

ELOÁ ALANO DAMIAN

**DUAS METODOLOGIAS APLICADAS À CLASSIFICAÇÃO DE
PRECIPITAÇÃO CONVECTIVA E ESTRATIFORME COM RADAR
METEOROLÓGICO: SVM E K-MEANS**

CURITIBA

SETEMBRO 2011

ELOÁ ALANO DAMIAN

**DUAS METODOLOGIAS APLICADAS À CLASSIFICAÇÃO DE
PRECIPITAÇÃO CONVECTIVA E ESTRATIFORME COM RADAR
METEOROLÓGICO: SVM E K-MEANS**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciências, pelo Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Área de Concentração em Programação Matemática, Setores de Tecnologia e Ciências Exatas da Universidade Federal do Paraná.

Orientador: Prof. Dr. Arinei Carlos Lindbeck da Silva

CURITIBA

SETEMBRO 2011

Termo de Aprovação

ELOÁ ALANO DAMIAN

DUAS METODOLOGIAS APLICADAS À CLASSIFICAÇÃO DE PRECIPITAÇÃO CONVECTIVA E ESTRATIFORME COM RADAR METEOROLÓGICO: SVM E K-MEANS

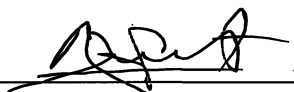
Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Ciências, pelo Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Área de Concentração em Programação Matemática, Setores de Tecnologia e Ciências Exatas da Universidade Federal do Paraná, pela seguinte banca examinadora:



Prof. Dr. Arinei Carlos Lindbeck da Silva
Universidade Federal do Paraná



Prof. Dr. Volmir Eugênio Wilhelm
Universidade Federal do Paraná



Dr. Reinaldo Bomfim da Silveira
Instituto Tecnológico SIMEPAR

Dedicatória

A meus pais, Airton e Enoê,
minha irmã Alana,
meu namorado Ricardo.

Agradecimentos

Aos meus pais, Airton e Enoê, por serem a razão de tudo o que sou hoje e por estarem presentes com amor e paciência em todos os momentos da minha vida.

À minha irmã Alana, pelo carinho e força que me proporciona e por estar sempre comigo em grandes momentos.

Ao meu namorado Ricardo, pelo companheirismo e amizade, por me ajudar nas horas difíceis e por seu amor incondicional.

À minha amiga Vanessa, por sua amizade, companheirismo e momentos compartilhados.

A todos os colegas do SIMEPAR, em especial a Marco Aurélio e André, por todo apoio e amizade. À Ana Beatriz, pela colaboração e disponibilidade em ajudar com os dados do trabalho. Ao Cesar Beneti, pelos esclarecimentos, sugestões e motivação. Ao Reinaldo Silveira, pelas importantes colaborações.

Ao professor Arinei, por seus ensinamentos, orientações e paciência.

Aos professores e colegas do PPGMNE, que contribuíram decisivamente para esta grande etapa.

A todos, os meus sinceros agradecimentos.

Epígrafe

“Uma mente que se abre a uma nova ideia jamais volta ao seu tamanho original.”

Albert Einstein

Sumário

Lista de Figuras	viii
Lista de Tabelas	xiii
Lista de Siglas	xiv
Lista de Símbolos	xv
Resumoxviii
Abstract	xix
1 Introdução	1
1.1 Importância	1
1.2 Objetivos	2
1.3 Limitações	2
1.4 Estrutura do Trabalho	2
2 Conceitos Meteorológicos	4
2.1 Radar Meteorológico	4
2.1.1 Coleta de Dados	6
2.1.2 Variáveis Medidas	9
2.2 Interpolação dos Dados	12
2.3 Formas de Visualização e Produtos Gerados	13
2.3.1 PPI - Plan Position Indicator	14
2.3.2 CAPPI - Constant Altitude of Plan Position Indicator	14

2.3.3	Zmax - Refletividