-QAM. (b) Diagrama de constelação para 8-PSK. (c) Amostras ordenadas e não ordenadas 16-QAM. (d) Amostras ordenadas e não ordenadas 8-PSK.	38
2.8	Comparação entre exemplos de dois vetores com $D = 2N = 500$ amostras (250 magnitudes e 250 fases) para comparação entre 16-QAM e 8-PSK, $RSR = 15dB$	39

2.9	Exemplos de histogramas 2D com $N = 250$ símbolos (250 magnitudes e 250 fases) para modulações BPSK, 4-PAM, 16-QAM e 8-PSK, sem ruído. (a), (b), (c) e (d).	40
2.10	Exemplos de histogramas 2D com $N = 250$ símbolos (250 magnitudes e 250 fases) para modulações BPSK, 4-PAM, 16-QAM e 8-PSK, com RSR =15 dB. (a), (b), (c) e (d).	41
3.1	Hipóteses de hiperplano com hiperplano ótimo	46
3.2	Separação dos dados de treino por meio de reta.	47
3.3	Separação dos 4 dados de treino por meio de reta.	48
3.4	Identificação da margem e vetores de suporte.	49
3.5	Hiperplano separando padrões linearmente separáveis.	49
3.6	Exemplo de conjunto de dados linearmente não separável	50
3.7	Exemplo de conjunto de dados linearmente não separável	53
3.8	Exemplo de conjunto de dados linearmente não separável Equação 3.12el	54
3.9	Exemplo de padrão não linear	56
3.10	Utilização de uma função não linear sobre os dados de entrada	56
3.11	Técnica Um Contra Todos.	59
3.12	Técnica Um contra um.	60
4.1	Exemplos de histogramas com $N = 250$ símbolos (250 magnitudes e 250 fases) para modulações BPSK, 4-PAM, 16-QAM e 8-PSK, sem ruído. (a), (c), (e) e (g) Constelações. (b), (d), (f) e (h) Histogramas	65
4.2	Análise de complexidade amostral.	69
4.3	Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_1 . Cenário AWGN com faixa de RSR $[-5, 15]$ dB.	71
4.4	Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_2 . Cenário AWGN com faixa de RSR $[0, 20]$ dB.	72
4.5	Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_1 obtidos a partir do canal com desvanecimento plano e <i>offset</i> de frequência com faixa de RSR $[0, 25]$ dB.	73

4.6	Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_2 obtidos a partir do canal com desvanecimento plano e <i>offset</i> de frequência com faixa de RSR $[0, 25]$ dB.	74
4.7	Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_1 obtidos a partir do canal com desvanecimento plano e <i>phase jitter</i> com faixa de RSR $[0, 25]$ dB.	75
4.8	Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_2 obtidos a partir do canal com desvanecimento plano e <i>phase jitter</i> com faixa de RSR $[0, 25]$ dB.	76

Lista de Tabelas

2.1	Valores de cumulantes de um canal ideal para vários tipos de modulações. . . .	24
2.2	Resumo dos métodos CAM.	29
2.3	Trabalhos baseados em verossimilhança.	34
3.1	Pesos associados com operações matemáticas.	62
4.1	Exemplo de matriz de confusão.	67
4.2	Exemplo de ICC for HISTO-MVS e CSS-MVS.	77
4.3	Utilização de recursos do FPGA ALTERA na implementação do CSS-MVS e HISTO-MVS utilizando kernel linear.	78

Lista de Abreviaturas

ASK *Amplitude Shift keying*

ALRT *Average Likelihood Ratio Test*

Anatel Agência Nacional de Telecomunicações

ANN *Artificial Neural Network*

AWGN *Additive White Gaussian Noise*

BPSK *Binary Phase Shift Keying*

CAM Classificação Automática de Modulação

CAO Cumulante de Alta Ordem

CDF *Cumulative Distribution Function*

CSS *Concatenated Sorted Symbols*

DEC Densidade Espectral Cíclica

DLRT *Discrete Likelihood Ratio Test*

DDC *Digital Down-Converter*

DSB *Double Side Band*

DTI Domínio do Tempo Instantâneo

FAE Função de AutoCorência Espectral

FAC Função de AutoCorrelação Cíclica

FCC *Federal Communications Commission*

FDP *Função Densidade de Probabilidade*

FDAE *Função de Distribuição Acumulada Empírica.*

FI *Frequência Intermediária*

FM *Frequency Modulation*

FSA *Fixed Spectrum Allocation*

FSK *Frequency Shift Keying*

FPGA *Field Programmable Gate Array*

GLRT *generalised likelihood ratio test*

ICC *Índice de Complexidade Computacional*

JPEG *Joint Photographic Experts Group*

IEEE *Instituto de Engenheiros Eletrônicos e Eletricistas*

KNN *k-Nearest Neighbor*

KKT *Karush-Kuhn-Tucker*

LNA *emphLow Noise Amplifier*

LO *emphLocal Oscillator*

LUT *Look-Up Table*

MAO *Momento de Alta Ordem*

MLP *Multilayer Perceptron*

MPEG *Moving Picture Experts Group*

MVS *Máquina de Vetores de Suporte*

NCO *Numerically Controlled Oscillator*

JPEG *Joint Photographic Experts Group*

HLRT *Hybrid Likelihood Ratio Test*

PC Perfil Cíclico

PCC Probabilidade de Classificação Correta

PDS Processamento Digital de Sinais

PSK *Phase Shift Keying*

QAM *Quadrature Amplitude Modulation*

RBF *Radial Basis Function*

RC Rádio Cognitivo

RF Rádio Frequência

RSR Razão Sinal-Ruído

RDS Rádio Definido por Software

SVM *Support Vectors Machine*

SSB *Single Side-Band*

TAE Teoria de Aprendizado Estatístico

USB *Upper Side Band*

Capítulo 1

Introdução

Classificação automática de modulação (CAM) possui suas raízes históricas pautadas em cenários militares, onde a guerra eletrônica, vigilância e análise de ameaças requerem o reconhecimento das modulações de sinais com o objetivo de identificar as unidades de transmissão do adversário de modo a bloquear e interceptar as informações transmitidas.

Dentro deste contexto militar, a CAM se torna uma ferramenta primordial no campo da guerra eletrônica, a qual é composta por três componentes essenciais: suporte eletrônico (*electronic support*), ataque eletrônico (*electronic attack*) e proteção eletrônica (*electronic protect*) [1]. Através da CAM, o componente suporte eletrônico consegue interceptar informações transmitidas via rádio-frequência após a detecção do sinal transmitido. A descoberta da modulação no sinal transmitido nesse componente se estende para todos os componentes de um sistema de guerra eletrônica. Para se ter uma ideia da importância da classificação de modulação nesse componente, a descoberta da modulação pode ser usada para demodular o sinal interceptado a fim de recuperar a mensagem transmitida entre as unidades adversárias. Além disso, a informação de modulação por si só ajuda a cartografia eletrônica cujo papel é identificar as unidades de comunicações e suas possíveis localizações.

Em relação ao componente ataque eletrônico, essa parte do sistema de guerra eletrônica tem como objetivo evitar a comunicação entre as bases inimigas através de uma medida denominada de bloqueio (*jamming*). Há muitas técnicas de bloqueio disponíveis. A mais comum se baseia na implantação de um bloqueador (*jammer*) no canal de comunicação entre as bases do adversário e também a transmissão de sinais ruidosos ou sinais gerados usando o tipo de modulação detectado pelo classificador de modulação.

Sabe-se que neste cenário de guerra, além de ter o conhecimento das informações transmitidas pelo inimigo, também é muito importante proteger as informações transmitidas entre as Forças aliadas a despeito de formas de interferência intencionais e das tentativas de ataque

eletrônico oriundos de Forças inimigas. Tal requisito é o principal objetivo do componente proteção eletrônica. De acordo com [1], uma estratégia efetiva de proteção é monitorar, através de um dispositivo [2], a modulação do sinal de bloqueio do inimigo, a fim de que a informação transmitida entre as Forças aliadas possam ter um tipo de modulação distinto do sinal de bloqueio.

Com o passar dos anos, juntamente com a evolução tecnológica, nasce uma nova motivação para CAM, baseada no crescimento do uso do espectro de frequência. Todo esse crescimento implica em uma consequência: o espectro de frequência tornou-se um recurso muito escasso. Além da escassez física, outro problema é a forma de regulamentação que define a exclusividade no uso de faixa de espectro para uma determinada tecnologia, de modo que limita que potenciais usuários possam obter acesso a tal recurso quando não utilizado [3].

A *Federal Communications Commission* (FCC), órgão que regulamenta o uso do espectro de frequência nos Estados Unidos, realizou um estudo sobre a variação do uso desse espectro. Nesse estudo foi constatado que o espectro de frequência sofre uma variação de utilização entre 15% e 85% [4], ou seja, esse recurso é subutilizado e novas tecnologias são impedidas de usá-lo devido à política de regulamentação. Desta forma, detectou-se a necessidade de novos métodos com a finalidade de melhorar a utilização do espectro licenciado, especialmente porque há uma escassez de faixas livres para novos canais.

Sob esse cenário, o Rádio Cognitivo (RC) [5] apareceu como uma solução promissora, pois uma de suas principais características é a questão da adaptabilidade dos parâmetros de rádio (frequência, potência e largura de banda). Estes podem ser modificados dependendo das características do ambiente do rádio, condições de rede, geolocalização e etc. Além disso, tal dispositivo possui capacidade de prover técnicas de acesso dinâmico via arquiteturas de redes sem fio. Baseado nessas características, RC é definido formalmente como uma tecnologia de comunicação sem fio que interage com ambiente de operação de forma inteligente, a qual busca dois objetivos primários: acesso oportunista e utilização eficiente do espectro. Tal tecnologia foi sugerida originalmente em [5] e vem sendo bastante investigada desde então.

Uma das aplicações do RC é no reuso dos canais de televisão de forma dinâmica. Nesse tipo de aplicação, o RC permite que espectros não ocupados sejam utilizados sem interferência às operações licenciadas, ou seja, os usuários primários (usuários habilitados a usar a faixa de frequência) não podem sofrer interferência. Desta maneira, quando um usuário primário é detectado, o usuário secundário (usuários não licenciados que ocasionalmente utilizam uma faixa de frequência licenciada não ocupada) que utiliza a tecnologia de RC precisa liberar a faixa de frequência ocupada no momento. Uma vez feito esse procedimento, a tecnologia busca outra porção de espectro vazio para o usuário secundário, a fim de retomar o processo

de comunicação.

O IEEE 802.22 [6] é o primeiro padrão que está sendo desenvolvido pelo Instituto dos Engenheiros Eletricistas e Eletrônicos (IEEE) e que utiliza conceitos de RC. O objetivo deste novo padrão é o reuso de bandas subutilizadas da faixa de TV UHF/VHF entre 54 MHz e 862 MHz em áreas rurais.

A fim de suportar a capacidade de modificar parâmetros de transmissão e recepção de acordo com as condições ambientais no qual se encontra, em resumo, as principais funcionalidades do RC podem ser vistas na Figura 1.1.

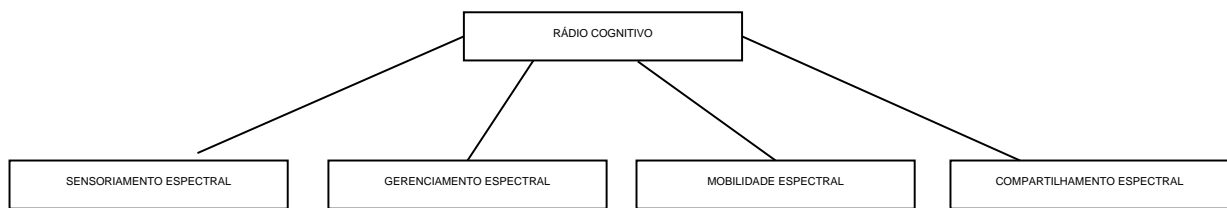


Figura 1.1: Módulos de um sistema de Rádio Cognitivo.

O sensoriamento espectral (*Spectrum Sensing*) tem o objetivo de detectar lacunas no espectro e compartilhá-los sem interferência para outros usuários. Já o gerenciamento espectral (*Spectrum management*) tem como objetivo fornecer, em tempo real, a melhor banda e a melhor frequência para o serviço desejado. Diferentemente dos outros dois, a mobilidade espectral (*Spectrum Mobility*) possui a responsabilidade de assegurar a comunicação do usuário durante a transição para a melhor banda espectro. Por último, o compartilhamento Espectral (*Spectrum Sharing*) busca fornecer um método de escalonamento de espectro, o qual seja justo entre usuários co-existentes [7].

Dentre esses módulos, o de sensoriamento espectral pode se beneficiar da classificação de automática modulação (CAM) [8]. Como citado, o processo de CAM é executado no receptor, baseado no sinal recebido antes da demodulação quando o esquema de modulação é desconhecido, ou seja, a classificação de modulação busca descobrir na recepção se um sinal é, por exemplo, ASK (*Amplitude Shift Keying*), PSK (*Phase Shift Keying*), QAM (*Quadrature Amplitude Modulation*), dentre outros. Uma vez tendo o conhecimento do esquema de modulação aplicado no sinal recebido, um sistema RC pode trabalhar nos parâmetros de configuração a fim de que não haja interferência na informação de modulação de um outro sistema de RC que esteja operando em um canal adjacente. Porém, vale ressaltar que a prioridade do canal é sempre do usuário licenciado para aquela determinada faixa de frequência que esteja sendo ocupada pelo RC. Caso as alterações de parâmetros não forem suficientes para diminuir o nível de interferência com outros usuários, o RC deve trocar de faixa. Há possibilidade de concepção de RCs mais simples, que não usem CAM, mas a informação sobre a modulação é

de valia para prover algumas funcionalidades ao RC [3]. Por isso, RC tornou-se, juntamente com as aplicações militares, outra área de aplicação de CAM.

1.1 Elementos básicos de um sistema de comunicação digital

A Figura 1.2 ilustra o diagrama funcional e os elementos básicos de um sistema de comunicação digital [9]. A parte superior da Figura 1.2 apresenta os blocos que constituem o transmissor. O primeiro bloco do transmissor é denominado de fonte da informação cuja saída pode ser tanto um sinal analógico, tais como um sinal de vídeo ou de áudio, ou um sinal digital. Por exemplo, a saída de um computador cuja característica do sinal no tempo é discreto e possui um número finito de caracteres de saída. É importante notar que grande parte da informação gerada no primeiro bloco é caracterizada por informação redundante, contribuindo assim no aumento da largura de banda. A fim de evitar esse problema, o segundo bloco denominado de codificação de fonte possui a função de retirar a redundância da informação oriunda da saída do primeiro bloco, caracterizando um processo de compressão digital. Um sinal de voz, por exemplo, cerca de 50% deste sinal é silêncio, ou seja, não há informação. Exemplos de técnica de compressão da fonte são MPEG (*Moving Picture Experts Group*) e JPEG (*Joint Photographic Experts Group*). Uma vez que no bloco de codificação da fonte, a informação sofre um processo de compressão de dados, a taxa de dados de saída é menor em relação a taxa de dados de entrada o que significa uma redução da largura de banda utilizada.

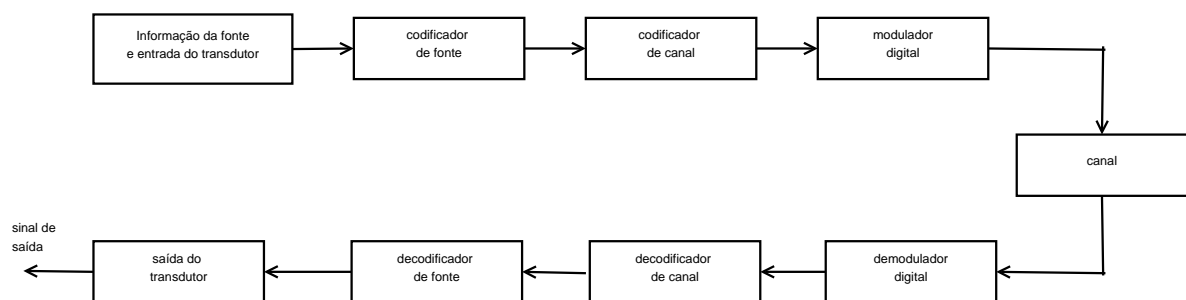


Figura 1.2: Elementos básicos de um sistema de comunicação digital.

O trem de bits oriundos do bloco de codificação de fonte é passado ao codificador de canal cuja a finalidade é introduzir, de maneira controlada, redundância na sequência de informação binária. A inclusão de redundância de maneira controlada tem como consequência a diminuição da taxa de transmissão de dados e também o aumento da largura de banda

do canal comparado ao sistema não codificado. É importante observar que tal inserção de informação pode ser utilizada no receptor a fim de tratar a interferência e os efeitos dos ruídos adicionados ao sinal durante a transmissão através do canal. Desta forma, a adição de redundância tem o objetivo de aumentar a confiabilidade dos dados recebidos, ou seja, ter maior certeza que os dados transmitidos e recebidos são os mesmos. Uma maneira trivial de realizar esse procedimento é repetir cada dígito binário m vezes, onde m é um valor positivo.

A sequência binária na saída do bloco de codificação de canal é passada ao bloco modulador digital, o qual serve como uma interface ao canal de comunicação, ou seja, tal bloco cria formas de onda que possam ser transmitidas pelo canal a fim de serem devidamente interpretadas no receptor. Assim, o objetivo principal do modulador digital é mapear a sequência de informação binária em formas de onda de sinal. Para realizar essa tarefa, a sequência de informação codificada deve ser dividida em grupos de modo que possa ser transmitida numa taxa R bits por segundo (bits/s). Por exemplo, o modulador digital pode simplesmente mapear o dígito binário 0 em uma forma de onda $s_0(t)$ e o dígito binário 1 em uma forma de onda $s_1(t)$. Desta maneira, cada bit oriundo do codificador do canal é transmitido separadamente. Esse procedimento é denominado de modulação binária. Alternativamente, um modulador pode transmitir b bits de informação codificada em um tempo usando $M = 2^b$ formas de ondas $s_i(t), i = 0, 1, \dots, M - 1$ distintas, ou seja, uma forma de onda para cada 2^b sequências possíveis de b -bits. Essa maneira alternativa é denominada de modulação M -ary ($M > 2$) [10]. Por último, deve-se notar que o modulador digital no transmissor não adiciona nem reduz a informação presente na entrada do bloco.

As M formas de ondas oriundas do modulador digital são passadas ao canal de comunicação, que representa o meio físico usado para enviar a informação do transmissor ao receptor. Há diferentes tipos de canal: ar, par trançado, fibra óptica, linha telefônica e etc. O canal de comunicação é responsável pela degradação do sinal através de algumas perturbações, por exemplo: atenuação (perda de potência), ruído, multi-percurso, interferência, distorção e retardo.

Uma vez descrito os blocos do processo de transmissão, analisa-se os blocos do receptor. Na recepção de um sistema de comunicação digital, o demodulador digital processa o sinal transmitido e corrompido pelo canal e o reduz para uma sequência de números que representam a estimação dos símbolos transmitidos (binários ou M -ary). Tal sequência de números é passada ao decodificador de canal, o qual tenta reconstituir a sequência original de informações a partir do código usado e a redundância contida nos dados recebidos.

A frequência dos erros ocorridos na sequência de decodificação representa uma medida que ajuda avaliar se o demodulador e o decodificador estão funcionando bem. Em outras

palavras, a probabilidade média de bit de erro na saída do decodificador representa uma medida de desempenho da combinação demodulador-decodificador. Em geral, a probabilidade de erro é uma função das características do código, dos tipos de formas de ondas utilizadas para transmitir a informação ao longo do canal, da potência da transmissão, das características do canal (por exemplo, a quantidade de ruído, a natureza da interferência), e dos métodos de demodulação e decodificação.

Como passo final, quando uma saída analógica está correta, o decodificador da fonte aceita a saída da sequência oriunda do decodificador do canal e, a partir do conhecimento do método aplicado na codificação de fonte tenta reconstituir o sinal original transmitido da fonte. Porém, devido aos erros de codificação de canal e possíveis distorções introduzidas pelo codificador de fonte, o sinal na saída do decodificador de fonte seja uma aproximação da saída da fonte do transmissor. A diferença entre o sinal original e o sinal reconstruído é uma medida de distorção introduzida pelo sistema de comunicação digital.

1.2 Motivações

Durante a revisão bibliográfica deste trabalho, percebeu-se que a maioria dos trabalhos sobre CAM, as propostas de classificação dependem do conhecimento de algumas informações, por exemplo, o valor da relação sinal-ruído a fim de que a classificação seja realizada. O conhecimento dessas informações torna os cenários menos desafiadores comparados a ambiente reais, onde o classificador não as possui. Desta forma, deve-se propor cenários de classificação mais desafiadores no sentido de avaliar a performance dos classificadores.

Além disso, poucos trabalhos são comparados com outras metodologias de classificação. Tal fato, torna a análise da eficiência dos resultados obtidos complicada, pois não se sabe o quanto estão próximos ou distantes dos resultados considerados como *baseline*.

Por fim, muitos trabalhos preocupam-se com a eficiência dos resultados das propostas, porém deixam de analisar o custo computacional. A análise do custo computacional é uma informação primordial na implementação dos classificadores em hardware, pois pode-se verificar se a proposta é viável em aplicações em tempo real.

1.3 Objetivos da Tese

A proposta desta Tese tem como objetivos realizar uma investigação de algoritmos de CAM baseados em aprendizagem discriminativa com objetivo de propor uma nova técnica

de classificação, a qual possa apresentar um custo computacional menor comparado as da literatura.

Além disso, analisar os resultados obtidos da proposta com outras metodologias de classificação, por exemplo, ALRT (*Average Likelihood Ratio Test*) cujos resultados são considerados como *baseline* para fins de comparação.

Ademais, formular cenários aos classificadores, os quais possam desconhecer algumas variáveis a fim de comparar os resultados de classificação em relação ao não conhecimento e conhecimento dessas informações. Tal objetivo tem como finalidade verificar a robustez do processo de classificação quanto ao desconhecimento de informações, as quais geralmente precisam ser estimadas para que haja a classificação de modulação.

1.4 Contribuições da Tese

A primeira contribuição deste trabalho é baseada na formulação de um novo extrator de parâmetros para CAM chamado HISTO, caracterizado pelo cálculo dos histogramas de magnitude e fase dos símbolos recebidos e, posteriormente, concatenação dos mesmos. O HISTO é proposta original e exclusiva do autor desta Tese. Os distintos valores de magnitude e fase de cada modulação, quando mapeados por meio de histogramas apresentam uma assinatura que permite distinguir uma modulação da outra, como mostram os resultados a serem apresentados. A partir da aplicação deste estudo, verificou-se que a proposta de classificação apresentou resultados bons quando comparados com outros trabalhos, e com um custo computacional muito menor, o qual pode ser comprovado na publicação do seguinte trabalho [11].

A avaliação do custo computacional é relegada a um segundo plano em muitas publicações na literatura. Contudo, a mesma é de suma importância para a viabilidade do algoritmo, em especial se o equipamento precisa ter, por exemplo, baixo consumo de potência. Assim, outra contribuição desta Tese é a análise do custo computacional de cada extrator de parâmetros baseada no índice de complexidade computacional (ICC). Tal análise informa o custo computacional de classificação de cada classificador baseado nos pesos atribuídos a cada operação matemática executada. Os resultados da análise são confirmados através da quantidade de recursos necessários na implementação dos classificadores em FPGA (*Field-Programmable Gate Array*). O foco deste trabalho não foi a implementação em FPGA, mas a parceria com pesquisadores atuando com FPGA que permitiu mapear os custos dos algoritmos propostos em FPGA.

Uma terceira contribuição também vem da atuação em parceria, desta vez com grupo da

Universidade Federal de Santa Catarina (UFSC). O resultado desta cooperação foi a concepção de um novo extrator de parâmetros para CAM baseado em um histograma 2D cujos resultados estão no trabalho [12] em avaliação para publicação. O autor desta Tese participou das discussões sobre o HISTO2D, forneceu todo código implementado durante sua pesquisa, e atuou na escrita de algumas seções do artigo. Em contrapartida, o grupo da UFSC originou a questão e ficou responsável pela implementação do classificador em FPGA, além de contribuir na escrita do artigo e outras atividades inerentes da cooperação. Este novo método será brevemente detalhado nesta Tese.

Além dessas contribuições, ressalta-se que o autor da Tese atuou na formulação de um extrator de parâmetros para CAM baseado na ordenação e concatenação dos valores de magnitude e fase dos símbolos recebidos chamado CSS. O autor participou das discussões e simulações, que redundaram em um trabalho conjunto do coautor e autor do trabalho [13]. No citado trabalho mostrou-se que a ordenação dos valores de magnitude e fase permite distinguir uma modulação da outra.

Por fim, esta Tese também contribuiu em melhor entendimento do uso de máquinas de vetores de suporte aplicadas a CAM. Apesar das mesmas já terem sido adotadas anteriormente neste problema, esta Tese é inovadora ao contrastar as MVS lineares com as não-lineares, mostrando que o custo computacional das lineares pode ser bastante reduzido. Tal fato não havia sido explorado anteriormente na literatura.

1.5 Estrutura da Tese

O restante do trabalho está organizado da seguinte forma:

- Capítulo 2 - descreve o modelo de sinal tanto quanto as abordagens relacionadas à CAM. Por último, descreve-se a concepção de duas propostas de extratores de parâmetros para classificação oriundas da cooperação com outros pesquisadores;
- Capítulo 3 - descreve a teoria sobre Máquinas de Vetores de Suporte (MVS), algoritmo escolhido para realizar a tarefa de classificação. Além disso, esse capítulo descreve a análise do custo computacional baseado no ICC;
- Capítulo 4 - apresenta a principal contribuição desta Tese baseada na concepção do extrator de parâmetros HISTO. Em seguida, a análise do desempenho do HISTO é realizada juntamente com outros algoritmos de classificação;
- Capítulo 5 - apresenta as considerações finais.

Capítulo 2

Abordagens para Classificação

Automática de Modulação

Neste capítulo, métodos de classificação de modulação são discutidos, os quais são definidos em dois grupos: o primeiro método baseado em verossimilhança e o segundo em características (*features*). Por motivos históricos, esses métodos são descritos e estudados de acordo com a área que inspirou o respectivo formalismo matemático: teoria da detecção/estimação [14] e reconhecimento de padrões [15,16]. Os métodos baseados na teoria de detecção utilizam teste de hipóteses para formular um problema de classificação [17,18]. Já os métodos baseados no reconhecimento de padrão podem ser divididos em dois subsistemas: extrator de características e o classificador [13,19]. Porém, antes de iniciar uma análise sobre os extratores de características, é fundamental descrever o processamento de sinais que precede a utilização dos mesmos.

2.1 Pré-processamento do sinal RF

A fim de se obter uma visualização do processamento de um sinal RF durante a tarefa de classificação de modulação, o diagrama abaixo ilustra os principais componentes que constituem essa tarefa:

$r(t)$ (sinal em RF) \rightarrow extrator de parâmetros $\rightarrow \mathbf{x}$ (características) \rightarrow classificador $\rightarrow y$ (classe)

os quais são representados pela extração de parâmetros juntamente com o bloco classificador. Na entrada do bloco de extrator de parâmetros, tem-se o sinal $r(t)$ em RF (por exemplo, em uma frequência de 1,8 GHz). Ao passar pelo bloco de extrator de parâmetros, $r(t)$ é convertido

para um vetor de D parâmetros $\mathbf{x} = \{x_d\}_{d=1}^D$. Realizada a conversão, o novo vetor \mathbf{x} oriundo do extrator de parâmetros é processado no classificador que decide a classe y dentre as Y classes possíveis.

Esse trabalho concentrou-se no processamento do sinal após sua digitalização. Entretanto, é fundamental descrever uma arquitetura que represente uma cadeia de blocos que produza um sinal digital a partir de um sinal analógico em rádio-frequência, de modo que mais adiante descreva-se o problema de classificação de modulação.

Os conversores analógicos/digitais (A/D) podem trabalhar com sinais tanto na banda base quanto com sinais cujo espectro é centrado na frequência intermediária (FI), por exemplo, 70 MHz. O princípio moderno de rádio definido por software ambiciona posicionar o conversor A/D próximo à antena, de modo a digitalizar diretamente o sinal em RF, por exemplo, em 1,8 GHz. Desta maneira, o sinal em RF já seria processado com técnicas de Processamento Digital de Sinais (PDS). Todavia, com a tecnologia existente, é usualmente mais comum utilizar uma heterodina e digitalizar o sinal em uma frequência-intermediária (FI) ou usar conversão direta (*direct conversion*) e digitalizar o sinal em banda base após a conversão para baixas frequências.

A parte prática deste trabalho foi implementada de modo que o sistema possa ser usado em conjunto com a USRP (*Universal Software Radio Peripheral*), admitindo, dessa forma, o uso de uma heterodina, como visto na Figura 2.1. Os testes com USRP estão fora do escopo desta Tese, mas como a mesma se insere no contexto de uma pesquisa do grupo, houve preocupação com a viabilidade de implementação do algoritmos e outros aspectos práticos.

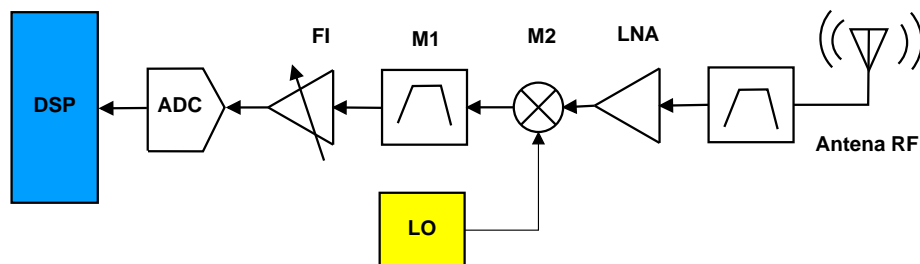


Figura 2.1: Sistema baseado em heterodina que converte o sinal em RF para FI para uma possível classificação de modulação.

Dentre as várias possibilidades, assume-se o seguinte cenário. Analisando-se a Figura 2.1, percebe-se que um sinal em RF é captado através da antena, e posteriormente passa por um filtro passa-banda a fim de reduzir o ruído. Após passar por um LNA (*low noise amplifier*), um misturador multiplica o sinal com uma portadora gerada por um LO (*local oscillator*.) O filtro passa-banda M1 é seguido por um amplificador e o sinal é convertido

via um chip ADC. A partir disso, já no domínio digital, o sinal passa por um processamento digital (DSP - *Digital Signal Processing*): um DDC (*digital down-converter*) posiciona o espectro do sinal aproximadamente (na prática, há um *offset* em frequência) em banda base (ou frequência DC). Assume-se que o LO pode ser programado de maneira que permita varrer a faixa de frequência de interesse.

Por exemplo, se um ADC de frequência de amostragem $f_s = 96$ MHz é usado para digitalizar sinais com banda de 20 MHz, o LO sintoniza na banda de 20 MHz desejada. O filtro M1 se encarrega de limitar a banda em 20 MHz, nesse caso, e o LO de “sintonizar” o sistema de tal forma que a frequência central dessa banda seja a desejada. O DDC faz posteriormente o “ajuste fino”. Continuando o exemplo, assume-se que o classificador de modulação deva atuar varrendo a banda na faixa de 800 MHz a 900 MHz. Nesse caso, o LO pode ser programado para um dado intervalo de tempo de modo que ADC digitalize o sinal na banda, por exemplo, de 800 a 820 MHz a fim de que seja realizado o processo de classificação de modulação. Depois o LO seria reprogramado para que a faixa em observação movesse para 820 a 840 MHz e assim por diante.

A fim de controlar a varredura do sinal digitalizado pode-se usar um algoritmo de detecção baseado em energia [20]. Continuando com o exemplo anterior, a banda original de 20 MHz pode ser fragmentada de modo a identificar quais sub-faixas de 2 MHz devem ser investigadas por apresentarem uma energia média relativamente alta comparada ao ruído de fundo. Uma vez escolhida a faixa, sua frequência central pode ser adotada como a frequência f_c da possível portadora sinal de interesse $\mathbf{r} = \{r_n\}_n^{N-1}$ a ser classificado. A banda de sinal útil em torno da portadora é um limite superior à taxa de símbolos do sinal, como indicado pelo Critério de Nyquist de modo que não haja interferência entre os símbolos [21]. Desta forma, assume-se que \mathbf{r} seja o sinal reposicionado aproximadamente em banda base. No DDC, um NCO (*numerically-controlled oscillator*) pode ser usado para centralizar o espectro de \mathbf{r} no nível DC. Após isso, \mathbf{r} é repassado ao bloco de extrator de parâmetros que o converte em um vetor \mathbf{x} de parâmetros, em seguida, entra no bloco de classificador. Nota-se que \mathbf{r} tem uma banda no máximo igual à banda adotada para cada sub-faixa.

Com base no que foi discutido nos parágrafos anteriores, um extrator de parâmetros pode ser modelado como:

$$r(t) \rightarrow \boxed{\text{FILT/DOWN}} \rightarrow c(t) \rightarrow \boxed{\text{A/D}} \rightarrow \mathbf{c} \rightarrow \boxed{\text{PDS}} \rightarrow \mathbf{r} \rightarrow \boxed{\text{extraí parâmetros}} \rightarrow \mathbf{x}$$

onde FILT/DOWN denota operações como filtragem, *down-conversion* e amplificação, $c(t)$ é o sinal de entrada do A/D, \mathbf{r} é obtido a partir de $\mathbf{c} = \{c_n\}_n^{N-1}$ através de processamento digital de sinais (por exemplo, *down-conversion usando NCO*) e o último bloco extraí parâmetros, consistindo na etapa na qual esta Tese irá se concentrar. Ressalta-se que o sinal \mathbf{r} é carac-

terizado por possuir uma taxa de amostragem $f_2 \in [R_{sym}, f_s]$ no intervalo entre a taxa de símbolos R_{sym} e a frequência f_s de amostragem do A/D.

Realizando-se uma comparação com o processo de classificação usando extração de parâmetros visto anteriormente, o modelo de classificação baseado na verossimilhança não utiliza o último bloco do diagrama anterior, ou seja, o sinal $\mathbf{r} = \mathbf{x}$. Desta forma, o processo de classificação baseado em verossimilhança é baseado no diagrama abaixo:

$$r(t) \rightarrow \boxed{\text{FILT/DOWN}} \rightarrow c(t) \rightarrow \boxed{\text{A/D}} \rightarrow \mathbf{c} \rightarrow \boxed{\text{PDS}} \rightarrow \mathbf{r} = \mathbf{x},$$

onde \mathbf{r} é tipicamente o envelope complexo [21] com taxa de amostragem R_{sym} . Nesse caso, o bloco PDS realiza o processamento e re-amostragem de \mathbf{c} para uma taxa de amostragem R_{sym} que coincide com a taxa de sinalização (em bauds). Esse tipo de modelo é bastante usado na literatura.

2.2 Modelo do sinal

A finalidade de um sistema de comunicação sem fio é transmitir sinais com informação por meio de um canal até o receptor [22,23]. Considerando o sinal ser propagado por um canal apresentando ruído qualquer $\mathcal{V}(t)$, a expressão do sinal recebido na banda base é caracterizada por

$$r(t) = \mathcal{C}(t) + \mathcal{V}(t), \quad (2.1)$$

onde $\mathcal{C}(t)$ é o envelope complexo do sinal modulado cujas características são constituídas por símbolos transmitidos s_n^y do y -ésimo tipo de modulação no tempo nT_s e a função pulso conformador $p_{pulso}(t)$, avaliada em detalhes em [24]. Para N símbolos com periodicidade T_s , a forma geral da expressão do envelope complexo no receptor é:

$$\mathcal{C}(t, \mathbf{u}_y) = a_c e^{j(2\pi f_0 t + \theta_c)} \sum_{n=0}^{N-1} s_n^{y,k} e^{j\phi_n} g(t - nT_s - \varepsilon T_s), \quad (2.2)$$

onde a notação $\mathcal{C}(t, \mathbf{u}_y)$ foi admitida neste trabalho com a finalidade de enfatizar a dependência do sinal em relação a parâmetros desconhecidos, de modo que, \mathbf{u}_y é um vetor constituído por esses parâmetros sob y -ésimo tipo de modulação

$$\mathbf{u}_y = \{a_c, \theta_c, \varepsilon, h(t), \{s_n^{y,k}\}_{k=1}^{M_y}, f_0, \{\phi_n\}_{n=0}^{N-1}\}. \quad (2.3)$$

Os símbolos usados nas Equações 2.2 e 2.3 são:

1. $g(t)$ representa o efeito composto entre a resposta impulso do canal denotado como $h(t)$ juntamente com o filtro conformador $p_{pulso}(t)$. Esse efeito é o resultado da saída da convolução $g(t) = h(t) * p_{pulso}(t)$;

2. $T_s = 1/R_{sym}$ representa o período do símbolo;
3. N número de símbolos para cada bloco;
4. f_0 representa o *offset* de frequência;
5. a_c representa o ganho introduzido pelo canal, que em uma situação de *flat-fading* é um mesmo valor para todos os N símbolos do bloco;
6. ϕ_n representa *jitter* de fase;
7. θ_c fase da portadora invariante no tempo;
8. ε o *offset* de tempo normalizado entre o transmissor e o receptor;
9. $s^{y,k}$ vetor de símbolos pertencendo ao y -ésimo tipo de modulação, os quais são assumidos serem independentes e uniformemente distribuídos. Cada modulação y possui uma constelação com k símbolos conhecidos pertencendo ao alfabeto M_y . Durante o n -ésimo intervalo de tempo, o símbolo transmitido é denotado por $s_n^{y,k}$. A potência média de cada símbolo é usualmente normalizada para uma unidade de maneira que:

$$\frac{1}{M_y} \sum_{k=1}^{M_y} \|s^{y,k}\|^2 = 1. \quad (2.4)$$

De acordo com a Equação 2.2, pode-se verificar que vários parâmetros precisam ser estimados de modo que haja uma recepção dos dados com sucesso. Desta forma, pode-se entender a complexidade de um receptor (Rádio Cognitivo) em um cenário não cooperativo, onde tal dispositivo não possui conhecimento a respeito desses parâmetros, de maneira que existe a necessidade de estimá-los. Devido esse fato, há uma variedade de trabalhos sobre CAM com cenários distintos e diversas propostas de classificação as quais, para realizar a tarefa de classificação, dependem apenas do conhecimento de alguns ou nenhum destes parâmetros. Por exemplo, o conhecimento da RSR é necessário para realizar CAM em [18, 25, 26].

Neste trabalho, o modelo do sinal usado na tarefa de CAM é a sequência do sinal recebido em banda base na saída do filtro de recepção (não se pode usar um filtro casado pois o receptor não conhece $p_{pulso}(t)$), cujas amostras são definidas como em [27, 28]:

$$r_n = a_c e^{j2\pi f_0 T_s n + j\phi_n} s_n^{(y)} + v_n, \quad (2.5)$$

onde

1. r_n é o n -ésimo símbolo recebido;

2. a_c representa o ganho introduzido pelo canal;
3. o *offset* de fase ϕ_n é uma variável aleatória que muda de símbolo para símbolo. Note que o valor de ϕ_n na Equação 2.5 não corresponde exatamente ao valor numérico de ϕ_n na Equação 2.2. Na Equação 2.5, esse valor incorpora a fase da portadora θ_c e de $g(t)$;
4. de forma similar, f_0 é o *offset* em frequência, modelado como uma variável aleatória constante para cada realização de $\mathbf{r} = \{r_n\}_{n=1}^N$;
5. T_s é o período do símbolo.

O modelo assumido na Equação 2.5 é mais realista do que o modelo AWGN (*additive white Gaussian noise*) [21] pois inclui erro na frequência f_0 e na fase θ_n , além do ganho a_c correspondente a uma atenuação constante (*flat fading*). Contudo, a Equação 2.5 não lida explicitamente com o erro de sincronismo modelado por ε na Equação 2.2. Pode-se entender então que, na Equação 2.5, assume-se que já foi realizado o sincronismo através de alguma das técnicas existentes na literatura [21]. Como esta tarefa não é trivial [29], assumir sincronismo é uma fragilidade do modelo, mas o mesmo foi adotado por ser extensivamente usado na literatura de CAM [27, 28].

Por último, assume-se também que os símbolos $s_n^{(y)}$ são independentes e identicamente distribuídos (i. i. d.), com valores extraídos de acordo com o alfabeto do y -ésimo tipo de modulação. Isto é razoável pois a maioria dos sistemas usa símbolos i. i. d.

Em algumas situações, o modelo assumido na Equação 2.5 é simplificado, assumindo-se que os *offsets* ϕ_n e f_0 são nulos. Ou mesmo um modelo AWGN é adotado para, por exemplo, facilitar comparação com resultados disponíveis na literatura [8]. Ressalta-se que há vários modelos na literatura, tais como os que adotam múltiplas entradas e saídas [30], mas a Equação 2.5 é representativa quando se trata da literatura de CAM.

2.3 Abordagem baseada no reconhecimento de padrão

Conforme pode ser visto na Figura 2.2, a classificação automática de modulação baseada no reconhecimento de padrão é dividida em dois subsistemas: o extrator de características e um de reconhecimento de padrões (o classificador). O primeiro bloco extrai a informação relevante para a decisão, transformando os sinais recebidos (após passar pelo canal) em um conjunto chamado de características ou parâmetros. Essas informações são exploradas pelo classificador para escolher a classe que melhor descreve a modulação do sinal.

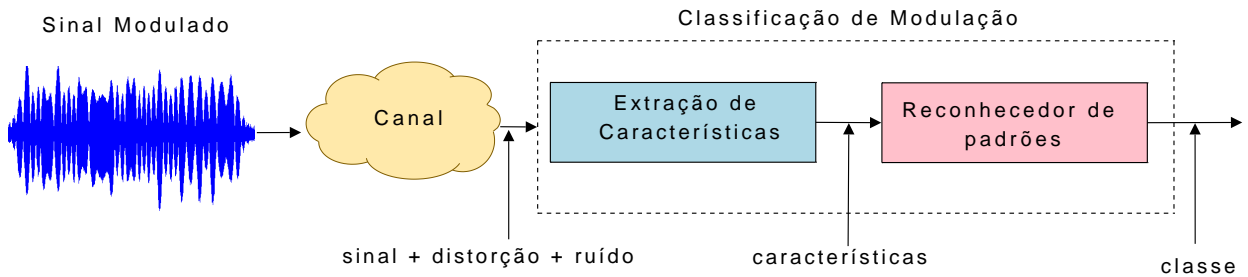


Figura 2.2: Diagrama das Etapas de classificação de modulação usando reconhecimento de padrão.

O extrator de características possui o objetivo de converter uma sequência de símbolos recebidos $\mathbf{r} = \{r_n\}_{n=1}^N$ em um vetor $\mathbf{x} = \{x_d\}_{d=1}^D$ com D características escalares. Após a conversão, o vetor \mathbf{x} é a entrada do classificador $f(\mathbf{x})$.

O classificador é definido matematicamente por um mapeamento $f : \mathbb{R}^D \rightarrow y = 1, \dots, Y$, onde D é a dimensão do vetor de entrada \mathbf{x} e a saída $y \in \mathbb{R}$ representa o rótulo da classe ou saída de dados. Na etapa de treinamento de um classificador convencional utilizando aprendizado supervisionado é imprescindível um conjunto de treino $\Psi = (\mathbf{x}_1, y_1), \dots, (\mathbf{x}_T, y_T)$ contendo T exemplos (\mathbf{x}, y) , ou seja, as amostras da etapa de treino são constituídas pelo conjunto de dados de entrada associados às suas respectivas classes pré-rotuladas. Uma vez treinado o classificador f , o objetivo é classificar um novo conjunto de dados, chamado conjunto de teste disjunto de Ψ , o qual pode ser não rotulado. Entretanto, caso o rótulo já exista, esse é escondido do classificador e usado apenas para calcular a taxa de erro.

Nota-se que a classificação baseada no reconhecimento de padrão exige, evidentemente, algumas características para representação da informação a fim de realizar uma decisão [31]. De acordo com [32], as características mais usadas na classificação de modulação podem ser categorizadas em cinco tipos: domínio de tempo instantâneo, domínio da transformada, estatística, formato da modulação e características de passagem por zero (*zero-crossing*).

Realiza-se a seguir uma breve descrição dessas características chaves. Apesar das mesmas não serem usadas diretamente nas simulações desta Tese, as descrições permitem melhor entender o estado da arte e os extratores de parâmetros propostos na Tese.

2.3.1 Características baseadas no domínio de tempo instantâneo (DTI)

Características baseadas no domínio de tempo instantâneo foram concebidas a partir do trabalho de Nandi e Azzouz para classificação de modulações analógicas e digitais [33–35]. Tais pesquisadores, na época, perceberam que modulações distintas apresentam diferentes propriedades na sua amplitude, fase e frequência, de modo que possibilitou que um grupo de modulações pudesse ser dividido em um conjunto ou subconjunto a partir dessas propriedades.

A primeira característica, σ_{ap} , é o desvio padrão do valor absoluto da componente não linear da fase instantânea:

$$\sigma_{ap} = \sqrt{\frac{1}{N_c} \left[\sum_{A_n[n] > t_{th}} \phi_{NL}^2[n] \right] - \left[\frac{1}{N_c} \sum_{A_n[n] > t_{th}} |\phi_{NL}^2[n]| \right]^2} \quad (2.6)$$

onde N_c é o número de amostras que atende à condição $A_n[n] > t_{th}$. A variável t_{th} é um valor limite que filtra as amostras que apresentam valores de amplitude pequenos devido sua alta sensibilidade ao ruído. O termo $\phi_{NL}^2[n]$ denota a n -ésima componente não linear da fase instantânea do sinal.

A segunda característica, σ_{dp} , é o desvio padrão da componente não linear da fase instantânea direta, representada por:

$$\sigma_{dp} = \sqrt{\frac{1}{N_c} \left[\sum_{A_n[n] > t_{th}} \phi_{NL}^2[n] \right] - \left[\frac{1}{N_c} \sum_{A_n[n] > t_{th}} \phi_{NL}^2[n] \right]^2} \quad (2.7)$$

onde todos os parâmetros são os mesmos da Equação 2.6.

A terceira característica, \mathcal{P} , é a avaliação da simetria do espectro em torno da frequência da portadora, avaliada por:

$$\mathcal{P} = \frac{\mathcal{P}_L - \mathcal{P}_U}{\mathcal{P}_L + \mathcal{P}_U} \quad (2.8)$$

onde as variáveis \mathcal{P}_L e \mathcal{P}_U são obtidas a partir de:

$$\mathcal{P}_L = \sum_{n=1}^{f_{cn}} |X_c[n]|^2 \quad (2.9)$$

$$\mathcal{P}_U = \sum_{n=1}^{f_{cn}} |X_c[n + f_{cn} + 1]|^2 \quad (2.10)$$

onde $X_c[n]$ é a transformada de fourier do sinal $x_c[n]$, $(f_{cn} + 1)$ é o número de amostras correspondentes à portadora de frequência f_c , onde f_{cn} é calculado a partir de

$$f_{cn} = \frac{f_c N}{f_s} - 1, \quad (2.11)$$

onde f_s é a frequência de amostragem e N representa o número de amostras.

A quarta característica, σ_{aa} é o desvio padrão do valor absoluto da amplitude instantânea normalizada e centrada, dada por:

$$\sigma_{aa} = \sqrt{\frac{1}{N} \left[\sum_{n=1}^N A_{cn}^2[n] \right] - \left[\frac{1}{N} \sum_{n=1}^N |A_{cn}[n]| \right]^2}, \quad (2.12)$$

onde A_{cn} é a amplitude instantânea normalizada e centrada do sinal recebido \mathbf{r} , e N é o número total de amostras. A normalização é obtida por meio de

$$A_{cn}[n] = A_n[n - 1], \text{ onde } A_n[n] = \frac{A[n]}{\mu_A} \quad (2.13)$$

e μ_A é a média da amplitude instantânea de um segmento do sinal, calculada a partir de

$$\mu_A = \frac{1}{N} \sum_{n=1}^N a[n]. \quad (2.14)$$

A quinta característica, σ_{af} , é o desvio padrão do valor absoluto da frequência instantânea normalizada e centrada, dada por

$$\sigma_{af} = \sqrt{\frac{1}{N_c} \left[\sum_{A_n[n] > t_{th}} f_N^2[n] \right] - \left[\frac{1}{N_c} \sum_{A_n[n] > t_{th}} |f_N[n]| \right]^2} \quad (2.15)$$

onde a frequência centrada instantânea f_m é normalizada pela frequência de amostragem f_s de tal forma que

$$F_N[n] = \frac{f_m[n]}{f_s} \quad (2.16)$$

se aplica.

A frequência instantânea é centralizada aplicando a média da frequência μ_f da seguinte forma:

$$f_m[i] = f[n] - \mu_f. \quad (2.17)$$

$$\mu_f = \frac{1}{N} \sum_{n=1}^N f[n]. \quad (2.18)$$

A sexta característica refere-se ao desvio padrão da amplitude normalizada e centralizada, σ_a , dada por:

$$\sigma_a = \sqrt{\frac{1}{N_c} \left[\sum_{A_n[n] > t_{th}} A_{cn}^2[n] \right] - \left[\frac{1}{N_c} \sum_{A_n[n] > t_{th}} A_{cn}[n] \right]^2}. \quad (2.19)$$

Outras características são baseadas na kurtosis, tanto da amplitude instantânea normalizada e centralizada quanto da frequência, dadas respectivamente por:

$$\mu_{42}^a = \frac{E\{a_{cn}^4[n]\}}{\{E(a_{cn}^2[n])\}^2} \quad (2.20)$$

e

$$\mu_{42}^f = \frac{E\{f_N^4[n]\}}{\{E(f_N^2[n])\}^2}. \quad (2.21)$$

Um vez definidas as características baseadas no domínio do tempo instantâneo, analisa-se cada uma, em seguida, revisando-se seu uso em classificação de modulação.

A característica σ_{ap} , Equação 2.6, mede a variação da fase instantânea absoluta [36]. Tal característica pode ser aplicada para distinguir dois conjuntos de modulações. O primeiro conjunto inclui as modulações FM, SSB, M-FSK, M-PSK ($M \geq 2$) e M-QAM, enquanto o segundo conjunto inclui AM, M-ASK e BPSK. Vale lembrar que a modulação BPSK não possui nenhuma informação na sua fase instantânea absoluta, pois nesse tipo de modulação a fase instantânea apresenta somente dois estados. Além disso, quando a fase instantânea está centrada no zero, os valores absolutos da fase são os mesmos.

A segunda característica σ_{dp} , também mede a variação da fase instantânea absoluta [37]. Assim como a característica σ_{ap} , a característica σ_{dp} pode ser aplicada para distinguir dois conjuntos de modulações. O primeiro conjunto inclui FM, SSB, M-FSK, M-PSK, M-QAM. O segundo conjunto de modulações inclui AM e M-ASK. Da mesma forma que a variável σ_{ap} , σ_{dp} possui a capacidade de distinguir a modulação BPSK das outras sem a informação da fase.

Já a característica \mathcal{P} fornece o critério para classificar diferentes modulações baseadas na amplitude com diferentes propriedades no domínio da frequência. A característica pode ser aplicada para classificar um conjunto de modulações que possuem espectros simétricos relativos à frequência da portadora, por exemplo: AM e *double-sideband modulation* (DSB). Além disso, também pode ser aplicada para classificar modulações que possuem densidade espectral assimétrica relativos à frequência da portadora, por exemplo: modulação *vestigial sideband* (VSB) e *upper sideband* (USB) [32].

A característica σ_{aa} mede a quantidade de informação contida na amplitude instantânea do sinal de modo que possui a capacidade de distinguir a modulação 2-ASK do restante da família M-ASK [36,38,39]. Isso pode ser explicado, pois o valor absoluto da amplitude instantânea normalizada e centrada do sinal 2-ASK é constante em relação ao restante da família M-ASK [32].

O desvio padrão do valor absoluto da frequência instantânea normalizada e centrada, σ_{af} , possui a capacidade de distinguir as modulações 2-FSK e 4-FSK [40]. A teoria que permite essa discriminação parte do mesmo princípio visto nas características σ_{ap} e σ_{aa} , onde a

informação da frequência instantânea normalizada e centrada da modulação 2-FSK é constante.

A característica μ_{42}^a mede quão compacta é a distribuição da amplitude instantânea. A modulação AM sendo um tipo de modulação analógica possui diferentes valores de amplitudes instantâneas, de maneira que possui uma distribuição menos compacta em comparação as modulações digitais, tais como a família M-ASK. Em [41], a característica μ_{42}^a serviu de entrada para uma árvore de decisão para distinguir sinais AM dos sinais 2-ASK/4-ASK.

Diferente da característica μ_{42}^a , a característica μ_{42}^f mede quão compacta é a distribuição da frequência instantânea. Assim como a modulação AM, a modulação FM possui uma distribuição menos compacta em comparação as modulações digitais, tais como a família M-FSK.

Embora as características instantâneas sejam relativamente fáceis de extrair, essas possuem a desvantagem de serem sensíveis ao ruído, de modo que são geralmente usadas em técnicas de reconhecimento de padrão quando a RSR possui valores relativamente altos.

2.3.2 Características baseadas em transformações

Como dito antes, as características baseadas em transformações são caracterizadas pela extração dessas através de uma transformada. Por exemplo, em trabalho recente [42], a transformada S foi utilizada com sucesso para CAM. Porém, as transformadas mais comuns são as de Fourier ou Wavelet.

Como exemplo do uso de Fourier, o valor máximo de uma estimativa da densidade espectral de potência da amplitude instantânea centrada normalizada pode ser obtida por:

$$\gamma_{max} = \frac{|\text{TDF}(A_{cn}(n))|^2}{N}, \quad (2.22)$$

onde TDF é a transformada discreta de Fourier, A_{cn} é a amplitude instantânea normalizada e centrada do sinal recebido \mathbf{r} e N é o número total das amostras do sinal. γ_{max} caracteriza as variações de amplitude, de maneira que ajuda a discriminar as modulações baseadas em amplitude das não baseadas tanto na forma analógica [17] quanto digital [40]. Aplicando um limiar predeterminado dessa característica, essa ajuda a diferenciar diferentes tipos de sinais AM e CW/FM [38]. Na modulação digital, essa característica é amplamente usada para discriminar modulações M-QAM/M-ASK e FSK/PSK [37,40,41]. Além disso, essa característica é utilizada como entrada em rede neurais [41] e MVS para classificação de diferentes tipos de modulações de ordens distintas.

A facilidade em localizar as oscilações na frequência instantânea, amplitude e fase de

um sinal recebido fizeram da transformada Wavelet um objeto de estudo para classificação automática de modulação [43], de modo que suas características aplicadas juntamente com rede neurais [36] ou MVS [44] ajudam na classificação de diferentes tipos de modulação, incluindo modulações QAM, FSK e PSK. A transformada contínua Wavelet (TCW) de um sinal recebido $r(t)$ é definida pela Equação 2.23

$$\text{TCW}(a, \tau) = \int_{-\infty}^{\infty} r(t)\Psi_{a,\tau}^* dt, \quad (2.23)$$

onde $\Psi_{a,\tau}^*$ é a conjugada transposta da função wavelet ao longo do tempo. Entre as distintas funções Wavelet, a Wavelet de Haar geralmente é escolhida pelos pesquisadores devido sua simplicidade. Um exemplo de sua aplicação é quando a Wavelet de Haar é utilizada na amplitude dos sinais PSK, QAM e FSK, nota-se um comportamento distinto para essas modulações e conseqüentemente ajuda na identificação [45, 46]. Uma das vantagens da utilização dessa técnica é a capacidade de reduzir os efeitos dos ruídos [36, 44]. Apesar de conseguir distinguir as modulações PSK, QAM e FSK, a aplicação da transformada Wavelet não consegue diferenciar as modulações ASK e QAM de modo que o autor do trabalho [47] sugere o uso de momentos de alta ordem e cumulantes a fim de explorar as pequenas diferenças entre as duas modulações.

2.3.3 Informações baseadas em estatísticas do sinal

Nesta parte do trabalho, apresenta-se as características baseadas em estatísticas de alta ordem, mais especificamente nas características baseadas em momentos e cumulantes.

2.3.3.1 Características baseadas em momentos de alta ordem (MAO)

O trabalho [48] foi o primeiro adotar o momento de terceira ordem como recurso para classificação de modulação. Posteriormente, o uso de momentos na classificação de momentos foi investigado no trabalho [49], onde momentos de alta ordem foram aplicados em sinais M-PSK para fins de classificação de modulação. A classificação foi realizada através da derivação do k -ésimo momento do sinal de fase num canal Gaussiano, onde concluiu-se que os momentos são uma função crescente no que diz respeito a ordem da modulação M-PSK. Desta forma, à medida que se aumenta a ordem da modulação M-PSK, o valor do momento cresce, fornecendo condições para classificação de modulação de sinais M-PSK de diferentes ordens. Além desta última contribuição, o trabalho fez a observação que a diferença dos valores de momentos de baixa ordem entre as modulações de alta ordem não são distintos. Desta forma, a fim de que haja uma classificação efetiva da modulação M-PSK de alta ordem, há necessidade o calculo

de momentos de alta ordem. O calculo do momento de ordem k de um sinal de fase é definido por:

$$\mu_k(\mathbf{r}) = \frac{1}{N} \sum_{n=1}^N \phi^k[n] \quad (2.24)$$

onde $\phi[n]$ é a fase da n -ésima amostra do sinal. O trabalho [50] propôs a kurtosis da amplitude instantânea centrada normalizada μ_{42}^a e kurtosis da frequência instantânea centrada normalizada μ_{42}^f para classificação de sinais M-ASK e M-FSK. As equações para essas duas características podem ser vistas respectivamente nas Equações 2.20 e 2.21. O trabalho [51] generalizou as características baseadas em momentos, incluindo momentos de alta ordem para fase do sinal e magnitude de frequência. O trabalho [52] empregou momentos de alta ordem cíclicos como característica para classificação de modulações com funções de autocorrelação cíclicas idênticas. Neste trabalho, usa-se a Equação 2.25 para calcular o k -ésimo momento de um sinal complexo \mathbf{r} ,

$$\mu_{xy}(\mathbf{r}) = \frac{1}{N} \sum_{n=1}^N (r[n])^x (r^*[n])^y \quad (2.25)$$

onde $x + y = k$ e $r^*[n]$ é o complexo conjugado de $r[n]$.

2.3.3.2 Características baseadas em cumulantes de alta ordem (CAO)

O trabalho [27] sugeriu a aplicação de cumulantes de quarta ordem para sinais complexos como alternativa para classificação de sinais baseados nas seguintes modulações: M-PAM, M-MPSK e M-QAM. Trabalhos recentes seguiram a linha do uso de cumulantes [53]. Para um sinal recebido \mathbf{r} , o momento de segunda ordem pode ser definido de duas maneiras distintas através de:

$$C_{20} = E\{r^2[n]\} \quad (2.26)$$

e

$$C_{20} = E\{|r[n]|^2\}. \quad (2.27)$$

Do mesmo modo, os momentos e os cumulantes de quarta ordem podem ser aplicados de três maneiras diferentes, usando distintos posicionamentos de conjugação:

$$C_{40} = \text{cum}(r[n], r[n], r[n], r[n]) \quad (2.28)$$

$$C_{41} = \text{cum}(r[n], r[n], r[n], r^*[n]) \quad (2.29)$$

$$C_{40} = \text{cum}(r[n], r[n], r^*[n], r^*[n]) \quad (2.30)$$

onde $\text{cum}(\cdot)$ é função de cumulante conjunta definida por:

$$\text{cum}(w, x, y, z) = E(wxyz) - E(wx)E(yz) - E(wy)E(xz) - E(wz)E(xy). \quad (2.31)$$

Enquanto isso, a estimativa dos cumulantes de segunda e quarta ordem são obtidas através dos seguintes processos como mostrado em:

$$\hat{C}_{20} = \frac{1}{N} \sum_{n=1}^N r^2[n] \quad (2.32)$$

$$\hat{C}_{21} = \frac{1}{N} \sum_{n=1}^N |r[n]|^2 \quad (2.33)$$

$$\hat{C}_{40} = \frac{1}{N} \sum_{n=1}^N r^4[n] - 3\hat{C}_{20} \quad (2.34)$$

$$\hat{C}_{41} = \frac{1}{N} \sum_{n=1}^N r^3[n]r^*[n] - 3\hat{C}_{20}\hat{C}_{21} \quad (2.35)$$

$$\hat{C}_{42} = \frac{1}{N} \sum_{n=1}^N |r[n]|^4 - |\hat{C}_{20}|^2 - 2\hat{C}_{21}^2 \quad (2.36)$$

Através da Tabela 2.1, pode-se verificar claramente que diferentes modulações têm distintos valores de cumulantes em um canal ideal, de modo que a classificação dessas modulações pode ser realizada com relativa facilidade quando a RSR é alta.

2.3.3.3 Características baseadas na cicloestacionariedade

A teoria de cicloestacionariedade está presente em diversas áreas de conhecimento. Especificamente em telecomunicações, essa técnica surge devido à combinação de mensagens (sinais estacionários) com portadoras e trem de pulsos, e também, pode ocorrer como resultado de outros processos usados na geração de sinais tais como amostragem e multiplexação [54, 55].

O trabalho de Gardner [56] foi um dos pioneiros a demonstrar o potencial da análise de sinais cicloestacionários para sistemas de comunicação sem fio. Depois desse trabalho, o estudo sobre análise cicloestacionária cresceu consideravelmente para uma grande variedade

	C_{20}	C_{21}	C_{40}	C_{41}	C_{42}
2-PAM	1.0	1.0	-2.0	-2.0	-2.0
4-PAM	1.0	1.0	-1.360	-1.36	-1.36
8-PAM	1.0	1.0	-1.2381	-1.2381	-1.2381
BPSK	1.0	1.0	-2.0	-2.0	-2.0
QPSK	0.0	1.0	1.0	0.0	-1.0
8-PSK	0.0	1.0	0.0	0.0	-1.0
4-QAM	0.0	1.0	1.0	0.0	-1.0
16-QAM	0.0	1.0	-0.68	0.0	-0.68
64-QAM	0.0	1.0	-0.61	0.0	-0.6191

Tabela 2.1: Valores de cumulantes de um canal ideal para vários tipos de modulações.

aplicações, incluindo detecção de sinais [57], classificação [58], sincronização e equalização [59]. Entre suas vantagens são a redução da sensibilidade ao ruído e interferência de sinais; a capacidade de extrair parâmetros do sinal tais como frequência da portadora e taxa de símbolos.

Um processo é conhecido como cicloestacionário quando algumas de suas características estatísticas são periódicas, por exemplo: médias, autocorrelação, espectro de potência, função densidade de probabilidade (FDP), etc. [60]. Para entender melhor essa definição, pode-se fazer uma comparação com sinais estacionários, os quais apresentam estatísticas que são invariantes com o tempo. Entretanto, essa invariância não acontece para sinais cicloestacionários, ou seja, esses sinais são não-estacionários e apresentam estatísticas que variam periodicamente com tempo, os quais ajudam a compor métodos excelentes para detecção e identificação de sinais de comunicação.

Uma das diversas maneiras que a cicloestacionariedade possui para distinguir as modulações é através do espectro cíclico [61, 62]. Nota-se, claramente, através da Figura 2.3, a magnitude $|S_x(f, \alpha)|$ do espectro cíclico dos sinais BPSK e QPSK, onde f é a frequência convencional e α é o “ciclo” [56].

Tais diferenças vistas na Figura 2.3 servem de entrada para algoritmos de reconhecimento de padrão [63], os quais possuem bons resultados de classificação em cenários onde a relação RSR é baixa.

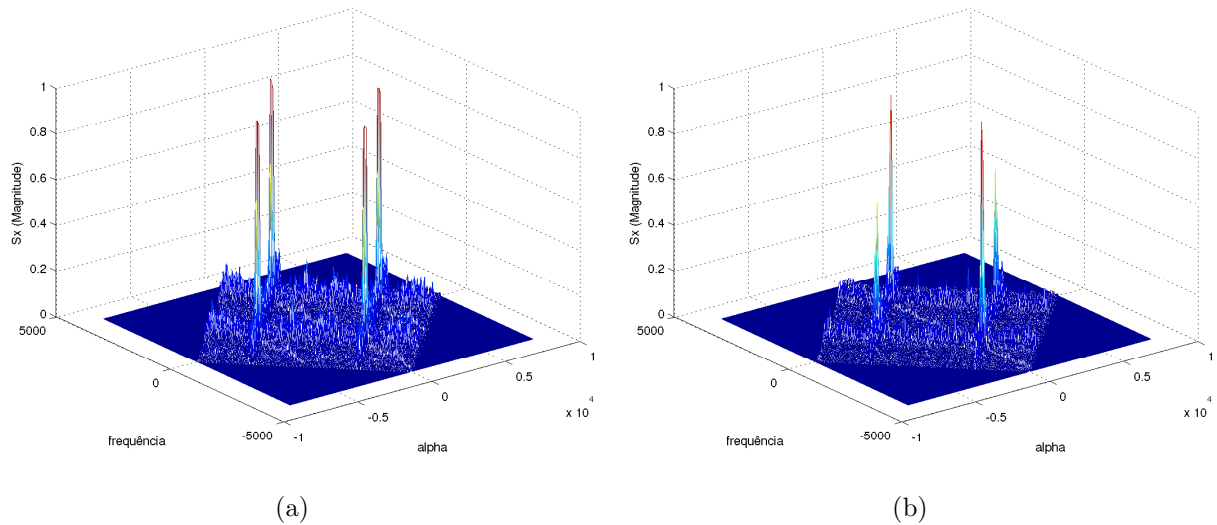


Figura 2.3: Exemplos de espectros cíclicos. (a) Espectro cíclico do sinal BPSK. (b) Espectro cíclico do sinal QPSK.

2.3.4 Características baseadas no formato da constelação

A forma geométrica de diagramas de constelação caracteriza cada formato de modulação. Tal característica está associada com o número de símbolos e sua localização na constelação, como pode ser visto na Figura 2.4. Pela Figura 2.4, percebe-se que cada ponto do diagrama tem um distância em relação à origem e uma fase. Em [64], foi proposta uma abordagem de classificação de modulação baseado em diagramas de constelação, comparando referências dos pontos da constelação. Tal estudo investigou o efeito do ruído sobre a taxa de classificação. Em outros trabalhos, as características foram baseadas em um conjunto de amplitudes e distâncias euclidianas entre os pontos das constelações a fim de realizar a classificação entre sinais PSK e QAM [64].

2.3.5 Informações baseadas na passagem por-zero (*Zero Crossing*)

Técnicas baseadas na passagem por zero possuem uma variedade de aplicações [65]. Uma delas é servir como uma ferramenta para classificação de modulação [66]. Tal técnica é caracterizada por registrar o número de vezes que o sinal recebido passa pelo nível de voltagem zero. A vantagem desta técnica é fornecer uma informação útil relacionada à frequência e transição de fase do sinal recebido com baixo custo computacional.

O trabalho de Callaghan [67] aplicou essa técnica para classificar sinais AM, FM, SSB, e FSK. Além disso, tal técnica tem sido aplicada para determinar a ordem de sinais FSK e

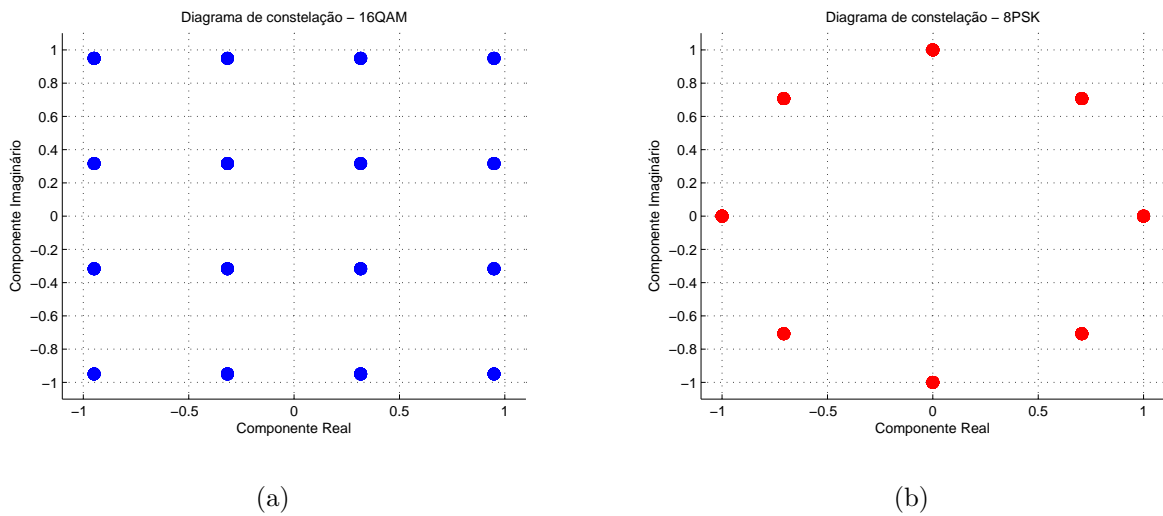


Figura 2.4: Exemplos de diagramas de constelações para modulação 16-QAM e 8-PSK. (a) Diagrama de constelação para 16-QAM. (b) Diagrama de constelação para 8-PSK.

distinguir as modulações PSK e FSK [68].

2.3.6 Classificadores

Após a extração das características do sinal recebido, na abordagem de reconhecimento de padrão, o próximo passo é a identificação da modulação através de algoritmos de reconhecimento de padrão. No processo de classificação por meio de características, o processo de decisão é realizado principalmente métodos de inteligência computacional. Uma primeira categoria de métodos são os baseados em Árvore de Decisão [36, 38, 41]. Outra categoria são os que usam Redes Neurais [41, 69, 70] ou Máquina de Vetores de Suporte (MVS) [13, 43, 44]. Também são usadas combinações de uma ou mais técnicas de inteligência computacional com a finalidade de otimizar a solução. O principal objetivo é sempre aumentar a probabilidade de classificação em cenários onde o valor da RSR é baixa [8].

Um exemplo de combinação de técnicas de Inteligência Artificial é o trabalho de Aslam [71], onde o algoritmo KNN (*K-nearest neighborhood*) juntamente com Programação Genética classificaram sinais 2-PSK, QPSK, 16-QAM e 64-QAM. Nas próximas sub-seções, este trabalho irá focar nos classificadores mais frequentemente utilizados na classificação automática de modulação.

2.3.6.1 Rede Neurais Artificiais

As Redes Neurais Artificiais são um dos algoritmos de reconhecimento de padrão mais utilizados na classificação automática de modulação. Tal característica é devido sua estrutura flexível, de modo que torna sua implementação facilitada. Além disso, as Redes Neurais podem se adaptar e aprender vários tipos de modulações distintas. Do ponto de vista de algoritmos de aprendizagem, as Redes Neurais são categorizadas em dois grupos: redes supervisionadas e não supervisionadas. Em relação às Redes Neurais supervisionadas, uma parte do conjunto de dados é usado para fins de aprendizagem e a outra parte é usada para testes. Diferente das Redes Neurais supervisionadas, as Redes Neurais não supervisionadas realizam o agrupamento de dados de entrada, de modo a identificar as classes distintas.

As Redes Neurais baseadas no aprendizado supervisionado fornecem resultados mais precisos. A maioria das Redes Neurais Artificiais que foram testadas na pesquisa sobre classificação automática de modulação utilizam técnicas de aprendizado supervisionado, incluindo Redes Neurais MLP (*Multi-Layer-Perceptron*) e Redes Neurais RBF (*Radial Basis Function*).

As Redes Neurais MLP possuem um atrativo aos projetistas devido suas implementações necessitarem de pouca memória, de maneira que a maioria dos trabalhos baseados em Redes Neurais aplicados a classificação de modulação utilizam uma única Rede Neural MLP como em [72]. Entretanto há alguns trabalhos que sugerem três Redes Neurais MLP em cascata. A última técnica usa a saída da primeira Rede Neural MLP como entrada para segunda ou terceira Rede Neural.

O uso de Rede Neural MLP apresenta uma performance melhor quando comparada com árvores de decisão, considerando as mesmas características [41]. De acordo com trabalho de [32], não há mudanças significativas de classificação quando Redes Neurais MLP ou RBF são aplicadas em uma tarefa de classificação de modulação.

2.3.6.2 Máquina de Vetores de Suporte

Embora as Redes Neurais Artificiais sejam amplamente aplicadas na pesquisa de Classificação Automática de Modulação, limitações como um ajuste demasiado à base de treinamento conhecido como sobre-ajuste (*overfitting*) e/ou mínimos locais prejudicam o desempenho desse algoritmo durante a etapa de teste. Essas limitações são superadas pelas MVS, de modo que esse classificador fornece maior capacidade de generalização, mesmo em cenário onde o valor da RSR possui um valor baixo. O principal objetivo de uma MSV é projetar hiperplanos que separam duas classes com a máxima *margem* [13, 73]. Um estudo sobre essa técnica é visto no próximo capítulo deste trabalho.

Geralmente, os trabalhos de CAM aplicam diferentes kernels a uma MVS, tais como RBF [44, 74, 75] e funções polinomiais [44, 74]. As MVS alcançam uma maior probabilidade correta de classificação em relação a algoritmos baseados em Árvore de Decisão e Redes Neurais Artificiais [74, 75], devido principalmente essa técnica não sofrer com os problemas citados associados às Redes Neurais Artificiais e também não ser restrita às condições requeridas na configurações baseadas em Árvore de Decisão (regiões de decisão dadas pelos limiares escolhidos pela árvore).

Os resultados obtidos com a MVS são tipicamente excelentes, mesmo tendo características simples na entrada, como por exemplo no trabalho [13], onde os valores de amplitude e fase do sinal foram aplicados para classificação de sinais BPSK, 4-PAM, 8-PSK e 16-QAM. Ademais, dependendo das características de entrada no classificador, uma MVS é capaz de obter resultados melhores de classificação em relação à abordagens baseada em verossimilhança, como por exemplo, os resultados obtidos neste trabalho. Por fim, Algoritmos Genéticos podem ser utilizados juntamente com a MVS com a finalidade de otimizar os parâmetros dos kernels, como visto no trabalho [76].

Apesar das MVS já terem sido adotadas em CAM, este trabalho é inovador ao contrastar as MVS lineares com as não-lineares, mostrando que o custo computacional das lineares pode ser bastante reduzido. Tal fato não havia sido explorado anteriormente na literatura.

2.3.6.3 Árvore de Decisão

A classificação por meio da Árvore de Decisão é baseada na ideia de selecionar valores de limiares específicos a fim de separar as diversas modulações distintas [37, 38, 41]. A principal vantagem da utilização deste classificador é a facilidade de sua implementação [68], de modo que a adição de modulações na tarefa de classificação ocorre sem dificuldade por meio da inserção de ramos de decisão. Além disso, a interpretabilidade das árvores é bastante útil em algumas situações.

A Tabela 2.2 resume os métodos discutidos anteriormente.

2.4 Abordagem baseada na Teoria de Decisão

Sabe-se que o principal objetivo da CAM é descobrir a modulação de um sinal \mathbf{r} entre y possíveis modulações. Claramente esse objetivo pode ser visto como um problema de decisão entre várias hipóteses, de maneira que, do ponto de vista da teoria de decisão, tais hipóteses podem ser interpretadas do tipo “a sequência recebida \mathbf{r} pertence à modulação y ”.

Tabela 2.2: Resumo dos métodos CAM.

Tipo de Característica	Classificador			Modulação
	Redes Neurais	MVS	Árvore de Decisão	
DTI	[41, 74, 77]	[37, 71, 74]	[37–40, 68]	FSK
	[41, 74, 77]	[13]	[37–40, 68]	ASK
	[41, 74, 77]	[13, 74]	[37–40, 68]	PSK
	[41, 74, 77, 78]	[13]	[37–40, 68]	QAM
	[41, 74, 77]	[74]	[40]	AM, FM
	[41]		[40, 68]	MC
TF	[41]	[37]		ASK
	[41]		[79]	PSK
	[41]		[80]	FSK
	[41]			QAM
TW	[27, 47, 81, 82]	[43, 44, 71, 76]	[36, 38, 68]	ASK
	[47, 81, 83, 84]	[43, 71, 76]	[36, 38, 70]	PSK
	[47, 81, 83, 84]	[43, 44, 76]	[36, 38, 70]	FSK
	[47, 84, 85]	[43, 44]	[36, 38, 70]	QAM
MAO e CAO	[74, 77, 86]	[38, 75, 87, 88]	[27, 38]	ASK
	[70, 72, 74, 77, 89, 90]	[71, 75, 87]	[27, 38, 40, 68, 70, 91–93]	PSK
	[70, 72, 74, 86, 90]	[75]	[27, 38, 40, 68, 70]	FSK
	[70, 72, 74, 86, 90]	[71, 75, 87]	[27, 38, 40, 68, 70, 93]	QAM
			[40, 68]	MC
Cicloestacionariedade			[94]	ASK
			[95]	CPM
		[96]	[64, 85, 86, 94]	PSK
			[85, 86, 94]	QAM
			[25, 85, 86, 94]	FSK
			[85, 86, 88]	AM
Constelação		[64]		QAM
Passagem por zero			[97, 98]	PSK, FSK

Desta forma, a CAM baseada na teoria de detecção é caracterizada por um processo estatístico, o qual estima-se a modulação de um sinal desconhecido \mathbf{r} através de múltiplas hipóteses [99]. Tal medida utiliza a regra de Bayes, que pode ser obtida a partir da definição de probabilidade condicional, para calcular decisões baseadas em probabilidades previamente estimadas a partir de:

$$P(y|\mathbf{r}) = \frac{P(\mathbf{r}|y)P(y)}{P(\mathbf{r})}, \quad (2.37)$$

onde $P(y|\mathbf{r})$ é a probabilidade a posteriori, $P(\mathbf{r}|y)$ a verossimilhança, $P(y)$ a priori e $P(\mathbf{r})$ a evidência [15].

Nota-se um fato importante: a regra de Bayes é toda baseada na probabilidade a posteriori, em outras palavras, na probabilidade de uma hipótese (modulação) ser escolhida condicionada ao recebimento de uma sequência \mathbf{r} . Para conhecer essa probabilidade é necessário, obviamente, conhecer a probabilidade $P(\mathbf{r}|y)$ (verossimilhança), ou seja, uma medida que nos informe a probabilidade associada a uma sequência \mathbf{r} dado que ela pertença à classe y . A probabilidade $P(y)$ é a probabilidade (*prior*) de uma certa modulação y [15].

Os classificadores Bayesianos concentram-se em escolher o rótulo

$$\arg \max_{y=1,\dots,Y} P(\mathbf{r}|y)P(y)$$

que maximiza a probabilidade a posteriori. Todavia, os valores $P(\mathbf{r}|y)$ e $P(y)$ não são necessariamente conhecidos. Logo, o classificador F deve utilizar valores estimados de $\hat{P}(\mathbf{r}|y)$ e $\hat{P}(y)$ a fim de maximizar a equação 2.38.

$$F(\mathbf{r}) = \arg \max_{y=1,\dots,Y} \hat{P}(\mathbf{r}|y)\hat{P}(y) \quad (2.38)$$

O valor de estimação da probabilidade a priori $P(y)$, na maioria dos casos, pode ser obtida a partir do conjunto de treino através da contagem do número de ocorrências de cada classe contido nele. Assim, assume-se aqui que $\hat{P}(y) = P(y)$. Entretanto, estimar $P(\mathbf{r}|y)$ é mais complicado.

Essa dificuldade é caracterizada pelo fato dos classificadores adotarem uma distribuição paramétrica $\hat{P}(\mathbf{r}|y) = \hat{P}(\mathbf{r}|y, \Theta_y)$ chamada de *modelo de verossimilhança*, onde Θ_y são os parâmetros da distribuição a ser determinada (e.g., média e matriz de covariância se o modelo de verossimilhança for Gaussiano).

Diversos modelos de verossimilhança vêm sendo empregados em classificadores de Bayes para CAM [99]. Desta forma, a partir da próxima seção, inicia-se uma descrição dos principais modelos de verossimilhança para CAM enfatizando-se o classificador ALRT, que atinge o desempenho ótimo quando as estimativas coincidem com as distribuições corretas.

2.4.1 Um Limite Superior para a Probabilidade de Classificação Correta

Um limite superior para performance de qualquer algoritmo de classificação de modulação pode ser obtido via abordagem de teste da razão entre verossimilhanças médias (ALRT, do inglês *average likelihood ratio test*). Nesta técnica, considera-se que o modelo seja caracterizado apenas pela contaminação do sinal por ruído AWGN, $f_0 = 0$, $\phi_n = 0$, $\forall n$ e, para simplificar, $\alpha = 1$, ou seja, somente o valor da potência do ruído não é conhecida. Além disso, assume-se que o tipo de modulação é linear, com os símbolos obtidos a partir de uma constelação complexa.

Assumindo que o n -ésimo símbolo transmitido é $s_n^{(y)}$ e r_n é o recebido, a FDP condicional do símbolo recebido $P(r_n|s_n^{(y)})$ é uma Gaussiana complexa $\mathcal{N}_c(r_n|s_n^{(y)}, \sigma_u^2)$ com média $s_n^{(y)}$ e variância por dimensão $\sigma_u^2/2$. Em CAM, o receptor não possui o conhecimento do símbolo transmitido, de modo que a FDP de um único símbolo recebido é uma mistura de Gaussianas descrita por:

$$P(r_n|y) = \frac{1}{M_y} \sum_{j=1}^{M_y} \mathcal{N}_c(r_n|s_j^{(y)}, \sigma_u^2), \quad (2.39)$$

onde $s_j^{(y)}$ é o j -ésimo dentre os M_y símbolos da y -ésima constelação. Uma vez que \mathbf{r} é o vetor com símbolos r_n , a hipótese de símbolos transmitidos serem i.i.d. conduz a $P(\mathbf{r}|y) = \prod_{n=1}^N P(r_n|y)$ cujo o valor da verossimilhança provido pelo ALRT para cada classe é descrito por:

$$P(\mathbf{r}|y) = \prod_{n=1}^N \frac{1}{M_y} \sum_{j=1}^{M_y} \frac{1}{2\pi\sigma_u^2} \exp\left(-\frac{|r_n - s_j^{(y)}|^2}{\sigma_u^2}\right). \quad (2.40)$$

Para facilitar a tarefa de classificação, ao invés de aplicar a Equação 2.40 é útil adotar a verossimilhança logarítmica:

$$\ln P(\mathbf{r}|y) = \sum_{n=1}^N \ln \left[\frac{1}{M_y 2\pi\sigma_u^2} \sum_{j=1}^{M_y} \exp\left(\frac{-|r_n - s_j^{(y)}|^2}{2\sigma_u^2}\right) \right].$$

Lembrando que para o ALRT o vetor \mathbf{x} de entrada do classificador coincide com os símbolos recebidos \mathbf{r} , a decisão da classificação de modulação é baseada em:

$$F(\mathbf{r}) = \arg \max_{y=1, \dots, Y} \ln P(\mathbf{r}|y), \quad (2.41)$$

onde assume-se que a prior é uniforme, em outras palavras, as Y classes possuem a mesma probabilidade.

Tendo como base o ALRT, os seguintes classificadores foram propostos:

- Differential ALRT com fase e portadora desconhecida [100];
- Quasi-ALRT com fase da portadora desconhecida [101–103];
- Quasi-ALRT com fase e offset da portadora desconhecidos [103];
- ALRT com nível de sinal desconhecido [104].

2.4.1.1 Teste Discreto da Razão entre Verossimilhanças

Nota-se que o cálculo da Equação 2.41, dependendo do tamanho do vetor \mathbf{r} , podendo envolver uma grande quantidade de multiplicações e cálculos exponenciais de maneira a tornar essa técnica impraticável para cenários em tempo real [18]. A fim de reduzir o custo computacional do ALRT, o teste discreto da razão entre verossimilhanças (DLRT, do inglês *discrete likelihood ratio test*) foi proposto em [18].

Basicamente, para cada tipo de modulação, a proposta usa um plano complexo quantizado, uma grade cartesiana, de tamanho $Q \times Q$, e uma faixa de valores de ruído do sinal recebido é quantizado por U intervalos. Os valores pré-calculados da DLRT são armazenados em *Look-up Tables* (LUT), tabelas na verdade, com seus respectivos endereços.

A combinação de índices (p, q, u) é considerado como um endereço para a LUT. Desta forma, para cada símbolo de entrada r_n , esse é mapeado para os índices e , uma vez mapeado, o valor da função de verossimilhança de r_n é identificado na LUT por [105]:

- Escolher na tabela correspondente o valor de σ_u^2 que se aproxima mais da estimativa da variância do ruído e ;
- Escolher o número complexo quantizado que mais se aproxima de r_n .

Os valores de cada entrada da LUT são calculados a partir de:

$$P_{(p,q,u)}^{(y)}(\hat{r}_n|y) = \ln\left\{\frac{1}{2\pi\sigma_u^2 M_y} \sum_{k=1}^{M_y} \exp\left[-\frac{(x_p - S_{I,k}^{(y)})^2 + (d_q - S_{Q,k}^{(y)})^2}{2\sigma_u^2}\right]\right\} \quad (2.42)$$

com

$$x_p = \frac{2 \times R_{limite}}{Q} \times p - R_{limite} \quad (2.43)$$

e

$$d_q = \frac{2 \times R_{limite}}{Q} \times q - R_{limite} \quad (2.44)$$

onde p e $q \in \{0, 1, \dots, Q\}$, R_{limite} é o limite superior da parte real ou imaginária do símbolo recebido, M_y é o número de símbolos da y -ésima constelação, $S_k^{(y)}$ é o k -ésimo símbolo da y -ésima modulação e r_n é o símbolo demodulado. A componente em fase do símbolo complexo recebido r_n é denotada como x_p e a componente em quadratura do respectivo símbolo é denotada como d_q . O desvio padrão do u -ésimo valor da RSR é representado por σ_u .

Assim, como mencionado, a função *likelihood* para cada símbolo r_n é recuperada da LUT selecionando o valor DLRT que quantiza o número complexo mais próximo para o símbolo r_n . A y -ésima modulação é escolhida por:

$$F(\mathbf{r}) = \arg \max_{y=1, \dots, Y} \sum_{n=1}^N P_{(p,q,u)}^{(y)}(\mathbf{r}|y). \quad (2.45)$$

Basicamente, a DLRT é mais rápida do que a ALRT porque pré-computa os valores quantizados das likelihoods. Contudo, exige uma tabela para cada valor da RSR. Além disso, a granularidade da quantização, especificada por Q , impacta os resultados e consumo de memória.

Uma vez descrito o conceito de CAM baseada em *likelihood*, na Tabela 2.3 tem-se um resumo de trabalhos baseados nesta metodologia.

Além de ALRT e DLRT, existem diversas outras abordagens baseadas em verossimilhança. As mesmas fogem do escopo dessa Tese mas podem ser estudadas a partir de referências como as em [8].

2.4.2 Teste de Kolmogorov-Smirnov

Diferente dos testes paramétricos que envolvem problemas caracterizados pelo conhecimento da distribuição da população, o teste de Kolmogorov-Smirnov [108], em estatística, é considerado um método não paramétrico, de maneira que não é necessário especificar a distribuição da população de onde provém a amostra e serve para medir o quanto um modelo se encaixa em um conjunto de observações; tal medição em inglês conhecida como “*goodness of fit*” e sumariza a discrepância entre os valores observados e os valores esperados sob um certo modelo. Desta forma, esse teste é usado para verificar se duas populações possuem a mesma distribuição ou não, correspondendo às hipóteses:

- $H_0 : F_1 = F_0$
- $H_1 : F_1 \neq F_0$

Autor(es)	Classificador(es)	Modulações	Parâmetros Desconhecidos	Canal
Sills [100]	ALRT	BPSK, QPSK, 16QAM, V29, 32QAM, 64QAM	fase da portadora Θ	AWGN
Wei e Mendel [106]	ALRT	16QAM, V29	-	AWGN
Wei Su [18]	ALRT e DLRT	16QAM, 32QAM	-	AWGN
Alyaoui [107]	ALRT	16QAM, 64QAM, 16PSK, BPSK, QPSK	-	AWGN
Xu [105]	DLRT	8PSK, V.29-8, V.29-16, 16QAM and 64QAM	-	AWGN

Tabela 2.3: Trabalhos baseados em verossimilhança.

Primeiramente, para realizar o teste, é necessário estimar a função de distribuição acumulada a partir das amostras:

$$\hat{F}_1(z) = \frac{1}{N} \sum_{n=1}^N \mathcal{I}(z_n \leq z), \quad (2.46)$$

onde $\mathcal{I}()$ é uma função lógica, cujo valor é um se a entrada for verdadeira, ou zero caso contrário.

No teste de Kolmogorov-Smirnov considera-se a estatística

$$\hat{\mathcal{D}} = \max_{1 \leq n \leq N} |\hat{F}_1(z_n) - F_0(z_n)|, \quad (2.47)$$

como uma medida da discrepância entre a função de distribuição acumulada empírica (FDAE) \hat{F}_1 e a função de distribuição proposta F_0 . Percebe-se que $\hat{\mathcal{D}}$ representa um ajustamento entre \hat{F}_1 e F_0 , pois representa a distância vertical máxima entre as imagens das respectivas funções. Assim, se H_0 for considerada verdadeira, a distância vertical máxima entre as imagens das duas distribuições não deve ser muito grande, de forma que, $\hat{\mathcal{D}}$ apresente um valor baixo.

Para classificação de modulação baseado no sinal recebido \mathbf{r} usando o teste de Kolmogorov-Smirnov, o trabalho [28] definiu a sequência estatística de decisão z_n a partir de \mathbf{r} , onde z_n pode ser tanto magnitude, fase ou as componentes reais ou imaginárias de \mathbf{r} e então calcula-se a correspondente FDAE \hat{F}_1 para cada possível modulação candidata y . Assim, o teste estatístico de Kolmogorov-Smirnov é então calculado por:

$$\hat{\mathcal{D}}_y = \max_{i=1,\dots,y} |\hat{F}_1(z_n) - F_0(z_n)|, \quad (2.48)$$

onde a decisão de classificação é baseada em:

$$\hat{y} = \arg \min_{i=1,\dots,y} \hat{\mathcal{D}}_y. \quad (2.49)$$

Após descritos os principais extratores de parâmetros e classificadores encontrados na literatura, são discutidos a seguir os extratores de parâmetros resultantes da cooperação com outros pesquisadores.

2.5 Extrator de parâmetros CSS (*Concatenated Sorted Symbols*)

O extrator de parâmetros CSS foi proposto no trabalho [13], onde os símbolos das constelações [21] servem como parâmetro de entrada ao classificador. A ideia do extrator é armazenar a magnitude e a fase dos símbolos recebidos \mathbf{r} (após a normalização) e ordená-los separadamente. Uma vez ordenados os dois vetores (magnitude e fase), esses são concatenados, resultando em um novo vetor \mathbf{x} com dimensão $D = 2N$, onde N é o número de símbolos. O vetor \mathbf{x} busca refletir uma *assinatura* individual para cada modulação.

Para refletir tais assinaturas, a Figura 2.5 apresenta um esquema que facilita a compreensão do processo do extrator de parâmetros proposto a partir de um exemplo contendo duas constelações distintas. As Figuras 2.5(a) e 2.5(b) apresentam as constelações das modulações 16-QAM e 8-PSK respectivamente, ambas geradas sem a presença do ruído. A partir dessas duas constelações, as Figuras 2.5(c) e 2.5(d) apresentam separadamente um exemplo de um vetor \mathbf{x} com amostras de símbolos recebidos, onde esses podem ser vistos antes e após a ordenação (magnitude e fase estão devidamente separadas em ambos os casos) e sem nenhum ruído adicionado. O comprimento de cada vetor corresponde a $N = 250$ símbolos, produzindo, após a concatenação, um comprimento total de $D = 2N = 500$ elementos em \mathbf{x} , correspondentes a 250 magnitudes e 250 fases, nesta ordem. Observa-se que uma assinatura é gerada após a ordenação, a qual pode ser usada na identificação do tipo de modulação do sinal recebido.

No exemplo da Figura 2.5, observa-se claramente na Figura 2.5(c) que existem apenas três valores possíveis de módulo e doze valores possíveis de fase na constelação da modulação 16-QAM. A Figura 2.5(d) mostra que há apenas um valor de módulo e oito de fases para a modulação 8-PSK. A distinção é clara para uma situação ideal, sem ruído.

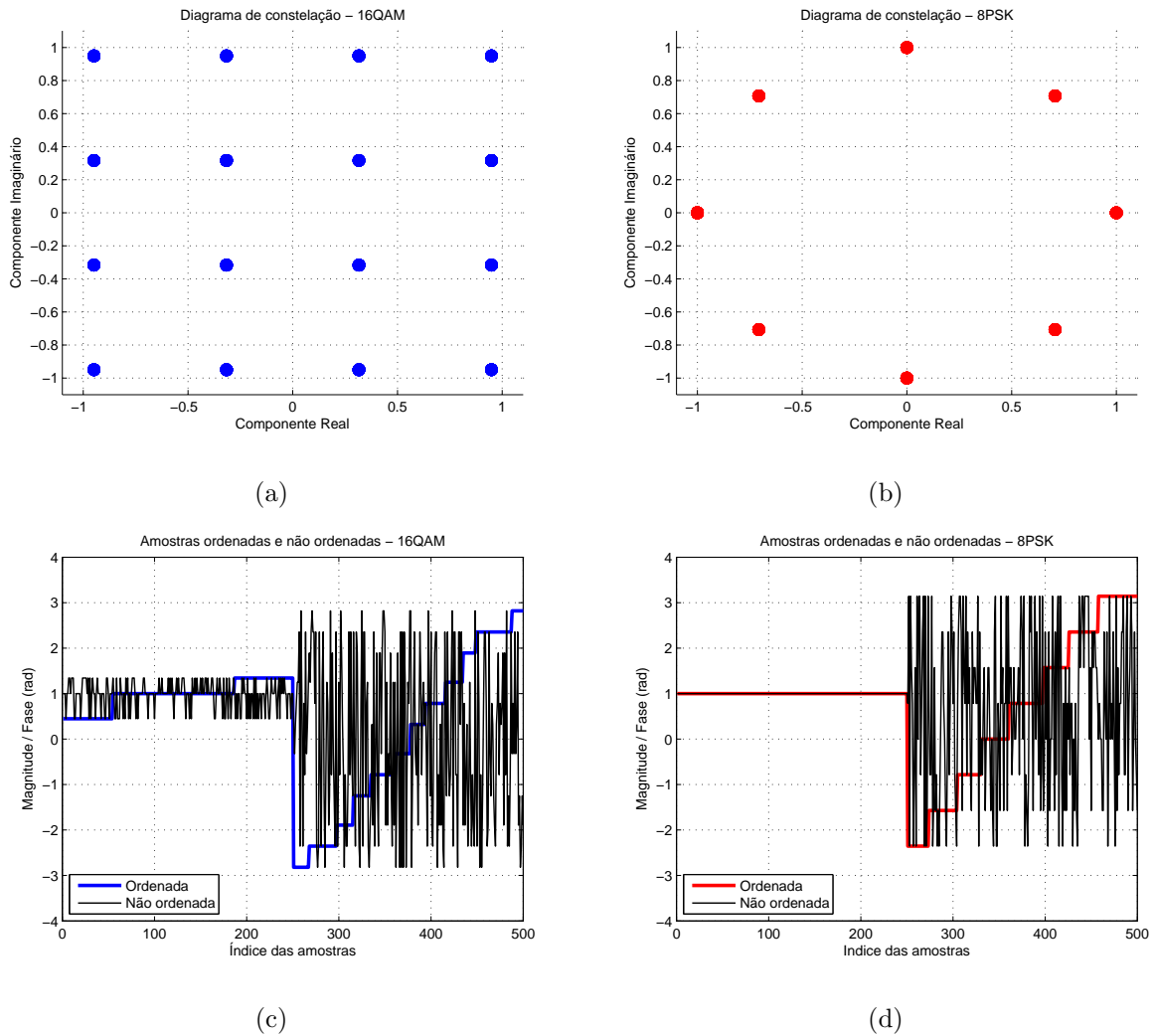


Figura 2.5: Exemplos de vetores com $D = 2N = 500$ elementos (250 magnitudes e 250 fases) para modulação 16-QAM e 8-PSK, sem ruído. (a) Diagrama de constelação para 16-QAM. (b) Diagrama de constelação para 8-PSK. (c) Amostras ordenadas e não ordenadas 16-QAM. (d) Amostras ordenadas e não ordenadas 8-PSK.

Uma outra forma de visualizar a distinção entre as modulações 16-QAM e 8-PSK é através da Figura 2.6, onde visualiza-se, de maneira sobreposta, a saída do extrator de parâmetros CSS de ambas as modulações.

Nota-se a distinção clara entre as assinaturas fornecidas pelo extrator de parâmetros

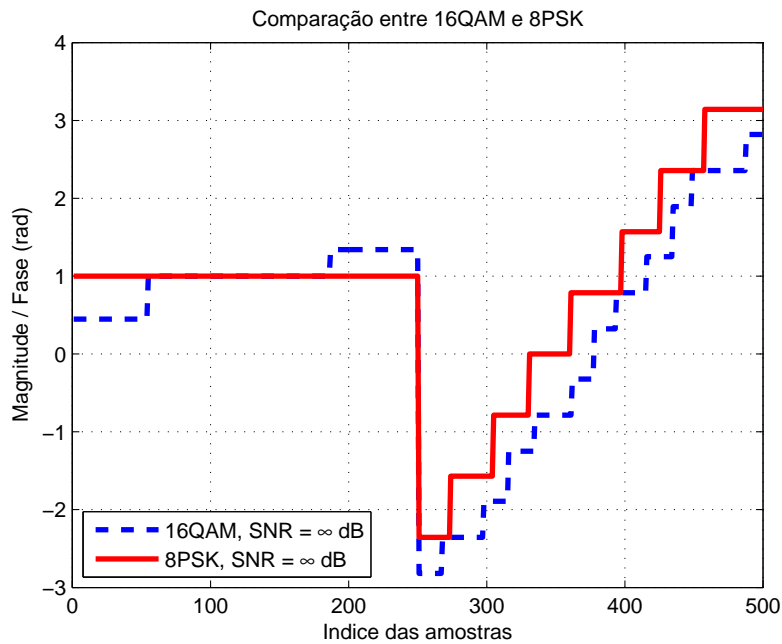


Figura 2.6: Comparação entre exemplos de dois vetores com $D = 2N = 500$ amostras (250 magnitudes e 250 fases) para comparação entre 16-QAM e 8-PSK, sem ruído.

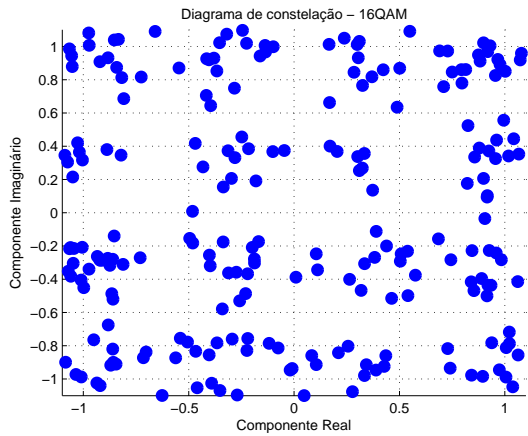
CSS para as duas modulações através da Figura 2.6, pois ruído algum foi adicionado.

Diferente do último exemplo, a Figura 2.7 apresenta uma sequência de símbolos contaminados por um ruído branco gaussiano com RSR= 15 dB. Percebe-se que a distinção, entre os dois tipos de modulação, não é mais tão óbvia como ilustra a Figura 2.8 uma vez que o ruído adicionado provoca uma distorção no sinal, prejudicando a distinção visual entre as modulações. Embora a distinção não seja tão evidente, em [13] é comprovado que ainda é possível distinguir as modulações mesmo com adição de ruído utilizando-se, por exemplo, o classificador MVS.

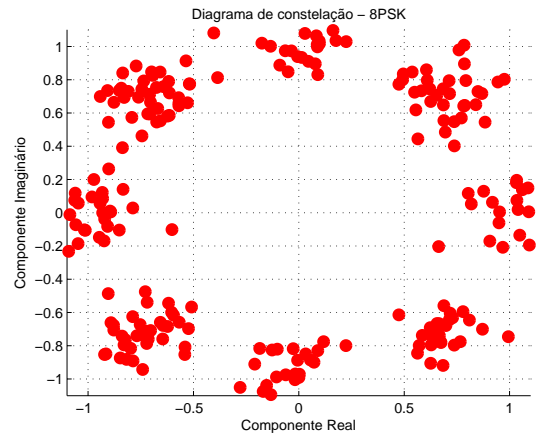
Por último, vale lembrar que a ordenação dos valores de magnitude e fase de maneira independente e a concatenação posteriormente provoca a perda de correlação entre esses componentes, entretanto a decorrelação não prejudica na tarefa de CAM.

2.6 Extrator de Parâmetros HISTO2D

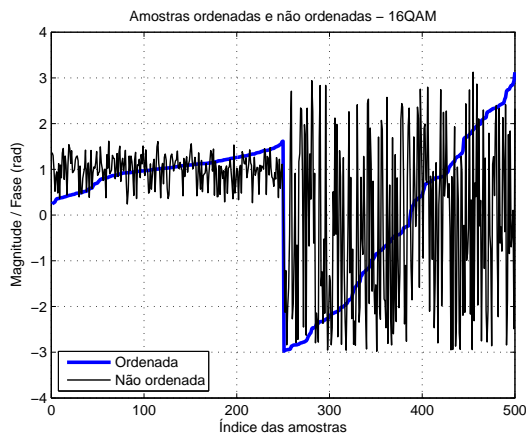
A motivação para o extrator de parâmetros veio da parceria com o grupo do professor José Luís Güntzel da Universidade Federal de Santa Catarina no trabalho [12], cujo primeiro autor é o Dr. Renan Alves. Neste trabalho, o extrator de parâmetros HISTO2D foi pro-



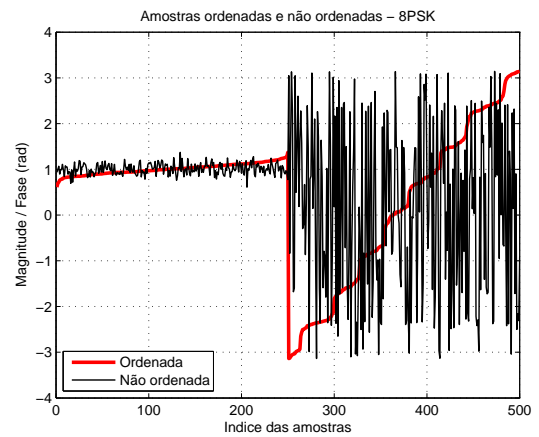
(a)



(b)



(c)



(d)

Figura 2.7: Exemplos de vetores com $D = 2N = 500$ elementos (250 magnitudes e 250 fases) para modulação 16-QAM e 8-PSK, $RSR = 15$ dB. (a) Diagrama de constelação para 16-QAM. (b) Diagrama de constelação para 8-PSK. (c) Amostras ordenadas e não ordenadas 16-QAM. (d) Amostras ordenadas e não ordenadas 8-PSK.

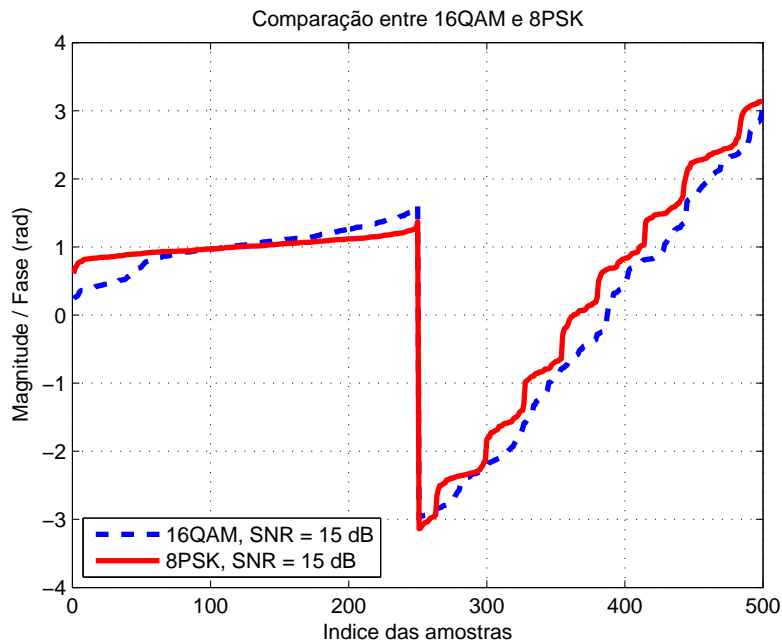


Figura 2.8: Comparaç o entre exemplos de dois vetores com $D = 2N = 500$ amostras (250 magnitudes e 250 fases) para comparaç o entre 16-QAM e 8-PSK, $RSR = 15dB$.

posto juntamente com sua implementaç o em FPGA. Nota-se que o extrator de par metros HISTO2D cria um grid bidimensional de bins, o qual explora as depend ncias estat sticas entre os eixos real e imagin rio de uma constelaç o, ou seja, o extrator de par metros HISTO2D preserva a correlaç o existente entre magnitude e fase.

Observando as sequ ncias de s mbolos recebidos sem ru do na Figura 2.9, onde cada eixo   caracterizado por $B = 32$ raias, totalizando $D = B^2 = 1024$ elementos, pode-se perceber, nitidamente, as distintas assinaturas obtidas das modulaç es BPSK, 4-PAM, 16-QAM e 8-PSK.

A Figura 2.10 apresenta as mesmas modulaç es e histogramas, por m acrescido de ru do com $RSR = 15$ dB. Apesar da adiç o de ru do provocar a distorç o do sinal recebido, percebe-se, de forma visual, que ainda   poss vel distinguir pela simples observaç o as diferenç as entre as assinaturas entre as modulaç es.

Por  ltimo, ressalta-se que a aplicaç o do HISTO2D quanto do CSS n o   exclusiva de um  nico classificador, ou seja, a aplicaç o desses extratores pode ser utilizada por qualquer algoritmo de reconhecimento de padr o.

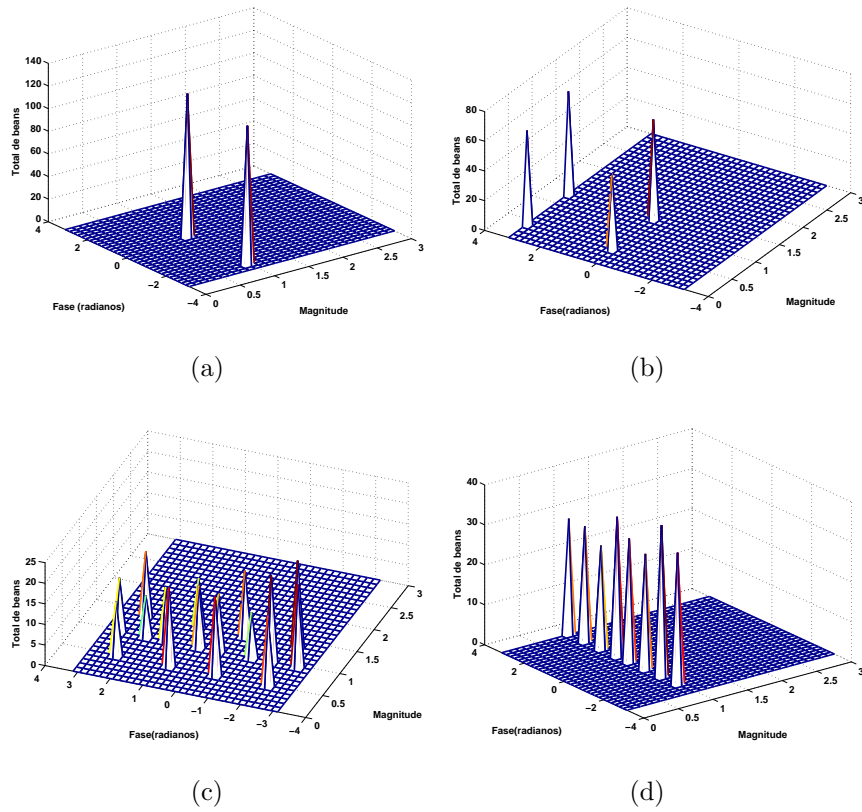


Figura 2.9: Exemplos de histogramas 2D com $N = 250$ símbolos (250 magnitudes e 250 fases) para modulações BPSK, 4-PAM, 16-QAM e 8-PSK, sem ruído. (a), (b), (c) e (d).

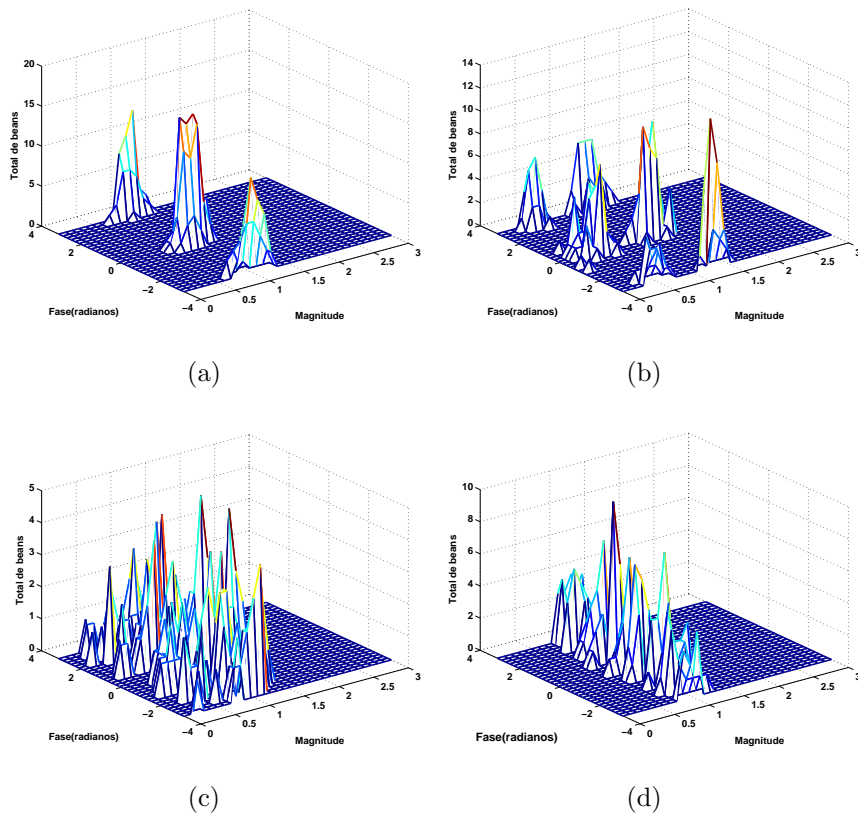


Figura 2.10: Exemplos de histogramas 2D com $N = 250$ símbolos (250 magnitudes e 250 fases) para modulações BPSK, 4-PAM, 16-QAM e 8-PSK, com $RSR = 15$ dB. (a), (b), (c) e (d).

2.7 Considerações Finais

Neste capítulo foi apresentada uma revisão sobre CAM. Primeiramente apresentamos o modelo de sinal, o qual é processado durante a tarefa de classificação. Uma vez obtido o sinal, necessita-se realizar um pré-processamento antes de executar a tarefa de classificação.

O pré-processamento é caracterizado por tarefas como redução do ruído, estimação da frequência, extração do período do símbolo, potência do sinal, equalização, etc. Nem todas essas tarefas são necessárias, dependendo da escolha do algoritmo de classificação.

A segunda etapa é caracterizada pela escolha da técnica de classificação. A literatura apresenta dois métodos: métodos baseados na teoria de detecção e métodos baseados no reconhecimento de padrão.

Embora os métodos baseados na teoria de detecção apresentem bons resultados, a complexidade computacional desses métodos é relativamente alta quando o número de parâmetros é desconhecido. Isso torna essa abordagem baseada em probabilidades muitas vezes impraticável para tarefas de classificação em tempo real.

Diferente dos métodos de teoria de detecção, os métodos baseados no reconhecimento de padrão utilizam algumas características para tarefa de classificação. Tais características são oriundas do primeiro subsistema da tarefa de classificação via reconhecimento de padrão: o extrator de características. Essas servem de entrada para o segundo subsistema: o classificador, o qual é responsável pela identificação da modulação. Foram apresentadas várias características tais como: amplitude instantânea, fase, frequência do sinal, a transformada de wavelet, e as estatísticas do sinal tais como momentos e cumulantes.

Buscou-se neste capítulo discutir a questão: “Qual o melhor método para tarefa de classificação automática?” Notoriamente a resposta para essa pergunta não é uma tarefa fácil, pois para que haja uma análise comparativa é necessário que o conjunto de modulações para tarefa de classificação seja o mesmo e, além disso, a maioria dos classificadores são propostos para trabalhar com parâmetros específicos desconhecidos. Desta forma, se não houver uma compatibilização desses fatores fica muito difícil de responder quem é o mais apropriado ou melhor. Com esta ressalva, o capítulo realizou extensiva descrição das propostas já apresentadas na literatura.

Além disso foi apresentado neste capítulo o extrator de parâmetros CSS que tem como característica a ordenação das magnitudes e fases dos símbolos recebidos e por último a concatenação destes. A proposta apresentou bons resultados quando utilizada com algoritmo MVS, os quais podem ser vistos em [13]. Tal trabalho serviu de motivação para implementação do CSS em *hardware* através do FPGA em [109], onde foi proposta uma arquitetura para clas-

sificação em tempo real.

Após a descrição do extrator de parâmetros CSS, o extrator HISTO2D foi apresentado. Diferente da primeiro extrator, o HISTO2D trabalha com histogramas bidimensionais de fase e magnitude a fim de distinguir as modulações. Uma vantagem do HISTO2D em relação ao CSS é que a abordagem baseada em histogramas não necessita realizar operações de ordenação, evitando assim, o aumento do custo computacional.

A seguir são discutidas as MVS, as quais permitem compor sistemas robustos quando usadas em conjunto com os extratores de parâmetros propostos.

Capítulo 3

Máquina de Vetores de Suporte - MVS

Neste capítulo, alguns tópicos referente a MVS são discutidos. Primeiramente apresenta-se uma introdução acerca do assunto. Depois são apresentados conceitos em relação à classificação de padrões linearmente separáveis e não-linearmente separáveis. Este aspecto não havia sido aprofundado na literatura concernente ao uso de MVS em CAM. Descreve-se também a classificação multi-classes voltada para MVS, o que consta como um segundo aspecto que esta tese aprofunda quando contrastada com o estado da arte de MVS em CAM. Por fim, realiza-se uma análise do custo computacional envolvido no processo de classificação utilizando o algoritmo MVS.

3.1 Introdução

MVS é um técnica de Reconhecimento de Padrão baseada na teoria do aprendizado estatístico, a qual foi desenvolvida por [73] com o objetivo de solucionar problemas de classificação de padrões.

Pelo fato do classificador MVS utilizar aprendizado supervisionado a partir de exemplos, necessita-se de uma introdução de alguns conceitos básicos e notação para referenciar entradas, saídas e etc.

Um classificador convencional é matematicamente um mapeamento $f : \mathbb{R}^D \rightarrow y = 1, \dots, Y$, onde D é a dimensão do vetor de entrada $\mathbf{x} \in \mathbb{R}^D$ e a saída $y \in \{1, \dots, Y\}$ onde Y representa o número possíveis de classe ou rótulos. [110]. As classes ou rótulos representam o fenômeno de interesse sobre o qual se deseja fazer provisões.

Ao treinar um classificador convencional utilizando aprendizado supervisionado é necessário um conjunto de treino $\Psi = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_T, y_T)\}$ contendo T exemplos (\mathbf{x}, y) [111],

ou seja, as amostras da etapa de treino são constituídas pelo conjunto de dados de entrada associadas às suas respectivas classes pré-rotuladas. Uma vez treinado o classificador f , o objetivo é classificar um novo conjunto de dados, conjunto de teste, não rotulado e disjuncto em relação ao um conjunto de dados de treino.

Inicialmente, o classificador MVS foi desenvolvido para resolver problemas de classificação binária [73], de maneira que dado um conjunto de treino Ψ , contendo exemplos de duas classes, uma MVS construirá um hiperplano ótimo como superfície de decisão a fim de que haja a separação dos exemplos das classes distintas de forma máxima. Essa separação máxima é caracterizada pela margem, determinada por alguns exemplos das classes pertencentes ao conjunto Ψ que são denominados de vetores de suporte. Considerando padrões linearmente separáveis, o hiperplano é uma superfície de decisão que separa os exemplos das classes de forma máxima. Entretanto, para padrões não-linearmente separáveis, tanto o hiperplano quanto os vetores de suporte são obtidos através de um resultado de uma função Φ denominado de kernel, a qual realiza um mapeamento apropriado sobre os dados do conjunto que os torna separáveis [112]. Na próxima seção será comentado os fundamentos que caracterizam a matemática do classificador MVS.

3.2 Teoria do Aprendizado Estatístico

Os fundamentos de um classificador MVS são constituídos pela Teoria de Aprendizado Estatístico (TAE). De acordo com [113], a TAE estabelece as ferramentas matemáticas que permitem obter um bom desempenho de um classificador, ou seja, essa técnica estatística possui os recursos necessários para encontrar um bom classificador a partir do conjunto de dados de treino.

Descreve-se o desempenho desejado de um classificador f como sendo uma taxa que indica o menor erro durante a etapa de treinamento cujo erro é mensurado pelo número de predições incorretas de f , a qual é definida de risco empírico $R_{emp}(f)$ [114], fornecido pela Equação 3.1:

$$R_{emp}(f) = \frac{1}{\mathcal{T}} \sum_{i=1}^{\mathcal{T}} c(f(\mathbf{x}_i), y_i) \quad (3.1)$$

Na Equação 3.1, \mathcal{T} representa o número de exemplos contidos no conjunto de entrada Ψ e $c(f(\mathbf{x}_i), y_i)$ representa uma função custo relacionando a previsão de f quando a saída deseja é y . Um tipo de função frequentemente utilizada em problemas de classificação é a 0/1

definida por $c(f(\mathbf{x}_i), y_i) = \frac{1}{2}|y - f(\mathbf{x})|$, a qual retorna o valor 0 se \mathbf{x} é classificado corretamente e o valor 1 caso contrário.

Na Figura 3.1 pode-se visualizar um exemplo, o qual apresenta funções que podem ser utilizadas para separar exemplos de um conjunto de treino contendo duas classes. Desta forma, é necessário encontrar um classificador, uma função, entre as diversas possíveis contidos em um conjunto \mathcal{F} , a qual classifique melhor esses dados. Isso pode ser feito através da minimização do $R_{emp}(f)$, o qual é representado na Figura 3.1 como a reta contínua.

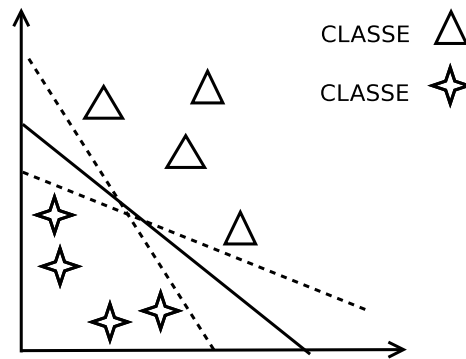


Figura 3.1: Hipóteses de hiperplano com hiperplano ótimo

Outra maneira de calcular a taxa de classificação incorreta de um classificador f utilizando dados de teste é através do Risco Esperado [113], o qual quantifica a capacidade de generalização de classificação. Nesse caso, os padrões de treinamento (\mathbf{x}_i, y_i) são gerados por uma distribuição de probabilidade de $P(\mathbf{x}_i, y_i)$, conforme é mostrado pela Equação 3.2:

$$R(f) = \int c(f(\mathbf{x}_i), y_i) dP(\mathbf{x}_i, y_i) \quad (3.2)$$

Analisando as duas formas, percebe-se que R_{emp} é facilmente obtido, ao contrário de $R(f)$, pois geralmente a distribuição de probabilidade P é desconhecida, e por esse motivo, não é possível minimizar o erro da Equação 3.2. Assim, pelo fato de ter-se somente os dados de treinamento, R_{emp} é normalmente calculado para inferir um classificador f que minimize o erro sobre os dados e conseqüentemente também sobre os dados de teste. Esse procedimento é chamado de princípio de minimização do risco empírico, o qual é um processo de indução com base nos dados de treinamento conhecidos [114].

Desta forma, pode-se interpretar a etapa de treino como um processo de estimação de um classificador (função) f . Caso não ocorra nenhuma restrição no processo de estimação de f , existe a possibilidade que o classificador obtenha um bom desempenho no conjunto de treinamento, entretanto pode não obter um bom desempenho no conjunto de dados de teste,

ou seja, os exemplos de treinamento não tinham informações suficientes para uma boa tarefa de aprendizado. Esse fenômeno é chamado *overfitting* [115]. Considerando essas possibilidades, a TAE possui características com o objetivo de prevenir a estimação de classificadores ruins através da aplicação de limites ao Risco Esperado de uma função, os quais podem ser empregados na escolha do classificador. Esses limites utilizam o conceito de dimensão VC.

Dado um conjunto \mathcal{F} de classificadores possíveis para uma tarefa de classificação onde $f \in \mathcal{F}$, define-se o conceito de dimensão VC como sendo o número máximo de exemplos contido no conjunto de treino, os quais podem ser divididos em duas classes pelos classificadores contidos em \mathcal{F} [116]. De modo a esclarecer essa definição, considere a Figura 3.2, onde nota-se que nos exemplos *a*, *b* e *c* é possível determinar retas capazes de separar os exemplos das classes distintas. Entretanto, quando existir 4 pontos no \mathbb{R}^2 , haverá cenários, os quais serão possíveis separar os dados das classes distintas e outros não, Figura 3.3. Quando o último ocorrer, será necessário obter funções de maior complexidade que retas para tentar contornar o problema [113].

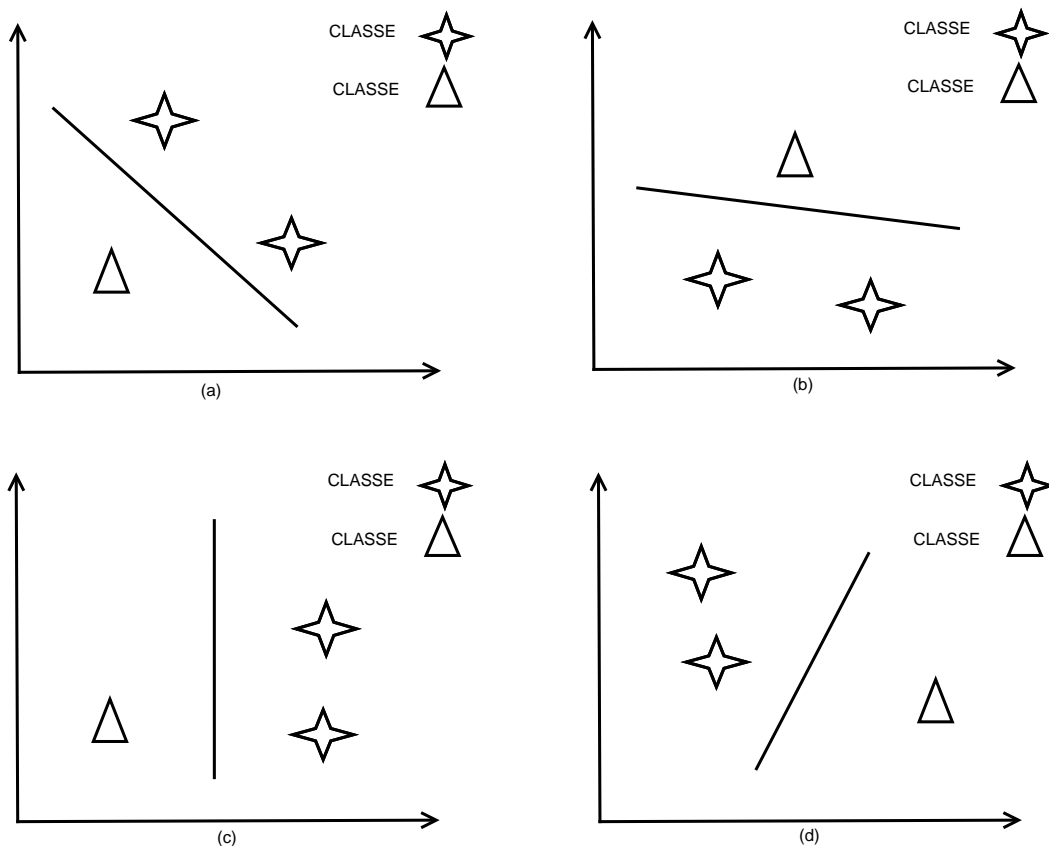


Figura 3.2: Separação dos dados de treino por meio de reta.

Desta forma, verifica-se que a dimensão VC de retas no caso \mathbb{R}^2 é três, pois este é o número de amostras, independente da classe que pertençam, que podem ser classificadas por

uma reta de maneira correta. A partir disso, a expressão matemática para dimensão VC para uma entrada qualquer com dimensão n pode ser vista na Equação 3.3.

$$VC(n) = n + 1 \quad (3.3)$$

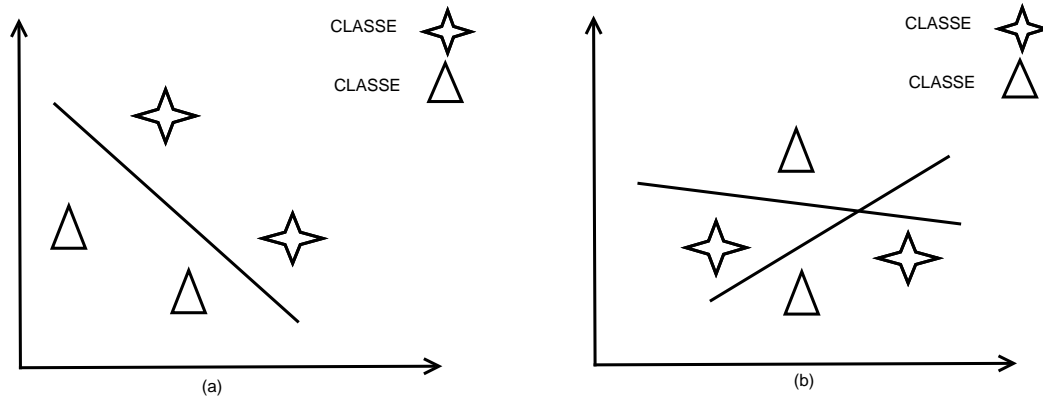


Figura 3.3: Separação dos 4 dados de treino por meio de reta.

Outro conceito da TAE relacionado a MVS é o conceito de margem. Como dito antes, MVS determina um hiperplano que divide os pontos de duas classes, o qual é constituído de pontos do conjunto de treinamento que satisfaz $\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$. A definição de margem de um classificador é caracterizada como a menor distância entre os exemplos do conjunto de treinamento e o hiperplano utilizado na separação dos exemplos das classes distintas [117]. Observando a Figura 3.4, verifica-se que há um classificador realçado pela cor preta representado pela reta contínua que maximiza a margem de classificação caracterizada pelos dois outros hiperplanos em tracejado, os quais passam por alguns pontos de ambas as classes. Esses pontos são chamados de vetores de suporte

3.3 Máquina de Vetores de Suporte Linear com Margens Rígidas

A aplicação de MVS com margens rígidas é aplicada quando um conjunto Ψ é considerado linearmente separável, de maneira que através de uma reta ou hiperplano é possível separar os exemplos dos padrões distintos contidos nele [118]. Considerando que os dados de treinamento são linearmente separáveis, uma MVS com margens rígidas determina um hiperplano que separa esses , o qual é representado matematicamente na Equação 3.4

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b = 0, \quad (3.4)$$

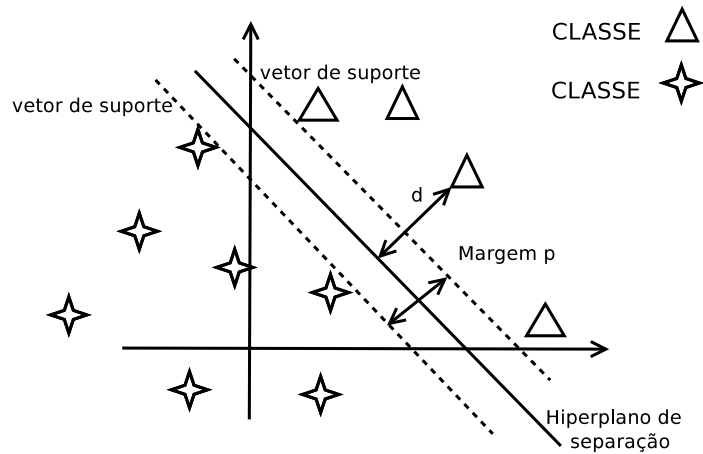


Figura 3.4: Identificação da margem e vetores de suporte.

onde $\langle \mathbf{w}, \mathbf{x} \rangle$ representa o produto interno entre os vetores \mathbf{w} e \mathbf{x} , em que $\mathbf{w} \in \Psi$ representa a orientação do hiperplano e $b \in \mathbb{R}$ é um termo “compensador”. Vale ressaltar que os valores tanto de \mathbf{w} quanto de b são determinados durante a etapa de treinamento.

A Equação 3.4 divide o espaço de entradas em duas regiões, levando à Equação 3.5. A interpretação desta nova Equação pode ser vista na Figura 3.5.

$$\begin{cases} \langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq +1 \text{ se } y_i = +1 \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1 \text{ se } y_i = -1 \end{cases} \quad (3.5)$$

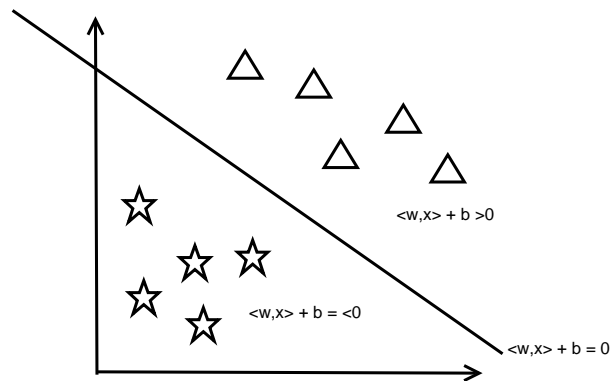


Figura 3.5: Hiperplano separando padrões linearmente separáveis.

Como nesta seção estamos trabalhando com padrões linearmente separáveis, assumi-se reescalar \mathbf{w} e b a fim de que os pontos mais próximos do hiperplano separador possam respeitar $|\langle \mathbf{w}, \mathbf{x} \rangle + b| = 1$ [114]. O resultado desta consideração é uma representação canônica do hiperplano, de modo a facilitar explicação na determinação do hiperplano ótimo.

A partir da consideração, mencionada anteriormente, essa implica na desigualdade 3.6 que caracteriza os classificadores separando as classes do conjunto de treinamento com uma margem positiva. A Figura 3.5 ilustra a interpretação geométrica desse sistema, onde não há pontos entre $\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$ e $\langle \mathbf{w}, \mathbf{x} \rangle + b = \pm 1$. Devido essa suposição, as MVS obtidas são denominadas de MVS com margens rígidas [119].

$$\begin{cases} \langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq +1 \text{ se } y_i = +1 \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1 \text{ se } y_i = -1 \\ i = 1, \dots, N \end{cases} \quad (3.6)$$

A fim de descobrir a distância entre os hiperplanos $\langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq +1$ se $y_i = +1$ e $\langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1$ se $y_i = -1$, a Figura 3.6 ilustra o ponto \mathbf{x}_1 sobre a reta $\langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq -1$ e o ponto \mathbf{x}_2 que está sobre a reta $\langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq +1$, formando a Equação 3.7. Além disso, pode-se perceber que \mathbf{x}_1 intercepta a reta perpendicular a \mathbf{x}_2 .

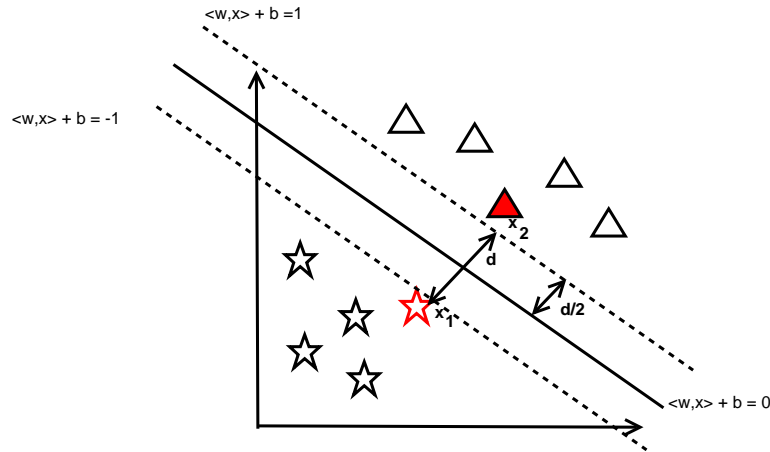


Figura 3.6: Exemplo de conjunto de dados linearmente não separável

$$\begin{cases} \langle \mathbf{w}, \mathbf{x}_1 \rangle + b = -1 \\ \langle \mathbf{w}, \mathbf{x}_2 \rangle + b = 1 \end{cases} \quad (3.7)$$

Desta forma, a partir do sistema da Equação 3.7, obtém-se a Equação 3.8.

$$\mathbf{w}(\mathbf{x}_2 - \mathbf{x}_1) = 2 \quad (3.8)$$

Pelo fato que \mathbf{w} e $\mathbf{x}_2 - \mathbf{x}_1$ serem ortogonais ao hiperplano separador, esses são paralelos entre si, podendo desta forma deduzir a Equação 3.9

$$|\mathbf{w}(\mathbf{x}_2 - \mathbf{x}_1)| = \|\mathbf{w}\| \times \|\mathbf{x}_2 - \mathbf{x}_1\|, \quad (3.9)$$

onde $\|\cdot\|$ representa a norma de um vetor. Assim, realizando a substituição da Equação 3.8 em 3.9 é obtida a Equação 3.10.

$$\|\mathbf{x}_2 - \mathbf{x}_1\| = \frac{2}{\|\mathbf{w}\|} \quad (3.10)$$

A distância ou margem entre os hiperplanos $\langle \mathbf{w}, \mathbf{x}_1 \rangle + b = -1$ e $\langle \mathbf{w}, \mathbf{x}_1 \rangle + b = +1$ é obtida através da norma $\|\mathbf{x}_2 - \mathbf{x}_1\|$, de modo que pode-se afirmar que a distância entre os hiperplanos é $2/\|\mathbf{w}\|$, Equação 3.10. A análise deste resultado tem uma grande importância na construção (por exemplo, aprendizagem) no projeto de um classificador com margem máxima. Isso levará a resolução de um problema programação quadrática, que será apresentado em breve.

Pode-se observar que a Equação 3.10 apresenta um resultado bastante interessante, pois a minimização da norma do vetor de pesos do hiperplano $\|\mathbf{w}\| = \sqrt{(\mathbf{w}^T \mathbf{w})}$ é equivalente a maximização da margem. Assim, o problema de aprendizagem está sujeito a minimização da Equação 3.11 sob as restrições introduzidas na Equação 3.12.

$$\min(\mathbf{w}) = \frac{1}{2} \mathbf{w}^2 \quad (3.11)$$

sujeito a

$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 \quad (3.12)$$

para $i = 1, \dots, N$. A restrição deste problema é resultado da combinação da Equação 3.6.

Nota-se que a definição acima é um problema de otimização quadrática clássico com restrições de inequações, o qual é resolvido através da função de Lagrange [117], caracterizada pela Equação 3.13 e definida em termos da variáveis \mathbf{w} , b e α , onde α_i são denominados de multiplicadores de Lagrange. Analisando a Equação 3.13, verifica-se que esta função é matematicamente expressa pela soma da função objetivo e uma combinação linear da função de restrição [120].

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^2 - \sum_{i=1}^T \alpha_i \{y_i[\langle \mathbf{w}, \mathbf{x}_i \rangle + b] - 1\} \quad (3.13)$$

A solução deste problema, a partir de agora, equivale minimizar $L(\mathbf{w}, b, \alpha)$ em relação \mathbf{w} e b e maximizar $L(\mathbf{w}, b, \alpha)$ em relação a α cuja solução pode ser feita tanto em um espaço primal (que é o espaço de parâmetros \mathbf{w} e b) ou no espaço dual (que é o espaço dos multiplicadores de Lagrange α_i). A segunda abordagem foi escolhida, pois fornece uma visão melhor da resolução deste problema, onde o mínimo do lagrangeano, usando as condições de KKT (*Karush-Kuhn-Tucker*), com respeito \mathbf{w}_0 e b_0 é obtido através da resolução de igualdades apresentadas nas Equações 3.14 e 3.15.

$$\frac{\partial}{\partial b} L(\mathbf{w}_0, b_0, \alpha) = 0 \Rightarrow \sum_{i=1}^{\mathcal{T}} \alpha_i y_i = 0 \quad (3.14)$$

$$\frac{\partial}{\partial \mathbf{w}_0} L(\mathbf{w}_0, b, \alpha) = 0 \Rightarrow \mathbf{w}_0 - \sum_{i=1}^{\mathcal{T}} \alpha_i \mathbf{x}_i y_i = 0 \Rightarrow \mathbf{w}_0 = \sum_{i=1}^{\mathcal{T}} \alpha_i \mathbf{x}_i y_i \quad (3.15)$$

Fazendo a substituição de 3.14 e 3.15 dentro da função de Lagrange primal 3.13, temos a função de Lagrange de variáveis duais $L_d(\alpha)$ dual:

$$\max_{(\alpha)} L_d(\alpha) = \sum_{i=1}^{\mathcal{T}} \alpha_i - \frac{1}{2} \sum_{i=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{T}} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (3.16)$$

com restrições:

$$\begin{cases} \alpha_i \geq 0, \forall_i = 1, \dots, N \\ \sum_i^{\mathcal{T}} \alpha_i y_i = 0 \end{cases} \quad (3.17)$$

Assim para encontrar o hiperplano ótimo, $L_d(\alpha)$ deve ser maximizada com relação $\alpha_i > 0$, respeitando as restrições. Além dessas restrições, a solução ótima deve também respeitar as condições complementares de KKT, de acordo com a Equação 3.18

$$\alpha_i (y_i (\langle \mathbf{w}_0, x_i \rangle + b_0) - 1) = 0 \quad (3.18)$$

Note-se que $L_d(\alpha)$ é expressa em termos dos dados de treinamento e depende somente do produto interno dos padrões de entrada $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$. As soluções ótimas α_{0i} do problema de otimização acima determinam os parâmetros \mathbf{w}_0 e b_0 a partir das Equações 3.15 e 3.18 respectivamente. Há duas observações importantes a respeito do calculo de \mathbf{w}_0 : primeira, o vetor de peso ótimo \mathbf{w}_0 é obtido na Equação 3.15 como uma combinação linear dos dados do conjunto de treinamento e segunda, \mathbf{w}_0 é calculado usando apenas os valores $\alpha_i > 0$, os quais são chamados de vetores de suporte (VS) [119]. Esses dados são considerados os mais importantes do conjunto de treino Ψ , pois somente esses participam na determinação do hiperplano separador [113].

Desta forma, um exemplo \mathbf{x}' desconhecido pode ser classificado através de um classificador $f(\mathbf{x}')$ dado por:

$$f(\mathbf{x}') = \text{sgn}\left(\sum_{x_i \in VS} \alpha_i y_i \langle \mathbf{x}_i, \mathbf{x}' \rangle + b_0\right) = \text{sgn}(\langle \mathbf{w}_0, x' \rangle + b_0) \quad (3.19)$$

A Equação 3.19 representa o hiperplano com maior margem que separa os dados de duas classes durante a etapa de treinamento da MVS, a qual representa entre os diversos classificadores possíveis, o classificador que possui maior capacidade de generalização de acordo com a TAE.

3.4 Máquina de Vetores de Suporte Linear com Margens Suaves

A aplicação de MVS linear com margens rígidas é válida somente para dados que respeitam a restrição da Equação 3.12, de modo que tais problemas são raros na prática, pois geralmente essa restrição não será obedecida, ou seja, na maioria dos casos haverá exemplos de treinamento entre as margens de separação das classes, os quais não podem ser classificados de maneira correta. Tal cenário pode ser visto na Figura 3.7.

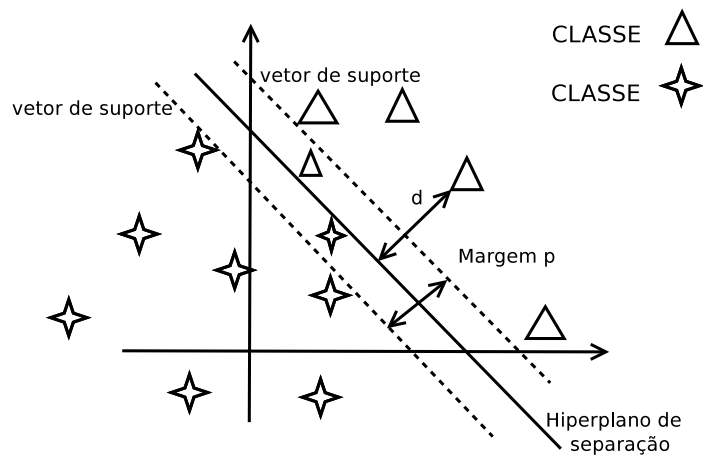


Figura 3.7: Exemplo de conjunto de dados linearmente não separável

Quando esses casos ocorrem, nota-se a impossibilidade de encontrar um hiperplano, o qual separe os dados do conjunto de treinamento sem erro de classificação [120]. Para esse problema, é necessário modificar o algoritmo apresentado na Seção 3.3, de forma que alguns dados alguns dados permaneçam entre os hiperplanos e também que ocorra alguns erros de classificação a fim de evitar o superajustamento (emphoverfitting). Assim, para lidar com esses casos, é necessário permitir que dados x_i de treinamento possam violar as restrições na determinação do hiperplano ótimo, Figura 3.8, através da inclusão de variáveis de folga u_i definida na Equação 3.20.

$$y_i(\langle \mathbf{w}, \mathbf{x} \rangle + b) \geq 1 - u_i, u_i \geq 0 \quad (3.20)$$

A inclusão das variáveis de folga u_i suaviza (*soft*) a margem do classificador linear, de modo que alguns dados do conjunto de treinamento permaneçam entre os hiperplanos de separação e também ocorra alguns erros de classificação. A soma de cada termo u_i representa um limite na quantidade de erros de treinamento. Desta forma, para se obter o hiperplano

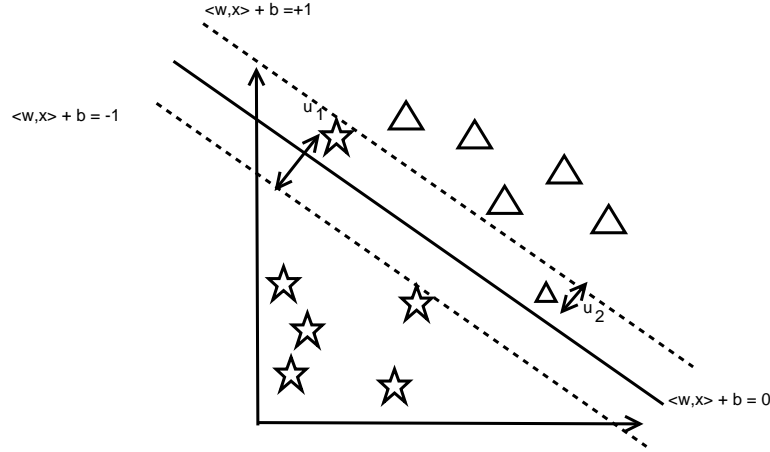


Figura 3.8: Exemplo de conjunto de dados linearmente não separável. Equação 3.12

canônico é necessário além de maximizar a margem, a qual é obtida através da minimização de \mathbf{w} , necessita-se minimizar o erro sobre os dados de treinamento, de modo que a Equação 3.11 é caracterizada por uma nova reformulação:

$$\text{Minimizar } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{\mathcal{T}} u_i, \quad (3.21)$$

onde o termo $C > 0$ é definido pelo usuário e estabelece um *trade-off* entre a maximização da margem e a minimização de erros [117].

A solução para o problema 3.21, sujeito às restrições 3.20 é caracterizada pelo ponto de sela do Lagrangeano Primal $L_p(\mathbf{w}, b, u, \boldsymbol{\alpha}, \boldsymbol{\beta})$:

$$L_p(\mathbf{w}, b, u, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2} \mathbf{w}^2 + C \left(\sum_{i=1}^{\mathcal{T}} u_i \right) - \sum_{i=1}^{\mathcal{T}} \alpha_i \{ y_i [\langle \mathbf{w}, \mathbf{x}_i \rangle + b] - 1 + u_i \} - \sum_{i=1}^{\mathcal{T}} \beta_i u_i, \quad (3.22)$$

onde α_i e β_i são os multiplicadores de Lagrange. Para encontrar a solução, deve-se, novamente, minimizar em relação \mathbf{w} , b e u e maximizar com respeito aos valores não negativos de α_i e β_i . A resolução deste problema pode ser realizada tanto no espaço primal quanto no espaço dual (espaço caracterizado pelos multiplicadores de Lagrange α_i e β_i). Como anteriormente, a solução no espaço dual é encontrada usando as condições padrões para ótimo de um função de restrição:

$$\frac{\partial L}{\partial \mathbf{w}_0} = 0, \text{ ou } \mathbf{w}_0 = \sum_{i=1}^{\mathcal{T}} \alpha_i \mathbf{x}_i y_i \quad (3.23)$$

$$\frac{\partial L}{\partial b_0} = 0, \text{ ou } \sum_{i=1}^{\mathcal{T}} \alpha_i y_i = 0 \quad (3.24)$$

$$\frac{\partial L}{\partial u_{i0}} = 0, \text{ ou } \alpha_i + \beta_i = C, \quad (3.25)$$

e as condições complementares de KKT

$$\alpha_i \{y_i [\langle \mathbf{w}, x_i \rangle + b] - 1 + u_i\} = 0, \quad i = 1, \dots, N \quad (3.26)$$

No intuito de encontrar o hiperplano ótimo, o Lagrangeano dual $L_d(\alpha)$ precisa ser maximizado com relação as restrições 3.28

$$\max_{(\alpha)} L_d(\alpha) = \sum_{i=1}^{\mathcal{T}} \alpha_i - \frac{1}{2} \sum_{i=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{T}} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (3.27)$$

com restrições:

$$\begin{cases} 0 \leq \alpha_i \leq C, \forall_i = 1, \dots, N \\ \sum_i^{\mathcal{T}} \alpha_i y_i = 0 \end{cases} \quad (3.28)$$

Pode-se notar a formulação do problema acima é muito semelhante ao problema de otimização envolvendo MVS com margens rígidas, diferindo na restrição para cada ponto α_i (multiplicadores de Lagrange) cujo valores são limitados por C [120]. Mais importante, o problema de aprendizagem é expresso somente em relação aos multiplicadores de Lagrange desconhecidos α_i e entradas x_i e saídas y_i conhecidas. Além disso, a otimização não depende exclusivamente das entradas \mathbf{x}_i , as quais podem apresentar uma dimensão grande, mas depende do produto escalar dos vetores de entrada \mathbf{x}_i [117].. Assim, a expressão para o classificador $f(\mathbf{x})$ com margem suave é caracterizada pela Equação 3.19, ou seja, a mesma para o caso onde os padrões são linearmente separáveis.

3.5 Máquina de Vetores de Suporte Não Linear

As máquinas de vetores de suporte discutidas até agora funcionam apenas quando o padrão é separável linearmente. Entretanto existem cenários, os quais não é possível separar os dados das classes distintas através de um hiperplano, ou seja, características de linearidade no conjunto de treinamento não são encontradas. Para exemplificar esse problema, pode-se visualizar a Figura 3.9, que ilustra dados unidimensionais, onde não é possível separar os dados das classes distintas através de um hiperplano.

A ideia básica para resolver esse problema é realizar um mapeamento dos dados de entrada $\mathbf{x}_i \in \mathbb{R}^D$ para um novo espaço de maior dimensão, denominado de espaço de características $\mathcal{G}(\mathbf{z} = \Phi(\mathbf{x}))$ [121], onde Φ representa o mapeamento $\mathbb{R}^D \rightarrow \mathbb{R}^D$. A escolha

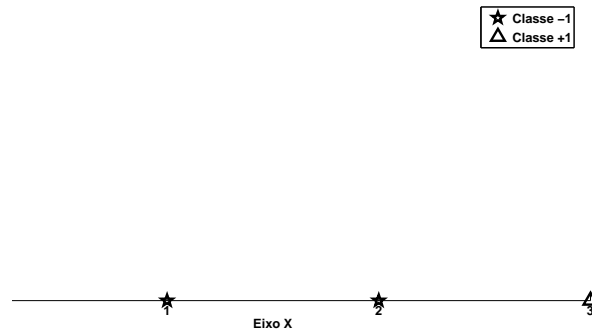


Figura 3.9: Exemplo de padrão não linear

apropriada de Φ torna os dados \mathbf{x}_i linearmente separáveis no novo espaço de características, de modo que os dados das classes distintas possam ser separados por uma MVS linear. Para exemplificar essa solução, aplica-se a seguinte função $\Phi(x_i) = (x_i, x_i^2)$ nos dados de entrada da Figura 3.9. Tal função converte o conjunto de entrada de \mathbb{R} para \mathbb{R}^2 , tornando os dados linearmente separáveis, como pode ser visto na figura 3.10.

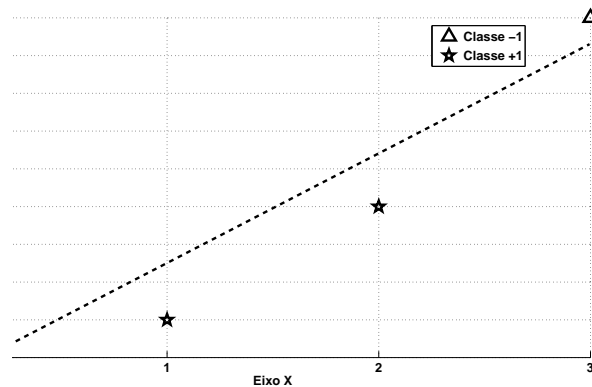


Figura 3.10: Utilização de uma função não linear sobre os dados de entrada

Entretanto, pode não ser apropriado executar a MVS nesse novo espaço de características, pois à medida que aumenta a dimensão de \mathbf{x}_i , o custo computacional da operação de mapeamento envolvida também aumenta. Esta explosão na dimensionalidade pode ser evitada, notando que o problema de otimização quadrática dado pelas Equações 3.16 e 3.27, tanto quanto a Equação 3.19 final do classificador, os dados de treinamento \mathbf{x}_i somente aparecem na forma de produto escalares $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$. Esses produtos são substituídos pelos produtos escalares $\langle \mathbf{z}_i, \mathbf{z}_j \rangle = [\Phi_1(\mathbf{x}), \Phi_2(\mathbf{x}), \dots, \Phi_n(\mathbf{x})][\Phi_1(\mathbf{x}_i), \Phi_2(\mathbf{x}_i), \dots, \Phi_n(\mathbf{x}_i)]^T$ no espaço de característica \mathcal{G} ,

e esta última é expressa usando uma função denominada de kernel

$$\mathcal{K}(x_i, x_j) = \langle \mathbf{z}_i, \mathbf{z}_j \rangle = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle. \quad (3.29)$$

Note que função kernel $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$ é uma função no espaço de entrada. Então, a vantagem básica no uso da função kernel $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$ é evitar que seja feito o mapeamento $\Phi(\mathbf{x})$. Como alternativa, os produtos escalares requeridos no espaço de característica $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$ são calculados diretamente, calculando $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$ para um dado conjunto de treinamento no espaço de entrada.

Para demonstrar esse conceito aplica-se um exemplo, o qual utiliza dois vetores $\mathbf{x}_i = (x_{i1}, x_{i2})$ e $\mathbf{x}_j = (x_{j1}, x_{j2})$ e kernel polinomial $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle)^d$. Utilizando o kernel polinomial com $d = 2$ tem-se:

$$\mathcal{K}(x_i, x_j) = (\langle x_i, x_j \rangle)^2 \quad (3.30)$$

$$= ((x_{i1}y_{j1}) + (x_{i2}y_{j2}))^2 \quad (3.31)$$

$$= x_{i1}^2 y_{j1}^2 + 2x_{i1}x_{i2}y_{j1}y_{j2} + x_{i2}^2 y_{j2}^2 \quad (3.32)$$

$$= (x_{i1}^2, \sqrt{2}x_{i1}, x_{i2}, x_{i2}^2) \cdot (x_{j1}^2, \sqrt{2}x_{j1}, x_{j2}, x_{j2}^2) \quad (3.33)$$

$$= \phi(x_i) \cdot \phi(x_j) \quad (3.34)$$

Nota-se nas equações acima, a desnecessidade de mapear os vetores de entrada \mathbf{x}_i e \mathbf{x}_j para um espaço de maior dimensão e realizar o produto interno entre esses, já que implicitamente a função kernel realiza essas etapas, as quais são comprovadas na Equação 3.33 que apresenta o produto interno entre dos vetores \mathbf{x}_i e \mathbf{x}_j mapeados em \mathbb{R}^3 através do kernel polinomial.

Há uma variedade de funções de kernel que podem ser aplicados numa tarefa de classificação utilizando MVS. Cada uma dessas funções constroem diferentes superfícies (*hypersurface*) de decisões não lineares no espaço de entrada. Caso o problema de classificação apresente dados de entrada \mathbf{x}_i linearmente separáveis, não é necessário mapear o espaço de entrada em um espaço de dimensão maior. Neste tipo de cenário, utiliza-se o kernel Linear:

$$\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle. \quad (3.35)$$

Entretanto, como discutido antes, há maioria dos problemas possuem dados que não são linearmente separáveis, de forma que é necessário mapear esses dados para um espaço de dimensão maior. A primeira opção é o kernel polinomial:

$$\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + 1)^d, \quad (3.36)$$

onde $d \in \mathcal{T}$ indica o grau do polinômio. Nota-se quando $d = 1$, o kernel polinomial torna-se igual o kernel linear mais 1, de modo que ajustando o termo b na função de decisão, é equivalente ao kernel linear.

Outra opção de kernel que pode ser utilizado quando os dados não são linearmente separáveis é o kernel Gaussiano ou RBF (*Radial Basis Function*):

$$\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \quad (3.37)$$

onde σ representa a variância, a qual é determinada pelo usuário.

Desta forma, finalmente pode-se analisar o processo de aprendizagem para classificadores não lineares cuja teoria segue a mesma ideia de encontrar um hiperplano de separação ótimo, sendo agora no espaço de características. Agora, em um espaço de característica, o Lagrangeano dual dados nas Equações 3.16 e 3.27, é

$$\max_{(\alpha)} L_d(\alpha) = \sum_{i=1}^{\mathcal{T}} \alpha_i - \frac{1}{2} \sum_{i=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{T}} \alpha_i \alpha_j y_i y_j \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) \quad (3.38)$$

sujeito as restrições

$$\begin{cases} \alpha_i \geq 0, \forall_i = 1, \dots, N \\ \sum_i^{\mathcal{T}} \alpha_i y_i = 0 \end{cases} \quad (3.39)$$

Deve-se atentar, que no caso mais geral, quando mesmo no novo espaço de característica, os dados são linearmente não separáveis, o classificador de margem suave não linear terá o problema de otimização quadrática dada pela Equação 3.38, porém sujeito as seguintes restrições:

$$\begin{cases} 0 \leq \alpha_i \leq C, \forall_i = 1, \dots, N \\ \sum_i^{\mathcal{T}} \alpha_i y_i = 0 \end{cases} \quad (3.40)$$

de modo que a função de classificação para um novo ponto x' sob restrição é dada por 3.41:

$$f(\mathbf{x}') = \text{sgn}\left(\sum_{x_i \in VS} \alpha_i y_i \mathcal{K}(\mathbf{x}_i, \mathbf{x}') + b_0\right) \quad (3.41)$$

3.6 Classificação Multi-classes

Como mencionado na introdução deste capítulo, MVS é um classificador binário utilizado para classificação de dados entre duas classes distintas, que podem ser denominadas de positiva e negativa respectivamente. Entretanto a maior parte dos problemas de classificação

envolvem $y > 2$ classes. Embora essa técnica seja binária, isso não inviabiliza sua aplicação em problemas de classificação multi-classes através de duas técnicas: Um Contra um (em inglês, *One Against One*) e Um contra Todos (em inglês, *One Against All*) [112].

“Um Contra Todos” é uma técnica para resolver problemas multi-classes proposto por [122]. Essa baseia-se na existência de y classes, as quais são particionadas em dois grupos: O primeiro grupo é constituído pelos exemplos de uma única classe e o segundo é constituído pelos exemplos das classes restantes. A Figura 3.11 ilustra bem essa característica, onde observa-se todas as possibilidades possíveis de partições para um problema envolvendo $y = 3$ classes. Desta forma para y classes são construídos f_y classificadores binários, de modo que a classificação para um exemplo x será a classe y representada por um dos classificadores f_y que obteve o valor máximo. Formalmente define-se pela método na seguinte expressão 3.42:

$$F(x) = \arg \max_{1, \dots, y} (f_y(x)) \quad (3.42)$$

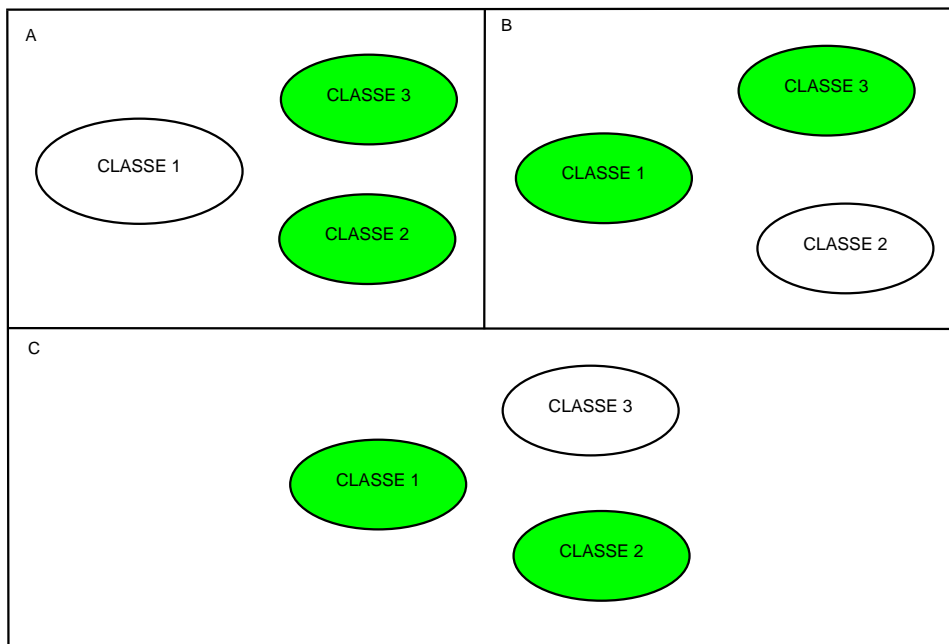


Figura 3.11: Técnica Um Contra Todos.

Uma vantagem da técnica Um Contra Todos é a quantidade reduzida de classificadores construídos para resolver problemas com número de classes $y > 2$, de modo que torna a tarefa de classificação mais rápida em casos onde o problema envolve poucas classes. Em contrapartida, a técnica apresenta algumas desvantagens. A primeira é que, durante a etapa de treinamento, a quantidade de memória requerida pode ser alta quando a base de dados de treino for grande, causando problemas de memória. Segunda, suponhamos que há Y

classes em um problema de classificação, cada classe possui o mesmo número de exemplos de treinamento. Desta forma, durante a etapa de treinamento, a razão do número de amostras do primeiro grupo constituído por uma única classe em relação ao segundo constituído pelas classes restantes será: $\frac{1}{Y-1}$, indicando que o número de amostras de treinamento não estão balanceadas.

A técnica Um Contra um é caracterizada por ser uma ideia bem simples e eficiente para o tratamento de problemas multi-classes, a qual é caracterizada da seguinte maneira: para cada par de classes de um problema com $y > 2$ classes é construído um classificador binário, onde cada classificador é construído pelos elementos do par de classes envolvidas, obtendo um total de $y(y-1)/2$ classificadores [123]. A Figura 3.12 ilustra o exemplo de um problema com

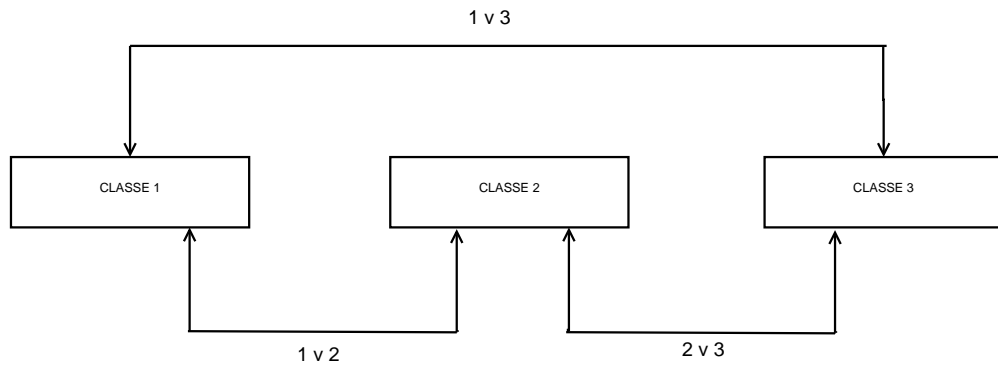


Figura 3.12: Técnica Um contra um.

$y = 3$ classes. Verifique que são construídos três classificadores binários envolvendo os pares das classes envolvidas. Uma vez construídos os f_y classificadores, a decisão de classificação é determinada através de um esquema de votação, onde cada classificador atribui uma das duas classes para um exemplo desconhecido. A classe mais votada determina a classe para o exemplo desconhecido. Caso haja um empate, uma estratégia de desempate pode ser adaptada. Uma estratégia comum é selecionar de maneira aleatória uma das classes que está vinculada ao ponto desconhecido.

O número de classificadores criado por esta técnica geralmente é muito maior em relação a técnica Um contra todos, entretanto o número de vetores de treinamento requerido é muito menor durante a etapa de treinamento, pois envolve somente os dados referentes ao par de classes de envolvidas, de modo que exige uma quantidade de memória menor comparado a técnica Um contra todos. Uma desvantagem desta técnica é o aumento do número dos classificadores com o aumento do número de classes. Por exemplo, um problema envolvendo sete classes distintas serão necessários vinte e um classificadores para solucionar o problema.

3.7 Custo Computacional

Estimar o custo computacional não é uma tarefa simples, porém é interessante tentar estimar o custo dos kernels utilizados neste trabalho durante a etapa de teste. É possível também estimar durante a etapa de treinamento, entretanto é geralmente menos importante devido essa etapa representar a configuração dos parâmetros do classificador e os resultados de classificação não serem utilizados.

A avaliação do custo computacional foi feita através do índice de complexidade computacional (ICC), que analisa diversas operações lógicas e matemáticas com seu desempenho normalizado em relação ao tempo necessário para execução de uma operação [124]. O estudo sobre ICC consiste sobre a avaliação do tempo necessário para executar uma determinada operação 10^{12} vezes em diversas máquinas com arquiteturas, hardwares e softwares diferentes. Esses tempos são normalizados de modo que possam ser aplicados a qualquer máquina utilizada. A média dos resultados de todas as máquinas foi feita e o ICC calculado. Conforme o trabalho [125], percebeu-se que as razões entre o ICC dos diversos procedimentos eram condizentes com a razão entre seus tempos de execução. Desta forma, o calculo do ICC é dada pela Equação 3.43:

$$ICC = \sum_{op} Q_{op} W_{op}, \quad (3.43)$$

onde Q_{op} representa quantidade de ocorrências de cada operação e W_{op} indica o peso da operação [124], conforme a Tabela 3.1.

O ICC assume que o peso da operação exponencial é $W_{exp} = 25$, enquanto adição, multiplicação e subtração tem peso $W = 1$. Assim, para classificar um exemplo usando a Equação 3.41 com kernel RBF, a estimativa é caracterizada por:

$$\begin{aligned} ICC_{RBF} &= N_{sv} [W_{exp} + W(3D + 1)] + W \\ &= N_{sv}(26 + 3D) + 1. \end{aligned} \quad (3.44)$$

Em geral, se o kernel é não-linear, o ICC escala linearmente com o número de vetores de suporte N_{sv} como na Equação (3.44). Entretanto, para uma MVS com kernel linear, o produto interno entre \mathbf{x}' e \mathbf{x}_i da Equação 3.41 pode ser reformulado como um *perceptron*, de modo que o somatório $\mathbf{w} = \sum_{x_i \in sv} \alpha_i y_i \mathbf{x}_i$ é pré-calculado, resultando na $f(\mathbf{x}') = \langle \mathbf{w}, \mathbf{x}' \rangle + b$. Portanto, para MSV lineares, o custo é reduzido para

$$ICC_{linear} = W(D + D) = 2D, \quad (3.45)$$

o qual não depende do número de SV.

Tabela 3.1: Pesos associados com operações matemáticas.

Operação	Exemplo	Peso
Adições	$()=()+()$	1
Multiplicações	$()=()*()$	1
Multiplicação-adição	$()=()*()+()$	1
Dados móveis	float, int	1
Operações lógicas	shift, módulo	1
Divisões	$()=()/()$	18
Raiz quadrada	$()=sqrt()$	25
Funções transcendentais	seno, log	25
Teste aritmético	se, se-então-senão	2
Adição-comparação-seleção	decodificação Viterbi	6

3.8 Considerações Finais

Neste capítulo, os conceitos básicos sobre a teoria de MVS foram descritos para problemas de classificação, os quais são baseados na teoria do Aprendizado Estatístico.

Uma das principais características desse classificador é a sua robustez diante de dados de grandes dimensões. Além dessa característica, a aplicação de funções kernel em dados não linearmente separáveis torna o classificador eficiente em espaços de alta dimensão devido permitir a construção de hiperplanos simples, deixando, do ponto de vista computacional, tratável.

Embora, originalmente, as MVS sejam formuladas para problemas de classificação binária, há métodos (Um Contra Todos e Um contra Um) que permitem resolver problemas de multi-classes. No método Um Contra Todos, a classificação é baseada na saída do valor máximo de acerto entre as y MVS. Enquanto o método Um Contra Um, a classificação é baseada em um sistema de votação, de modo que, a solução é dada pela classe que recebeu mais votos.

Por último, foi discutido a avaliação do custo computacional baseado no número de operações matemáticas necessárias para realizar a classificação. O resultado dessa discussão foi que a tarefa de classificação utilizando kernel linear não depende do número dos vetores de suporte. Tal independência diminui consideravelmente o custo computacional, pois tais vetores

são reunidos em um *perceptron*. Diferentemente do kernel linear, o kernel RBF depende do número de vetores de suporte a fim de realizar a classificação de um dado desconhecido. Essa característica faz aumentar consideravelmente, dependendo do cenário, o custo computacional de classificação quando comparado com kernel linear.

Capítulo 4

Sistema de CAM proposto e Avaliação de Desempenho

Nesta parte do trabalho, apresenta-se uma nova proposta de CAM baseada em histograma, a qual juntamente com MVS realiza o processo de classificação. Além disso, as descrições dos experimentos para CAM realizadas juntamente com os respectivos resultados são apresentadas. Por fim, analisa-se o custo computacional dos extratores de parâmetros que utilizam MVS.

4.1 Proposta da Tese: Extrator de Parâmetros HISTO

O HISTO é o extrator de parâmetros proposto nesta Tese. Diferentemente do CSS, visto no Capítulo 2, que possui um custo computacional relativamente alto devido à ordenação da magnitude e da fase dos símbolos recebidos, o extrator HISTO é mais eficiente e foi concebido para evitar esse procedimento de ordenação. Ao invés de ordenar, o extrator HISTO calcula dois histogramas com B raias (faixas ou *bins*): um histograma para magnitude e outro para fase. Com esses dois histogramas, oriundos do vetor \mathbf{r} de N símbolos recebidos, a concatenação desses dois vetores é realizada a fim de compor vetor de parâmetros \mathbf{x} que serve de entrada ao classificador.

No HISTO, os valores da distribuição das ocorrências são usados como os parâmetros \mathbf{x} de entrada do classificador. Para uma melhor compreensão, veja a Figura 4.1.

A Figura 4.1 ilustra exemplos de constelações das modulações BPSK, 4-PAM, 16-QAM e 8-PSK e respectivos histogramas para uma situação ideal, sem ruído. Cada gráfico foi gerado

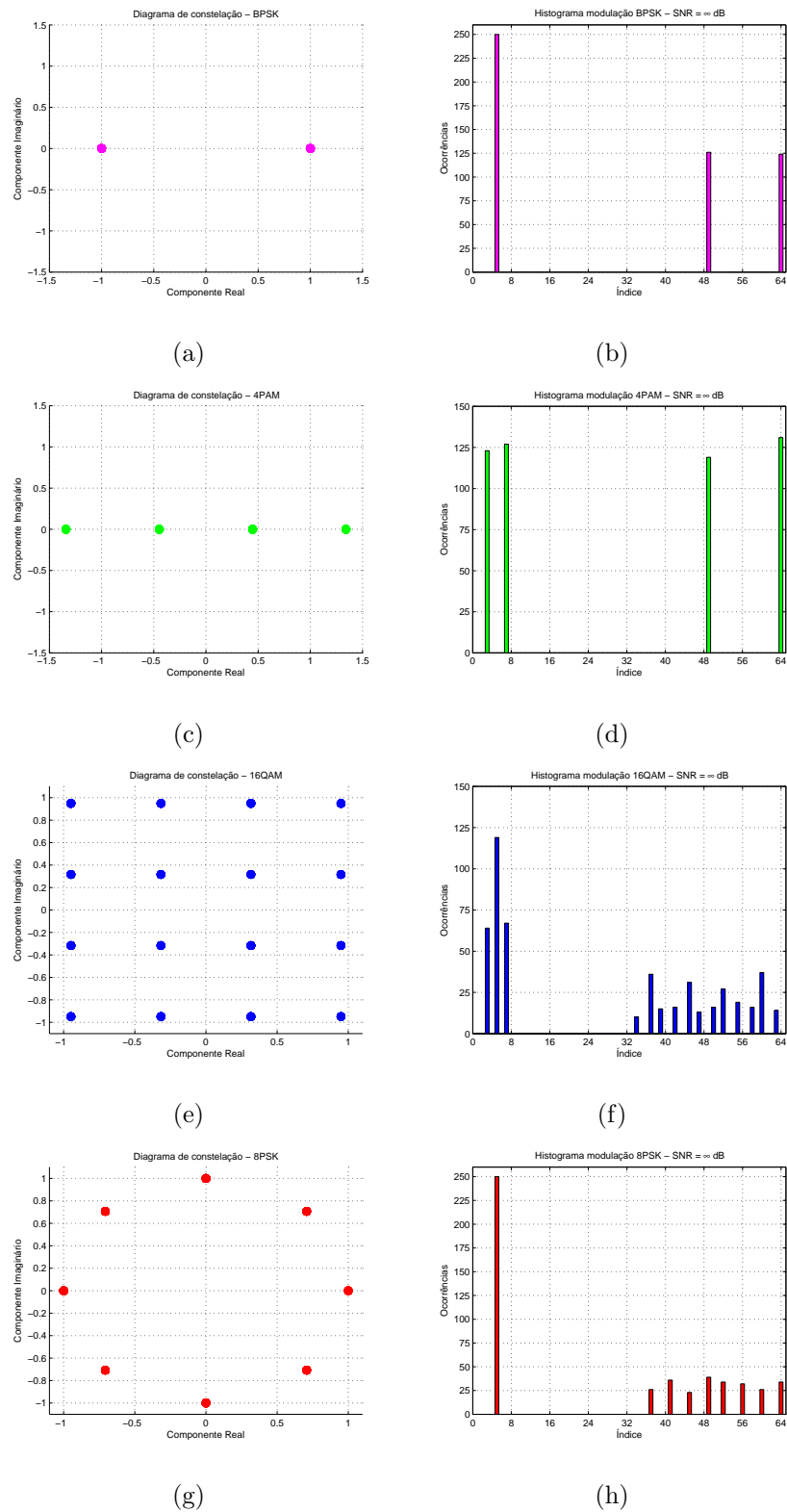


Figura 4.1: Exemplos de histogramas com $N = 250$ símbolos (250 magnitudes e 250 fases) para modulações BPSK, 4-PAM, 16-QAM e 8-PSK, sem ruído. (a), (c), (e) e (g) Constelações. (b), (d), (f) e (h) Histogramas

utilizando-se $N = 250$ símbolos da modulação. Para o cálculo dos histogramas, $B = 32$ raias foram utilizadas para o módulo, concatenado com mais 32 raias utilizadas para as fases, gerando um vetor com $D = 64$ elementos. Desta forma os 32 primeiros elementos representam a distribuição dos módulos e os últimos 32 elementos representam a distribuição das fases. Os possíveis valores de $-\pi$ a π são mapeados entre as raias 32 e 64. Em relação ao módulo, o menor valor possível é zero, porém não existe um valor máximo já que o mesmo depende das constelações usadas e do ruído estocástico. Neste caso, foi escolhido um valor empírico estimado do conjunto de treino como valor máximo.

Assim como no caso CSS, é possível observar claramente nos histogramas, Figura 4.1, as diferentes assinaturas obtidas das modulações. Observa-se neste exemplo que a modulação BPSK apresenta, idealmente, um valor possível de módulo e dois possíveis valores de fase, enquanto que a modulação 4-PAM apresenta dois valores de módulo e dois valores de fase. A modulação 16-QAM apresenta três valores para módulo e doze fases distintas. Por último, a modulação 8-PSK apresenta apenas um valor possível para o módulo e oito valores possíveis para a fase. Estas são situações onde não há ruído.

Deve-se notar que o HISTO é relacionado a métodos que usam distribuições de probabilidade. A inovação nesse caso é utilizar tais parâmetros como entrada para um classificador. Comparando-se com métodos como o teste *Kolmogorov-Smirnov (KS)* [28,126] ou os baseados em verossimilhança, um classificador como MVS permite um aumento na robustez quando há descasamento entre as estimativas usadas para o projeto do sistema e as condições de operação.

Os resultados apresentados, nesta Tese, confirmam que o HISTO é uma proposta efetiva, pois combina as características positivas de extratores de parâmetros com baixo custo computacional e classificadores discriminativos robustos.

4.2 Probabilidade de classificação correta

A probabilidade de classificação correta (PCC) foi a medida adotada para avaliar o desempenho dos classificadores na CAM [8], a qual depende da probabilidade condicional de classificação $P_c^{(y'|y)}$ onde o y' -ésimo tipo de modulação foi a opção do classificador, quando a modulação correta é y . Desta forma, probabilidades $P_c^{(y'|y)}$ para os pares do produto cartesiano (y, y') entre as Y modulações podem ser estimadas a partir de \mathcal{T} exemplos de treino. Essas estimativas (ainda não normalizadas) podem ser organizadas em uma matriz $Y \times Y$ denominada de *matriz de confusão* [15], de maneira que ao ser normalizada, a diagonal principal corresponde à $P_c^{(y|y)}$ para o y -ésimo tipo de modulação. Com o objetivo de exemplificar esta definição, a Tabela 4.2 ilustra um exemplo de matriz de confusão.

	4-QAM	16-QAM	64-QAM
4-QAM	60	20	20
16-QAM	10	70	20
64-QAM	30	30	40

Tabela 4.1: Exemplo de matriz de confusão.

Lendo a Tabela 4.2 observa-se: na primeira linha tem-se 100 exemplos de modulação 4-QAM, de maneira que somente 60 exemplos são classificados de maneira correta; na segunda linha, novamente, tem-se 100 exemplos da modulação 16-QAM, os quais 70 exemplos são corretos; por último, tem-se 100 exemplos da modulação 64-QAM, onde somente 40 exemplos são corretos.

Desta maneira, em um processo de CAM onde temos um conjunto de teste com Y modulações totalizando \mathcal{T} exemplos, a média da PCC dentre todas as classes é estimada por:

$$\text{PCC} = \sum_{y=1}^Y P(y) P_c^{(y|y)}, \quad (4.1)$$

onde $P(y)$ é a probabilidade da classe y , que tipicamente é $P(y) = 1/Y, \forall y$.

4.3 Descrição dos experimentos

Nesta seção, uma variedade de experimentos são descritos a fim de analisar o desempenho do HISTO juntamente com CSS e HISTO2D. Os três de extratores de parâmetros utilizam o algoritmo MVS como classificador, as quais são denominados neste trabalho de: CSS-MVS, HISTO-MVS e HISTO2D-MVS. O desempenho de cada extrator foi comparado com os seguintes métodos de classificação: *average likelihood ratio test* (ALRT) [127], *discrete likelihood ratio test* (DLRT) [18], Cumulantes [27] e o teste *Kolmogorov-Smirnov* (KS) [28].

Os desempenhos foram avaliados a partir dos resultados obtidos oriundos de dois grupos de modulações. Estes grupos foram extraídos da literatura. O primeiro grupo [13,27] é baseado em 4 tipos de modulações listados abaixo e representa um conjunto de classes que não são linearmente separáveis:

$$\mathcal{M}_1 = \{\text{BPSK}, \text{4-PAM}, \text{16-QAM}, \text{8-PSK}\}. \quad (4.2)$$

O segundo grupo [28] considera somente a família de modulações QAM:

$$\mathcal{M}_2 = \{\text{4-QAM}, \text{16-QAM}, \text{64-QAM}\} \quad (4.3)$$

e representa um conjunto de modulações linearmente separáveis. Contando-se com um conjunto que é linearmente separável e outro que não é, pode-se melhor avaliar o desempenho de classificadores projetados para problemas linearmente separáveis, como a MVS linear, com os que não possuem esta restrição.

A classificação via Teste KS foi realizada somente para o conjunto de modulação \mathcal{M}_2 devido este classificador ser restrito a classificar modulações do tipo M -QAM e M -PSK.

4.4 Treinamento da MVS

Nesta seção, descreve-se as configurações a respeito das bases de treino e teste criadas para realizar as simulações, assim como também, a seleção dos modelos de classificador obtidos para cada tarefa de classificação.

Uma das primeiras etapas na tarefa de classificação, utilizando aprendizagem supervisionada, é determinar o tamanho da base de treino (número de vetores de treino) de modo que o classificador atinja um valor de PCC satisfatório. Tal objetivo é alcançado através do gráfico de complexidade amostral, onde observa-se a variação do valor de PCC à medida que o valor do número de amostras de treino aumenta para um determinado valor de RSR.

A Figura 4.2 ilustra a complexidade amostral para uma MVS utilizando HISTO para quatro valores distintos de RSR onde fixou-se o número de vetores de teste e o número de vetores de treino foi variado. As curvas foram levantadas a partir de uma base de teste, contendo 1000 vetores, onde cada vetor foi composto por $N = 250$ amostras para cada modulação do conjunto \mathcal{M}_1 . Cada gráfico de complexidade amostral é caracterizado por uma abscissa indicando a quantidade de vetores de treino e a ordenada indicando o valor da PCC.

Analisando a Figura 4.2, verifica-se que o valor de PCC aumenta a partir do crescimento do valor da RSR. Vale notar que para valores baixos de RSR, nota-se uma variação do valor de PCC, porém tal fato não é observado para valores altos de RSR.

Desta forma, para as simulações deste trabalho, um conjunto de treino e outro de teste foram criados de maneira independente e disjunta para cada avaliação, os quais são gerados de acordo com a Equação 2.5. Para cada valor de RSR, 500 exemplos de cada modulação foram gerados para os conjuntos de treino e teste, onde cada exemplo é constituído por $N = 250$ amostras. Neste trabalho, dois tipos de kernel foram utilizados: linear e o RBF (*radial basis function*). A motivação é que o RBF é conhecido por conduzir a um bom desempenho, enquanto o linear permite a redução no custo computacional.

Os extratores de parâmetros HISTO e HISTO2D tiveram $Q = 32$ bins nas suas con-

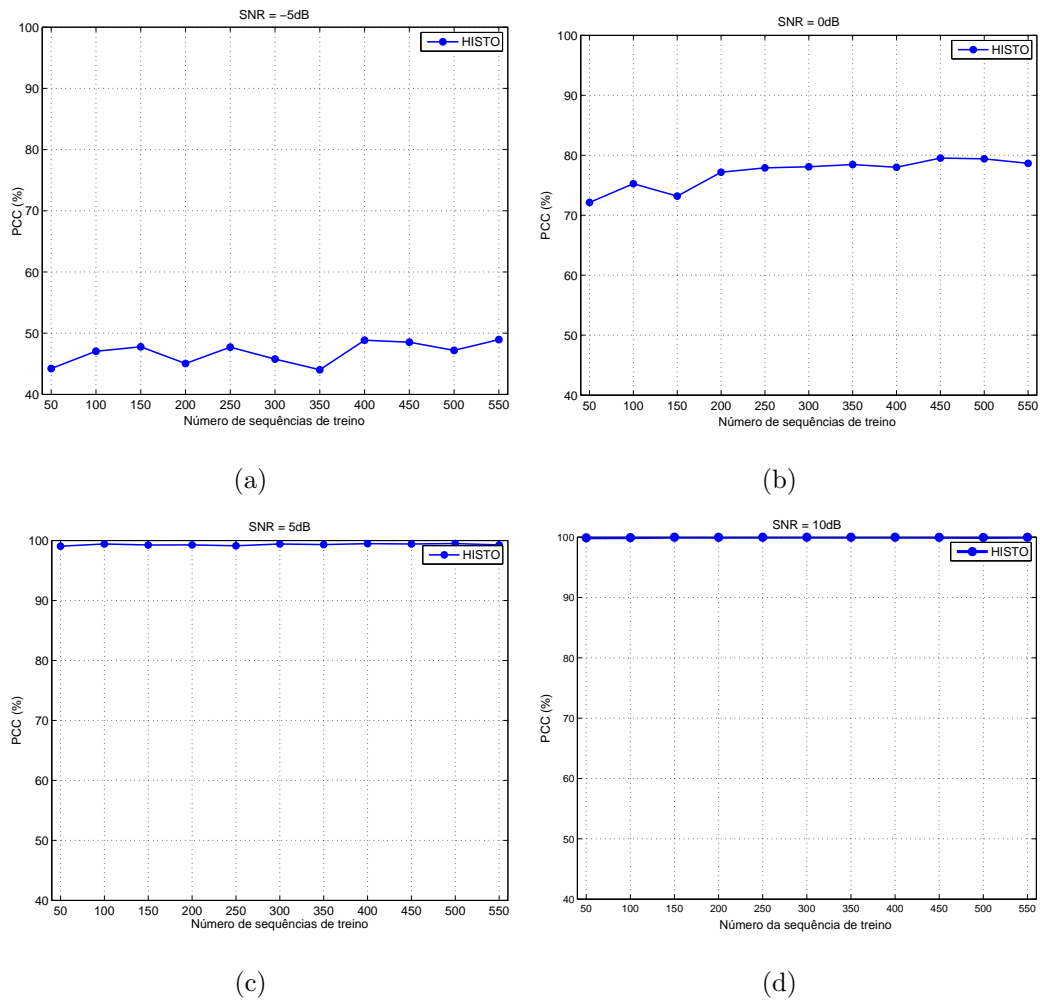


Figura 4.2: Análise de complexidade amostral.

figurações, de modo que resultaram em $D = 64$ e $D = 1024$ características de entrada para o classificador MVS. Já o extrator de parâmetros CSS fornece ao classificador $D = 500$ características. Outros valores foram avaliados, mas os descritos levaram a bons resultados.

Como citado, a tarefa de classificação do algoritmo MVS é baseada na aprendizagem supervisionada, a qual busca aprender com passado (treino) para poder generalizar sobre o futuro (teste). Um dos pontos fundamentais para o processo de aprendizagem é a seleção de parâmetros internos do classificador, cujo objetivo é a otimização dos resultados de classificação. Nesse contexto, a técnica denominada de *grid search* [128] foi utilizada neste trabalho.

A técnica *grid search* está relacionada à busca dos melhores parâmetros por meio da análise dos resultados obtidos através da execução do algoritmo para um intervalo de parâmetros. No caso da função do kernel linear, esse grid é unidimensional com sequências exponenciais crescentes $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$. Em relação ao kernel RBF, a busca pelos me-

lhores parâmetros é caracterizada por um grid bi-dimendisional constituído pela faixa valores de C e pela faixa $\lambda = 2^{-5}, 2^{-3}, \dots, 2^{15}$. Por fim, a técnica de seleção de parâmetros *grid search* foi utilizada com um conjunto de validação disjunto em relação a ambos conjuntos de treino e teste.

4.5 Cenários

4.5.1 Classificação de modulação em canais AWGN

No primeiro cenário, vamos analisar o desempenho das três propostas sob condições próximas das ideais, de maneira que na Equação 2.5 os valores dos parâmetros $f_0 = 0$, $\phi_n = 0$ e o ganho $a_c = 1$ são conhecidos e constantes. O ruído Gaussiano complexo aditivo v_n possui média zero e componentes reais e imaginárias com variância $\sigma_g^2/2$. A razão sinal-ruído (RSR) é definida como $\frac{1}{\sigma_g^2}$, considerado que as constelações tenham variância unitária.

Este primeiro cenário é considerado menos desafiador quando comparado aos demais. Com o objetivo de avaliar o desempenho dos classificadores em relação ao conhecimento ou não do valor da RSR, duas abordagens foram adotadas:

- Os resultados da primeira abordagem assumem perfeito conhecimento da RSR, pois tanto a etapa de treino quanto a etapa de teste usam valores “casados” da RSR, ou seja, os classificadores foram treinados e testados para cada valor específico de RSR;
- Para efeito de comparação dos resultados em relação à primeira abordagem, formulou-se uma segunda abordagem, a qual foi chamada de multi-condição (MC). Na etapa de treino, houve um único treinamento numa faixa de valores específicos de RSR em dB (somente para as MVSs), ao contrário do primeira abordagem que houve vários treinamentos para cada valor RSR específico. Na etapa de teste, o classificador resultante é utilizado por todos os vetores de teste, independente da sua RSR.

As descrições originais do classificador DLRT [18], Cumulantes [27] e do classificador via Teste KS [28] não incluem uma avaliação quando não há estimativa do valor da RSR sob um canal AWGN. Assim, para evitar avaliação injusta, esses três classificadores e o classificador ALRT não foram avaliados quando o parâmetro σ_g^2 não é perfeitamente conhecido. Além disso, o escopo dos resultados obtidos a partir dos cumulantes restringe-se ao conjunto de modulações \mathcal{M}_1 , considerando somente o ambiente AWGN devido apresentar resultados mais significativos.

Os primeiros resultados obtidos a partir de um canal AWGN para conjunto de modulação \mathcal{M}_1 é ilustrado na Figura 4.3. Analisando a performance dos classificadores, verifica-se que o classificador ALRT apresentou os melhores resultados, como já era esperado visto que o mesmo obtém o desempenho ótimo nesta condição. Porém, ressalta-se que esta performance não é muito superior. Por exemplo, o classificador ALRT obteve PCC=85.9% quando a RSR=0 dB. Tal resultado de PCC é alcançado pelos classificadores CSS, HISTO e HISTO2D (PCC=87.2%) quando a RSR=1 dB e pela classificação baseada em cumulantes (PCC=90.6%) quando a RSR=4 dB, ou seja, nessa faixa de valores de RSR no cenário considerado ,há perda de performance de apenas 1 dB em relações aos classificadores baseados em MVS e 4 dB no caso dos cumulantes conforme ilustra a Figura 4.3.

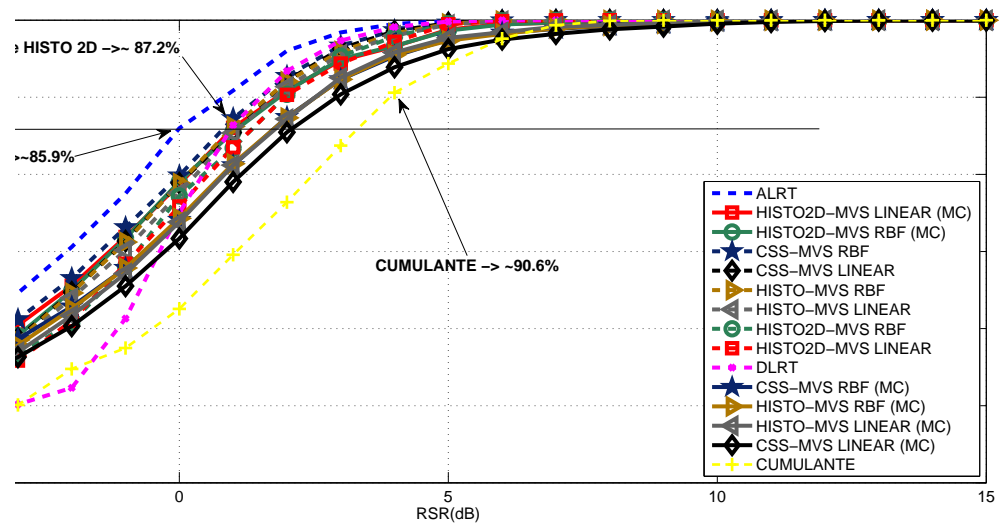


Figura 4.3: Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_1 . Cenário AWGN com faixa de RSR $[-5, 15]$ dB.

Além disso, verifica-se na Figura 4.3 que os classificadores baseados no desconhecimento do valor RSR (abordagem MC) obtiveram resultados melhores em relação a classificação baseada em cumulantes. Por último, nota-se também que os classificadores CSS-MVS, HISTO-MVS e HISTO2D-MVS baseados no kernel linear apresentaram bons resultados quando comparados com os resultados dos classificadores baseados no kernel RBF.

Já a Figura 4.4 ilustra os resultados relacionados ao conjunto de modulação \mathcal{M}_2 . Analisando a performance dos classificadores, tendo como base a performance obtida do classificador ALRT que obteve PCC= 100% quando a RSR=10 dB , pode-se verificar que os classificadores que possuem conhecimento do valor da RSR alcançaram melhor desempenho quando comparados com os classificadores em MC. Além disso, os classificadores baseados em MVS, quando a RSR é conhecida, obtiveram aproximadamente PCC=100% quando a RSR=12 dB enquanto

a classificação baseada no Teste KS atingiu o mesmo valor de PCC quando a RSR=15 dB, ou seja, a performance dos classificadores baseados em MVS obteve 3 dB de melhoria comparada a performance de classificação via Teste KS.

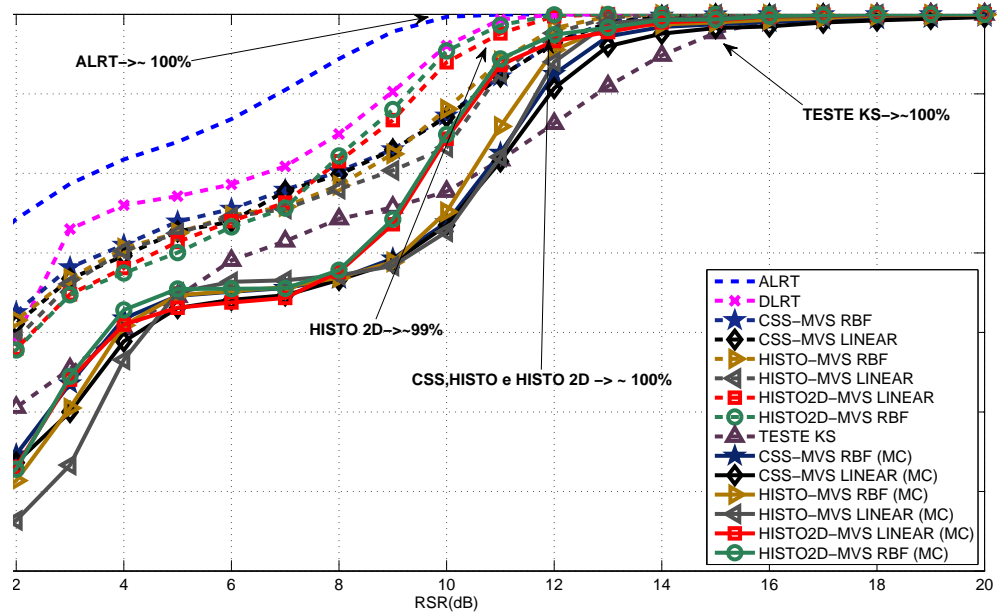


Figura 4.4: Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_2 . Cenário AWGN com faixa de RSR [0, 20] dB.

Um resultado interessante da Figura 4.4 é que, mesmo na abordagem MC, os classificadores baseados em MVS, a partir de 11 dB, superaram o desempenho do classificador Teste KS cujos resultados são baseados no conhecimento do valor RSR.

4.5.2 Classificação de modulação em canais com desvanecimento plano (*flat-fading*) e com *offset* de frequência

Neste segundo cenário, a classificação de modulação foi expandida para canais com desvanecimento plano e com *offset* de frequência f_0 , de modo que, a classificação é baseada em diferentes realizações do canal, cujo ganho é $a_c \sim \mathcal{N}_c(0, \sigma_{a_c}^2)$. Por sua vez, a RSR é definida como $\frac{\sigma_{a_c}^2}{\sigma_g^2}$ neste caso.

Nesta parte do trabalho, o parâmetro f_0 é uma variável aleatória constante para cada realização $r = \{r_n\}_{n=1}^N$, onde N é o número de símbolos de cada realização. Ainda levando-se em conta o modelo da Equação 2.5, o parâmetro $\phi_n = 0$ informa não haver jitter de fase.

A faixa de valores assumida para o *offset* de frequência normalizado $f_0 T_s$ foi definida, tanto para a base de dados de treino quanto para de teste, como obedecendo a uma distribuição uniforme no intervalo $[0; 0,001]$ [27]. Deve-se interpretar essa faixa como f_0 sendo no máximo 0,1% da taxa de símbolos R_{sim} . Este intervalo foi adotado por ter sido usado em [27], mas há situações onde considera-se que f_0 possui valores maiores [29].

As duas abordagens de classificação realizadas na seção 4.5.1 também foram feitas neste cenário. Deve-se atentar que durante a etapa de teste, o parâmetro f_0 não é conhecido em nenhuma abordagem. Por sua vez, o ganho a_c é desconhecido somente na abordagem multi-condição, ou seja, neste cenário, o classificador desconhece o valor da RSR e os parâmetros f_0 e a_c . Para esta abordagem, o desempenho de classificação dos classificadores ALRT e DLRT e Cumulantes não foram reportados devido aos mesmos apresentarem resultados não satisfatórios. Isto é esperado pois os mesmos requerem que as estimativas nas quais são baseados, sejam próximas das estatísticas dos dados usados no teste.

A Figura 4.5 ilustra os resultados obtidos a partir do grupo \mathcal{M}_1 , onde se verifica que o desempenho dos classificadores melhora com o aumento do valor da RSR, como esperado.

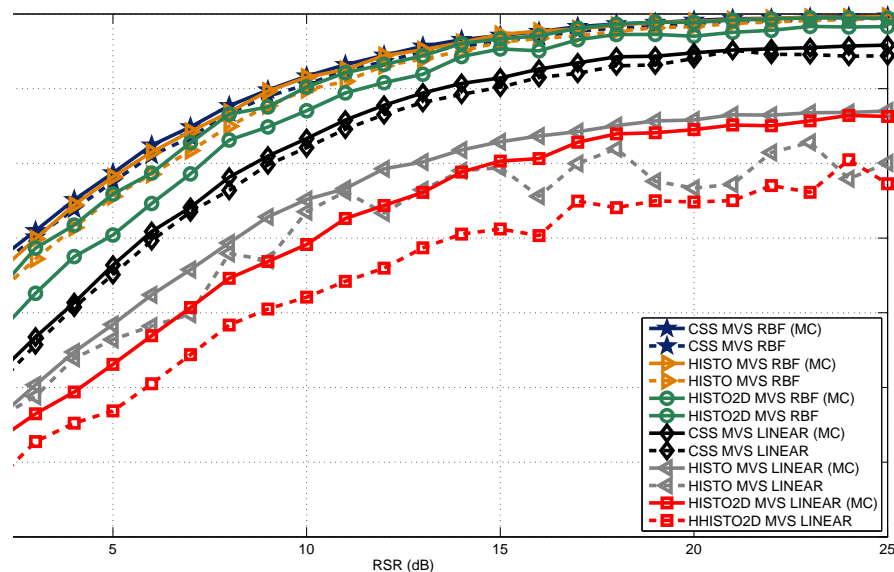


Figura 4.5: Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_1 obtidos a partir do canal com desvanecimento plano e *offset* de frequência com faixa de RSR $[0, 25]$ dB.

Os classificadores baseados no kernel RBF apresentaram um melhor desempenho quando comparados aos classificadores baseados no kernel linear. Uma razão para este fato é que a magnitude das modulações BPSK e 8-PSK apresentarem o mesmo valor, tornando

essas classes não linearmente separáveis. Deste modo, a maior parte dos erros de classificação dos classificadores baseados no kernel linear ocorreram entre as modulações BPSK e 8-PSK, mesmo para valores altos de RSR.

Por último, nota-se que os resultados obtidos a partir da abordagem MC são semelhantes aos resultados obtidos quando a RSR é conhecida praticamente para todos os classificadores, menos para os classificadores HISTO-MVS e HISTO2D-MVS com kernel linear.

Em relação à Figura 4.6 verifica-se que, diferentemente da Figura 4.5, o desempenho dos classificadores CSS-MVS, HISTO-MVS e HISTO2D-MVS para ambos os kernels são similares tanto na abordagem MC quanto na RSR conhecida.

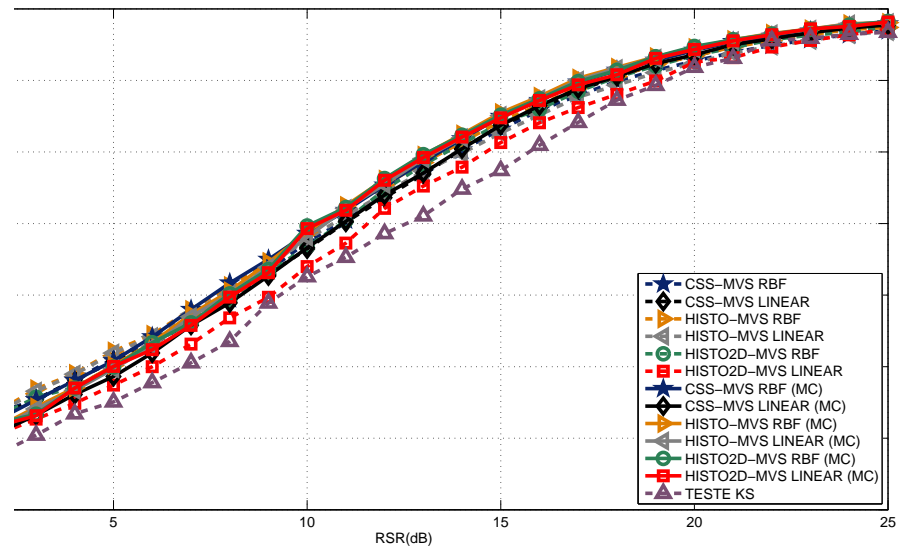


Figura 4.6: Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_2 obtidos a partir do canal com desvanecimento plano e *offset* de frequência com faixa de RSR $[0, 25]$ dB.

Além disso, pode-se verificar que apesar do classificador via Teste KS apresentar bons resultados, seu desempenho é inferior aos resultados obtidos pelos classificadores baseados em MVS. Levando-se em conta que o ALRT e DLRT também não funcionam bem nesse caso, fica reforçado o argumento sugerido por esta Tese de que extratores de parâmetros com baixo custo computacional podem ser combinados com eficácia a classificadores discriminativos como MVS.

4.5.3 Classificação de modulação em canais com desvanecimento plano (*flat-fading*) com *jitter* de fase

O terceiro cenário é caracterizado pelo desconhecimento do parâmetro ϕ_n em canais com desvanecimento plano. Diferentemente do cenário da Seção 4.5.2, o parâmetro ϕ_n é caracterizado por ser uma variável aleatória cujo valor varia de símbolo para símbolo.

A faixa de valores assumidos para *jitter* de fase ϕ_n foi definida também, tanto na base de dados de treino quanto de teste, como valores apresentando uma distribuição uniforme no intervalo $[-90^\circ, 90^\circ]$, como em [27].

Neste cenário, assim como foi feito na Seção 4.5.1, também houve duas abordagens distintas de treinar os classificadores e avaliar a classificação. Todavia, deve-se atentar que o ϕ_n não é conhecido em qualquer situação, e assim como no cenário anterior, o ganho a_c é desconhecido somente na abordagem multi-condição, ou seja, nesta situação, o classificador desconhece o valor da RSR e os parâmetros ϕ_n e a_c .

A Figura 4.7, ilustra os resultados em relação ao conjunto de modulação \mathcal{M}_1 quanto ao desconhecimento do parâmetro ϕ_n em canais com desvanecimento plano.

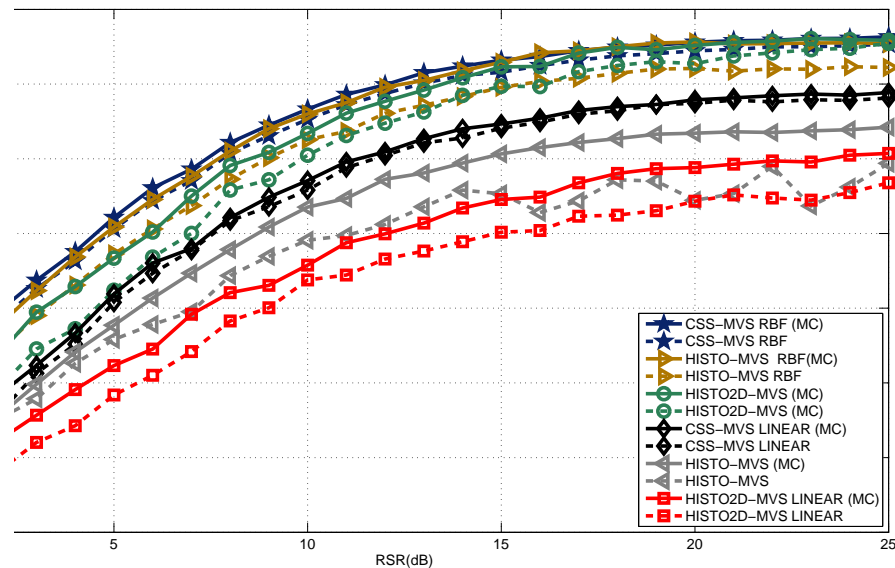


Figura 4.7: Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_1 obtidos a partir do canal com desvanecimento plano e *phase jitter* com faixa de RSR [0, 25] dB.

Analisando a Figura 4.7 nota-se que os classificadores baseados no kernel RBF apresentam performance superior em relação aos classificadores baseados no kernel linear dado que o

conjunto de modulações \mathcal{M}_1 possui classes que não são linearmente separáveis.

Os classificadores CSS-MVS RBF e HISTO-MVS RBF, tanto na abordagem MC quanto na abordagem quando a RSR é conhecida, obtiveram uma probabilidade de classificação correta de 95.3% em 25 dB. O classificador HISTO2D-MVS Linear, quando a RSR é conhecida, obteve o pior desempenho neste cenário.

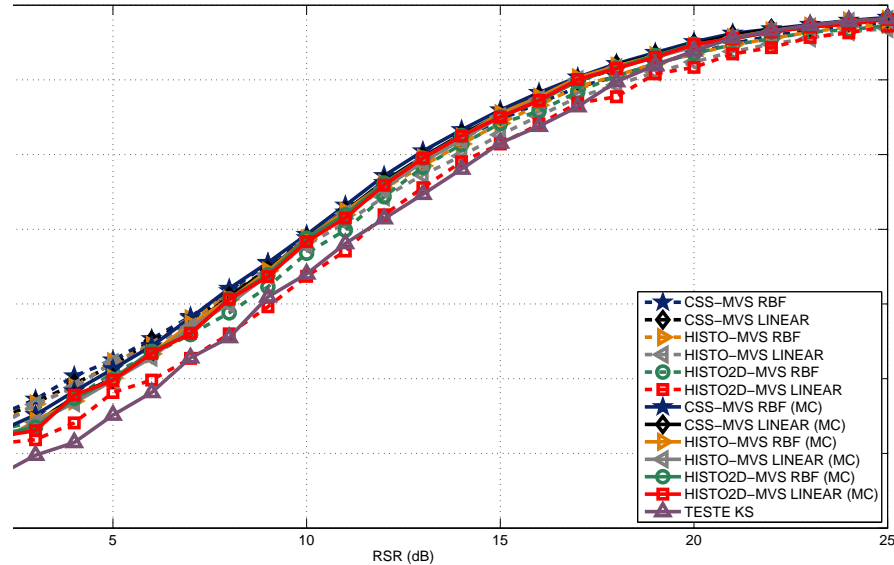


Figura 4.8: Probabilidade de classificação correta para o conjunto de modulação \mathcal{M}_2 obtidos a partir do canal com desvanecimento plano e *phase jitter* com faixa de RSR [0, 25] dB.

A Figura 4.8 ilustra os resultados relacionados ao conjunto de modulação \mathcal{M}_2 . Verifica-se que o classificador via Teste KS é superado pelos outros classificadores, os quais apresentam resultados similares tanto na abordagem MC quanto na de RSR conhecida.

Como conclusão preliminar, registra-se a robustez da combinação dos extratores propostos com MVS. O CSS é superior em termos de acurácia, mas a próxima seção irá discutir que o HISTO possui custo computacional bem menor.

4.5.4 Avaliação do custo computacional

Sabe-se que estimar um custo computacional não é uma tarefa trivial, porém pode-se obter uma estimativa aceitável através da atribuição de pesos para cada operação matemática. Desta forma, nesta parte do trabalho, através das Equações (3.44) e (3.45), é avaliado o custo computacional da etapa de teste dos classificadores CSS-MVS, HISTO-MVS e HISTO2D-MVS

Tabela 4.2: Exemplo de ICC for HISTO-MVS e CSS-MVS.

Classificador MVS	N_{sv}	D	ICC
HISTO-MVS Linear	33.896	128	256
HISTO-MVS RBF	32.183	128	$\sim 1,3 \cdot 10^7$
CSS-MVS Linear	17.700	500	1.000
CSS-MVS RBF	17.473	500	$\sim 2,6 \cdot 10^7$
HISTO2D-MVS LINEAR	33.700	1.024	2.048
HISTO2D-MVS RBF	33.558	1.024	$\sim 10 \cdot 10^7$

na abordagem multi-condição, especificamente para o conjunto de modulação \mathcal{M}_2 .

O conjunto de modulação \mathcal{M}_2 é constituído por três modulações, de modo que, três classificadores binários são treinados para realizar a classificação multi-classe. Desta forma, a Tabela 4.2 apresenta o custo computacional via ICC de somente um único classificador binário $f(\mathbf{z})$. Esses números devem ser multiplicados por três para obtermos a quantidade de operações finais realizadas de um classificador multi-classe $F(\mathbf{z})$ a ser usado para \mathcal{M}_2 . Raciocínio semelhante deve ser feito para outras classes.

Desta forma, analisando-se na Tabela 4.2 a quantidade de vetores de suporte N_{sv} juntamente com o número de parâmetros de entrada D de cada classificador, verifica-se que a classificação baseada no HISTO-MVS apresenta, tanto no kernel linear quanto RBF, um custo de operações menor comparado aos outros dois classificadores. O custo computacional do kernel RBF é significativamente maior em relação ao kernel linear para as três propostas baseadas em MVS, chegando a ser cinco ordens de grandeza maior.

Esta observação ainda não havia sido feita na literatura. Assim, uma contribuição deste trabalho é a proposta de uso de perceptrons obtidos a partir de MVS lineares no problema de CAM. A Tabela 4.2 mostra que mesmo com N_{sv} sendo grande para MVS lineares, o custo total é baixo pois os mesmos não são armazenados e sim convertidos no perceptron.

O custo menor do classificador HISTO-MVS é comprovado ao analisarmos o quantitativo dos recursos utilizados por um FPGA, após compilação no programa Altera Quartus II, para implementação dos classificadores CSS-MVS Linear e HISTO-MVS Linear [129]. Analisando a Tabela 4.3, pode-se observar que o classificador HISTO-MVS Linear apresentou vantagens em comparação ao classificador CSS-MVS Linear nos requisitos: quantidade de memória RAM e número de multiplicadores embarcados.

A menor quantidade de memória RAM do HISTO-MVS linear é justificada pelo fato

já citado do classificador não precisar armazenar os vetores de suporte. E por fim, a não utilização de multiplicadores embarcados é justificada por não ter sido necessária a operação de multiplicação para o funcionamento do classificador HISTO-MVS linear [11].

Tabela 4.3: Utilização de recursos do FPGA ALTERA na implementação do CSS-MVS e HISTO-MVS utilizando kernel linear.

Recurso	CSS-MVS	HISTO-MVS
Elementos lógicos (Total)	1.824 (10%).	911 (5%)
Funções combinacionais	1.586 (8%)	985 (5%)
Registradores lógicos dedicados	704 (4%)	462 (2%)
Total de bits de memória	64.000 (27%)	6.1 (3%)
Multiplicadores embarcados	24 (46%)	0 (0%)
Tempo para 200 classificações	138 ms	4 ms

4.6 Considerações Finais

Neste capítulo foram apresentados resultados representativos relacionados aos classificadores propostos: CSS-MVS, HISTO-MVS e HISTO2D-MVS. Esses resultados foram confrontados com outros classificadores do estado da arte, baseando-se a comparação em dois conjuntos de modulações e duas abordagens: a primeira foi baseada na classificação onde os classificadores tinham conhecimento do valor da RSR do sinal recebido; a segunda foi baseada no desconhecimento do valor da RSR e outras variáveis do sinal recebido, a qual foi denominada de multi-condição.

Tais abordagens foram realizadas nos dois grupos de modulação: um grupo com exemplos de modulações linearmente separáveis e outro grupo com modulações não linearmente separáveis. Os desempenhos dos classificadores foram avaliados por meio desses grupos através de um canal AWGN e um canal com desvanecimento plano. Os resultados com MVS foram obtidos com os kernels linear e RBF.

Através dos resultados obtidos, pôde-se verificar que as propostas apresentaram um bom desempenho nos cenários testados. É interessante notar que na abordagem MC, os classificadores obtiveram bons resultados e até melhores em relação aos classificadores que tinham conhecimento da RSR.

No grupo de modulação onde os exemplos são linearmente separáveis, pôde-se verificar

que os resultados de classificação baseados no kernel linear são muito próximos aos resultados obtidos via kernel RBF. Isso significa, que caso as modulações sejam linearmente separáveis, o uso do kernel linear é eficiente na tarefa de classificação pois diminui bastante o custo computacional. Isso se dá pois a classificação neste caso não depende do número de vetores de suporte gerados durante a etapa de treino. Em contrapartida, os resultados oriundos do conjunto de modulação não linearmente separável via kernel RBF obtiveram resultados melhores em relação aos resultados obtidos via kernel linear, porém o custo computacional do processo de classificação é muito maior devido à dependência deste kernel em relação ao número de vetores de suporte.

Em termos de desempenho dos extratores de parâmetros, o CSS apresentou melhor acurácia, seguido pelo HISTO. Em alguns casos esta diferença foi desprezível.

Capítulo 5

Conclusões

É evidente que umas das principais motivações na pesquisa atual sobre CAM veio do advento do RC, que tem como um dos principais objetivos o compartilhamento do espectro de frequência de maneira justa e eficaz. A fim de tornar essa tarefa possível, a classificação automática de modulação de sinais é essencial na descoberta do tipo de modulação empregado pelo sinal recebido. Através dessa informação, o RC pode configurar os parâmetros de recepção de modo a estabelecer um processo de comunicação eficiente com transmissor e gerenciar o compartilhamento do espectro com outros usuários. Outra aplicação importante de CAM é na área militar, que se destaca por tipicamente não haver informação prévia sobre o sinal a ser classificado.

Em qualquer caso, realizar a tarefa de classificação de modulação em um ambiente real não é trivial, principalmente quando o sistema trabalha em um cenário onde muitos parâmetros do sinal recebido são desconhecidos. Tal cenário é abordado em poucos trabalhos na literatura recente, pois a maioria descreve propostas de classificação no âmbito de RC, onde os classificadores possuem conhecimento de todos os parâmetros ou quase todos. Em contraste, este trabalho teve como foco ambientes não cooperativos.

Baseado numa pesquisa abrangente, esse trabalho descreveu as duas principais abordagens para CAM: a primeira é caracterizada pela verossimilhança e a segunda baseada no reconhecimento de padrão; ambas com vantagens e desvantagens. Embora a abordagem baseada na verossimilhança proporcione um desempenho que pode ser o ótimo, é difícil obter uma solução analítica exata quando o número de parâmetros desconhecidos aumenta. Além disso, quando uma solução analítica fechada existe, seu uso às vezes torna-se impraticável em cenários reais devido à complexidade computacional. Algoritmos como ALRT e DLRT foram exemplos dessa abordagem neste trabalho, os quais serviram de comparativo para as propostas apresentadas.

Na abordagem baseada no reconhecimento de padrão, algumas características do sinal

são utilizadas a fim de identificar o formato da modulação. Embora essa abordagem seja sub-ótima, algoritmos baseados no reconhecimento de padrão são mais fáceis de implementar e relativamente robustos. Muitos desses algoritmos utilizam amplitude, frequência e fase instantânea, transformada wavelet do sinal e estatística do sinal como momentos, cumulantes e cicloestacionariedade para distinguir os diversos formatos de modulação.

Esta Tese sugere o uso da sinergia que há entre extratores de parâmetro eficazes e classificadores discriminativos. Mais especificamente, avaliou-se a classificação de modulação baseada na aprendizagem discriminativa via algoritmo MVS utilizando três formas diferentes de extratores de parâmetros: HISTO, CSS e HISTO2D. O HISTO é uma proposta original e exclusiva do autor da Tese, enquanto os dois outros foram desenvolvidos no âmbito de cooperações científicas. Tais propostas apresentaram bons resultados em cenários simples como AWGN e também em cenários mais próximos à realidade, caracterizados por um canal com desvanecimento plano juntamente com variações decorrentes de *phase jitter* e *offset* de frequência.

Verificou-se que os resultados obtidos utilizando kernel linear são tão bons quanto os resultados através do kernel RBF quando o conjunto de modulação é caracterizado por modulações linearmente separáveis, diminuindo consideravelmente o custo computacional do processo de classificação. Este tipo de sistema ainda não havia sido destacado pela literatura. Uma grande vantagem do mesmo é que se torna aplicável em cenários de tempo real e baixo consumo de potência devido ao seu baixo custo computacional, de várias ordens de grandeza quando comparado ao uso tradicional de MVSs. Entretanto, quando o conjunto de modulação não é linearmente separável, percebeu-se que os resultados utilizando kernel RBF apresentam resultados melhores em termos de acurácia.

Além disso, este trabalho demonstrou que os algoritmos de classificação propostos são capazes de classificar modulações mesmo não sabendo o valor de RSR. Em contraste, o conhecimento da RSR é pré-requisito para o bom funcionamento de classificadores como a DLRT.

Ademais, foi realizada a avaliação do custo computacional dos extratores de parâmetros propostos. Nesta avaliação foi levado em conta o número de parâmetros D passados ao classificador MVS e o número de vetores de suporte quando o kernel RBF é utilizado. Pelo fato do kernel linear não depender do número de vetores de suporte, a estimativa do seu custo computacional atingiu valor muito menor quando comparado ao kernel RBF.

Em resumo, dentre as várias técnicas de classificação, o HISTO-MVS mostrou-se competitivo devido ao relativo pequeno custo computacional, em especial se as classes forem linearmente separáveis, e boa acurácia.

5.1 Trabalhos Futuros

Alguns possíveis trabalhos futuros são:

- Testes práticos utilizando os classificadores CSS-MVS, HISTO-MVS e HISTO2D-MVS com sinais de comunicação reais, obtidos com o uso de equipamentos como a *Universal Software Radio Peripheral* (USRP);
- avaliar esquemas hierárquicos de CAM, expandindo-se ao máximo o número de modulações suportadas pelo sistema;
- testar o uso de diferentes classificadores aos extratores propostos, utilizando aprendizagem discriminativa.

Referências Bibliográficas

- [1] Z. Zhu and A. Nandi, *Automatic modulation classification : principles, algorithms, and applications*. Chichester, West Sussex, United Kingdom: Wiley, 2015.
- [2] C. Weber, M. Peter, and T. Felhauer, “Automatic modulation classification technique for radio monitoring,” *Electronics Letters*, vol. 51, no. 10, pp. 794 – 796, 2015.
- [3] S. Haykin, “Cognitive radio: Brain-empowered wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 144 – 150, 2005.
- [4] F. C. Commission, “Spectrum policy task force,” *Rep. ET Docket*, vol. 02 n. 135, Nov. 2002.
- [5] J. Mitola, “The software radio architecture,” *IEEE Commun. Mag.*, vol. 33, no. 5, pp. 26 – 38, 1995.
- [6] IEEE, “Ieee 802.22 wireless regional area networks.” <http://www.ieee802.org/22/>.
- [7] F. Ian, L. Won-Yeol, C. Mehmet, and M. Shantidev, “Next generation dynamic spectrum access cognitive radio wireless networks: A survey,” *Computer Networks*, vol. 50, no. 13, pp. 2127 – 2159, 2006.
- [8] O. Dobre, A. Abdi, Y. Bar-Ness, and W.Su, “Survey of automatic modulation classification techniques: Classical approaches and new trends,” *IET Commun.*, vol. 1, no. 2, pp. 137 – 156, Apr. 2007.
- [9] B. Sklar, *Digital communications: Fundamentals and applications*, 2nd ed. Prentice Hall, 2001.
- [10] J. Proakis and M. Salehi, *Digital Communications*, 5th ed. McGraw-Hill, 2007.
- [11] C. Cardoso, A. Castro, and A. Klautau, “An efficient FPGA IP core for automatic modulation classification,” *IEEE Embedded Syst. Lett.*, vol. 5, no. 3, pp. 42 – 45, 2013.

-
- [12] R. Alves, E. Sorato, E. Fronza, C. Cardoso, R. Netto, P. Michel, A. Castro, A. Klautau, and J. Güntzel, “Robust and low cost modulation classification using svm and two-dimensional histogram (submetido),” Sept. 2015.
- [13] F. Muller, C. Cardoso, and A. Klautau, “A front end for discriminative learning in automatic modulation classification,” *IEEE Commun. Lett.*, vol. 15, no. 4, pp. 443 – 445, Apr. 2011.
- [14] H. V. Trees, *Detection, estimation, and modulation theory*. Wiley, 2003.
- [15] R. Duda, P. Hart, and D. Stork, *Pattern classification*. Wiley, 2001.
- [16] B. D. O. Ozdemir, P. Varshney, and W. Su, “A novel approach to dictionary construction for automatic modulation classification,” *Journal of the Franklin Institute*, vol. 351, no. 5, pp. 2991 – 3012, 2014.
- [17] Y. Yang, J. Chang, J. Liu, and C. Liu, “Maximum log-likelihood function-based QAM signal classification over fading channels,” *Wireless Pers. Commun.*, vol. 28, no. 1, pp. 77 – 94, 2004.
- [18] W. Su, J. Xu, and M. Zhou, “Real-time modulation classification based on maximum likelihood,” *IEEE Commun. Lett.*, vol. 12, pp. 801–803, Nov. 2008.
- [19] E. Avci and D. Avci, “Using combination of support vector machines for automatic analog modulation recognition,” *Expert Systems with Applications*, vol. 36, no. 2, Part 2, pp. 3956–3964, Mar. 2009.
- [20] A. Sahai, N. Hoven, and R. Tandra, “Some fundamental limits on cognitive radio,” in *The Allerton Conference on Communication, Control and Computing*, 2004.
- [21] J. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2001.
- [22] S. Haykin, *Introduction to Analog and Digital Communications*. Wiley Sons, 2007.
- [23] M. Azarbad, A. Ebrahimzadeh, and J. Addeh, “A new intelligent approach for recognition of digital satellite signals,” *Journal of Signal Processing Systems*, vol. 79, no. 1, pp. 75 – 88, 2015.
- [24] M. Bari and M. Doroslovacki, “Robust recognition of linear and nonlinear digital modulations of RRC pulse trains,” in *Conference on Information Sciences and Systems (CISS)*, Mar. 2015, pp. 1 – 6.

-
- [25] F. Wang, R. Xu, and Z. Zhong, “Low complexity Kolmogorov-Smirnov modulation classification,” in *Proc. Wireless Communications and Networking Conference*, Mar. 2011, pp. 1607 – 1611.
- [26] A. Sherme, “A novel method for automatic modulation recognition,” *Appl. Soft Comput.*, vol. 12, no. 1, pp. 453 – 461, Jan. 2012.
- [27] A. Swami and B. Sadler, “Hierarchical digital modulation classification using cumulants,” *IEEE Trans. on Commun.*, vol. 48, no. 3, pp. 416 – 429, Mar. 2000.
- [28] F. Wang and X. Wang, “Fast and robust modulation classification via Kolmogorov-Smirnov test,” *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2324 – 2332, Aug. 2010.
- [29] F. Gini and B. Georgios, “Frequency offset and symbol timing recovery in flat-fading channels: A cyclostationary approach,” *IEEE Transactions on Communications*, vol. 46, pp. 400–411, March 1998.
- [30] Y. Liu, A. Haimovich, W. Su, and E. Kanterakis, “Theoretical bound on modulation classification for multiple-input multiple-output (MIMO) systems over unknown, flat fading channels,” in *Conference on Information Sciences and Systems (CISS)*, Mar. 2015, pp. 1–5.
- [31] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and H. Witten, “The WEKA data mining software: an update,” *SIGKDD Explor. Newsl.*, vol. 11, pp. 10 –18, 2009.
- [32] A. Hazza, M. Shoaib, S. Alshebeili, and A. Fahad, “An overview of feature-based methods for digital modulation classification,” in *Communications, Signal Processing, and their Applications*, Feb. 2013, pp. 1 – 6.
- [33] E. Azzouz and A. Nandi, “Automatic identification of digital modulation types,” *Signal Processing*, vol. 47, no. 1, pp. 55 – 69, 1995.
- [34] E. Azzouz and A. Nandi, “Procedure for automatic recognition of analogue and digital modulations,” *IEE Proceedings - Communications*, vol. 143, no. 5, pp. 259–266, Oct. 1996.
- [35] A. Nandi and E. Azzouz, “Automatic analogue modulation recognition,” *Signal Processing*, vol. 46, no. 2, pp. 211 – 222, 1995.
- [36] K. Maliatsos, S. Vassaki, and P. Constantinou, “Interclass and intraclass modulation recognition using the wavelet transformation,” in *Proc. PIMRC*, Sept. 2007, pp. 1 – 5.

-
- [37] L. Youyong, L. Guolong, X. Xiaoka, and L. Xinxin, "The methods of recognition for common used M-ary digital modulations," in *Proc. WiCom*, Oct. 2008, pp. 1 – 4.
- [38] J. Yuan, Z. Zhao-Yang, and Q. Pei-Liang, "Modulation classification of communication signals," in *Proc. MILCOM*, Oct. 2004, pp. 1470 – 1476.
- [39] H. Fuqing, Z. Zhi, X. Yi-tao, and R. Guo, "Modulation recognition of symbol shaped digital signals," in *Proc. ICCAS*, May 2008, pp. 328 – 332.
- [40] L. D. Vito, S. Rapuano, and M. Villanacci, "Prototype of an automatic digital modulation classifier embedded in a real-time spectrum analyzer," *IEEE Transactions on Instrum. and Meas.*, vol. 59, no. 10, pp. 2639 – 2651, Oct 2010.
- [41] A. Nandi and E. Azzouz, "Algorithms for automatic recognition of communication signals," *IEEE Trans. Commun.*, vol. 46, no. 4, pp. 431 – 436, Apr. 1998.
- [42] U. Satija, M. Mohanty, and B. Ramkumar, "Automatic modulation classification using S-transform based features," in *Signal Processing and Integrated Networks (SPIN)*, Feb. 2015, pp. 708 – 712.
- [43] D. Wu, X. Gu, and Q. Guo, "A new scheme of automatic modulation classification using wavelet and WSVM," in *Proc. International Conference on Mobile Technology, Applications and Systems*, Nov. 2005, pp. 1 – 5.
- [44] P. Cheol-Sun, C. Jun-Ho, N. Sun-Phil, and J. Won, "Automatic modulation recognition of digital signals using wavelet features and SVM," in *Proc. 10th ICACT*, Feb. 2008, pp. 387 – 390.
- [45] L. Hong and K. Ho, "Identification of digital modulation types using the Wavelet transform," in *Proc. IEEE MILCOM*, 1999, pp. 427 – 431.
- [46] K. Ho, W. Prokopiw, and Y. Chan, "Modulation identification of digital signals by the Wavelet transform," in *Proc. IEE Radar, Sonar and Navig.*, 2000, pp. 169 – 176.
- [47] K. Hassan, I. Dayoub, W. Hamouda, and M. Berbineau, "Automatic modulation recognition using Wavelet transform and neural networks in wireless systems." *EURASIP J. Adv. Sig. Proc.*, vol. 2010, 2010.
- [48] E. Hipp, "Modulation classification based on statistical moments," in *Military Communications Conference*, vol. 2, Oct. 1986, pp. 20 – 26.
- [49] S. Soliman and S. Hsue, "Signal classification using statistical moments," *IEEE Trans. Commun.*, vol. 40, pp. 908 – 916, 1992.

-
- [50] E. Azzouz and A. Nandi, *Automatic modulation recognition of communication signals*. Kluwer Academic, 1996.
- [51] A. Hero and H. Mahram, “Digital modulation classification using power moment matrices,” in *Proc. ICASSP*, 1998, pp. 3285 – 3288.
- [52] C. Spooner, “Classification of co-channel communication signals using cyclic cumulants,” in *Conference Record of the Twenty-Ninth Asilomar Conference on Signals, Systems and Computers*, vol. 1, Oct. 1995, pp. 531–536.
- [53] D. Chang and P. Shih, “Cumulants-based modulation classification technique in multipath fading channels,” *IET Communications*, vol. 9, no. 6, pp. 828 – 835, 2015.
- [54] L. Freitas, “Classificação de modulação para rádio cognitivo,” Master’s thesis, UFPA, 2009.
- [55] Z. Sun, Y. Chen, S. Liu, and W. Wang, “Cyclostationarity-based joint domain approach to blind recognition of SCLD and OFDM signals,” *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, 2014.
- [56] W. Gardner and C. Spooner, “Measurement of spectral correlation,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, no. 5, pp. 1111 – 1123, Oct. 1986.
- [57] K. Kim, I. Akbar, K. Bae, J. Um, C. Spooner, and J. Reed, “Cyclostationary approaches to signal detection and classification in cognitive radio,” in *New Frontiers in Dynamic Spectrum Access Networks*, Apr. 2007, pp. 212–215.
- [58] S. Jang, J. Gu, and J. Kim, “A cyclostationary based signal classification using 2d pca,” in *Wireless Communications, Networking and Mobile Computing*, Sept. 2011, pp. 1–4.
- [59] W. Gardner, “Exploitation of spectral redundancy in cyclostationary signals,” *IEEE Signal Processing Magazine*, vol. 8, no. 2, pp. 14–36, Apr. 1991.
- [60] S. Hosseini, H. Amindavar, and J. Ritcey, “A new cyclostationary spectrum sensing approach in cognitive radio,” in *Proc. IEEE SPAWC*, Jun. 2010.
- [61] C. da Silva, W. Headley, J. Reed, and Z. Youping, “The application of distributed spectrum sensing and available resource maps to cognitive radio systems,” in *Information Theory and Applications Workshop*, 2008, pp. 53 – 57.
- [62] K. Kyouwoong, I. Akbar, K. Bae, U. Jung-sun, C. Spooner, and J. Reed, “Cyclostationary approaches to signal detection and classification in cognitive radio,” in *New Frontiers in Dynamic Spectrum Access Networks*, April 2007.

-
- [63] L. Freitas, C. Cardoso, F. Muller, J. Costa, and A. Klautau, "Automatic modulation classification for cognitive radio systems: Results for the symbol and waveform domains," in *LATINCOM*, Sept. 2009, pp. 1 – 6.
- [64] S. Gulati and R. Bhattacharjee, "Automatic blind recognition of noisy and faded digitally modulated M-QAM signals," in *Proc. Annual India Conference*, Sept. 2006, pp. 1 – 6.
- [65] J. Abrahams, "A survey of recent progress on level-crossing problems for random processes," in *Communications and Networks*, I. Blake and H. Poor, Eds. Springer New York, 1986, pp. 6 – 25.
- [66] D. Chu, "Phase digitizing sharpens timing measurements," *IEEE Spectrum*, vol. 25, no. 7, pp. 28–32, Jul. 1988.
- [67] T. Callaghan, J. Perry, and J. Tjho, "Sampling and algorithms aid modulation recognition," *Microwaves*, vol. 24, pp. 117–119, Sep. 1985.
- [68] L. D. Vito and S. Rapuano, "Validating a method for classifying digitally modulated signals," *Measurement*, vol. 42, no. 3, pp. 427 – 435, 2009.
- [69] A. Hazza, M. Shoaib, A. Saleh, and A. Fahd, "Robustness of digitally modulated signal features against variation in HF noise model," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, p. 24, 2011.
- [70] P. Prakasam and M. Madheswaran, "Digital modulation identification model using wavelet transform and statistical parameters," *Journal of Computer Systems, Networks, and Communications*, vol. 2008, no. 6, pp. 1 – 6, 2008.
- [71] M. Aslam, Z. Zhu, and A. Nandi, "Automatic modulation classification using combination of genetic programming and KNN," *IEEE Trans. on Wireless Commun.*, vol. 11, no. 8, pp. 2742 – 2750, Aug. 2012.
- [72] J. Li, C. He, J. Chen, and D. Wang, "Automatic digital modulation recognition based on euclidean distance in hyperspace," *IEICE Trans. Commun.*, vol. E89-B, Oct 2006.
- [73] V. Vapnik, *The nature of statistical learning theory*. Springer Verlag, 1995.
- [74] J. Popoola and R. van Olst, "A novel modulation-sensing method," *IEEE Veh. Technol. Mag.*, vol. 6, no. 3, pp. 60 – 69, Aug. 2011.
- [75] X. Zhou, Y. Wu, and B. Yang, "Signal classification method based on support vector machine and high-order cumulants." *Wireless Sensor Network*, vol. 2, no. 1, pp. 48–52, 2010.

-
- [76] A. Engin, "Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm-support vector machines: HGASVM." *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1391–1402, 2009.
- [77] M. Wong and A. Nandi, "Automatic digital modulation recognition using artificial neural network and genetic algorithm," *Signal Processing*, vol. 84, no. 2, pp. 351 – 365, 2004.
- [78] H. Cheng, H. Han, L. Wu, and L. Chen, "A 1-dimension structure adaptive self-organizing neural network for QAM signal classification," in *Third International Conference on Natural Computation*, vol. 3, Aug 2007, pp. 53 – 57.
- [79] F. Xie, C. Li, and G. Wan, "An efficient and simple method of MPSK modulation classification," in *Wireless Communications, Networking and Mobile Computing*, Oct 2008, pp. 1–3.
- [80] Z. Yu, Y.-Q. Shi, and W. Su, "M-ary frequency shift keying signal classification based-on discrete Fourier transform," in *IEEE MILCOM*, vol. 2, Oct. 2003, pp. 1167 – 1172.
- [81] E. Avci and D. Avci, "The performance comparison of discrete wavelet neural network and discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition," *Expert Systems with Applications*, vol. 35, no. 1 - 2, pp. 90 – 101, 2008.
- [82] G. Zhang, "Neural networks for classification: a survey," *IEEE Trans. Syst., Man and Cyber.*, vol. 30, no. 4, pp. 451 – 462, Nov. 2000.
- [83] Y. Hou and H. Feng, "Study of modulation recognition algorithm based on wavelet transform and neural network." *Journal of Computers*, vol. 6, no. 7, pp. 1511 – 1518, 2011.
- [84] P. Prakasam and M. Madheswaran, "Intelligent decision making system for digital modulation scheme classification in software radio using wavelet transform and higher order statistical moments." *Wireless Personal Communications*, vol. 50, no. 4, pp. 509 – 528, 2009.
- [85] A. Malady and A. Beex, "AMC improvements from robust estimation." in *GLOBECOM*. IEEE, 2010, pp. 1 – 5.
- [86] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Cyclostationarity-based blind classification of analog and digital modulations," in *IEEE Conference on Military Communications*. IEEE Press, 2006, pp. 2176 – 2182.

-
- [87] G. Han, J. Han, and D. Lu, "Study of modulation recognition based on HOCs and SVM," in *Proc. VTC*, vol. 2, May 2004, pp. 898 – 902.
- [88] O. Dobre, S. Rajan, and R. Inkol, "Joint signal detection and classification based on first-order cyclostationarity for cognitive radios." *EURASIP J. Adv. Sig. Proc.*, vol. 2009, 2009.
- [89] M. Pedzisz and A. Mansour, "Automatic modulation recognition of MPSK signals using constellation rotation and its 4th order cumulant." *Digital Signal Processing*, vol. 15, no. 3, pp. 295 – 304, 2005.
- [90] H. Lei, D. Tiandong, and F. Xiangqian, "A novel method of modulation classification for digital signals." in *IJCNN*. IEEE, 2006, pp. 2435 – 2438.
- [91] L. Shen, S. Li, C. Song, and F. Chen, "Automatic modulation classification of MPSK signals using high order cumulants," in *Proc. ICSP*, vol. 1, Nov. 2006.
- [92] M. Wong and A. Nandi, "Efficacies of selected blind modulation type detection methods for adaptive OFDM systems," in *Proc. Int. Conf. on Signal Processing and Communication Systems*, vol. 1, Dec. 2007.
- [93] V. Orlic and M. Dukic, "Multipath channel estimation algorithm for automatic modulation classification using sixth-order cumulants," *EEE Commun. Lett.*, vol. 46, no. 19, pp. 1348–1349, Sept. 2010.
- [94] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Cyclostationarity-based modulation classification of linear digital modulations in flat fading channels," *Wireless Personal Communications*, vol. 54, no. 4, pp. 699 – 717, Sept. 2010.
- [95] D. Boutte and B. Santhanam, "A hybrid ICA-SVM approach to continuous phase modulation recognition," *IEEE Signal Processing Letters*, vol. 16, no. 5, pp. 402 – 405, May 2009.
- [96] S. Theodoridis and K. Koutroumbas, *Pattern recognition*. Academic Press, 2009.
- [97] S. Hsue and S. Soliman, "Automatic modulation classification using zero crossing," *IEEE Radar and Signal Processing*, vol. 137, no. 6, pp. 459 – 464, Dec. 1990.
- [98] S. Hsue, "Automatic modulation recognition of digitally modulated signals," in *Proc. MILCOM*, Oct. 1989, pp. 645 – 649.
- [99] J. Xu, W. Su, and M. Zhou, "Likelihood-Ratio approaches to automatic modulation classification," *IEEE Trans. on Sys.*, vol. 41, no. 4, pp. 455 – 469, Jul. 2011.

-
- [100] J. Sills, "Maximum-likelihood modulation classification for PSK/QAM," in *IEEE MILCOM*, 1999, pp. 57 – 61.
- [101] K. Kim and A. Polydoros, "Digital modulation classification: the BPSK versus QPSK case," in *Proc. IEEE MILCOM*, 1988, pp. 431 – 436.
- [102] C. Long, K. Chugg, and A. Polydoros, "Further results in likelihood classification of QAM signals," in *Proc. IEEE MILCOM*, 1994, pp. 57 – 61.
- [103] C. Huang and A. Polydoros, "Likelihood methods for MPSK modulation classification," *IEEE Trans. Commun.*, vol. 43, pp. 1493 – 1504, 1995.
- [104] L. Hong and K. Ho, "Classification of BPSK and QPSK signals with unknown signal level using the Bayes technique," in *Proc. IEEE ISCAS*, 2003, pp. IV.1 – IV.4.
- [105] J. Xu, S. Wei, and Z. MengChu, "Software-defined radio equipped with rapid modulation recognition," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1659–1667, May 2010.
- [106] W. Wei and J. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 189 – 193, Feb. 2000.
- [107] N. Alyaoui, H. B. Hnia, A. Kachouri, and M. Samet, "The modulation recognition approaches for software radio," in *Proc. SCS*, Nov. 2008, pp. 1 –5.
- [108] Massey and J. Frank, "The Kolmogorov-Smirnov test for goodness of fit," *J. Amer. Stat. Assoc.*, vol. 46, no. 253, pp. 68 – 78, 1951.
- [109] A. Castro, L. Freitas, C. Cardoso, J. Costa, and A. Klautau, *Cognitive Radio Systems*. In-Teh, 2011, ch. Modulation Classification in Cognitive Radio.
- [110] I. Witten and E. Frank, *Data mining: practical machine learning tools and techniques with java implementations*. Morgan Kaufmann, 2005.
- [111] J. Morais, Y. Pires, C. Cardoso, and A. Klautau, "Data mining applied to the electric power industry: Classification of short-circuit in transmission lines," in *Proc. IEEE ISDA*, 2007.
- [112] C. Nello and S. John, *An introduction to support vector machines : and other kernel-based learning methods*, 1st ed. Cambridge University Press, March 2000.
- [113] A. Lorena, "Investigação de estratégias para a geração de máquinas de vetores de suporte multiclassés," Ph.D. dissertation, Instituto de Ciências Matemáticas e de Computação - USP, 2006.

-
- [114] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, “An introduction to kernel-based learning algorithms,” *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181 – 201, Mar. 2001.
- [115] X. Wang and Y. Zhong, “Statistical learning theory and state of the art in SVM,” in *Proc. of the Second IEEE International Conference on Cognitive Informatics*, 2003, pp. 55 – 59.
- [116] Burges and C. J.C., “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 1 – 43, 1998.
- [117] K. Stroeh, “Uma abordagem para correlação de eventos de segurança baseada em técnicas de aprendizado de máquinas,” Master’s thesis, Unicamp, 2009.
- [118] S. Haykin, *Redes Neurais, Princípios e prática*, 2nd ed. Bookman, 1999.
- [119] Wang, *Support Vector Machines: Theory and Applications*, ser. Studies in Fuzziness and Soft Computing, Lipo, Ed. Springer Berlin, 2005, vol. 177.
- [120] G. Bisognin, “Utilização de máquinas de suporte vetorial para predição de estruturas terciárias de proteínas,” Master’s thesis, Universidade do Vale do Rio dos Sinos, 2007.
- [121] V. Kecman, *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. Cambridge, MA, USA: MIT Press, 2001.
- [122] V. Vapnik, *Statistical learning theory*. New York: Wiley, 1998.
- [123] R. Rifkin and A. Klautau, “In defense of one-vs-all classification,” *J. Machine Learning Research*, vol. 5, pp. 101–141, 2004.
- [124] J. B. de Marca, “An LSF quantizer for the north-american half-rate speech coder,” *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 413–419, 1994.
- [125] T. Bragatto, “Uma nova metodologia para categorização de veículos terrestres baseada em momentos e métricas de desempenho e complexidade computacional,” Master’s thesis, Faculdade de Tecnologia. Universidade de Brasília - UNB, 2009.
- [126] G. Corder and D. Foreman, *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*. Wiley, 2009.
- [127] F. Hameed, O. Dobre, and D. Popescu, “On the likelihood-based approach to modulation classification,” *IEEE Trans. Wirel. Commun.*, vol. 8, pp. 5884 – 5892, 2009.

- [128] C. Chang and C. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [129] A. Castro, “Implementação em FPGA de algoritmos para classificação de modulação em rádio cognitivo,” Ph.D. dissertation, Universidade Federal do Pará - UFPA, 2011.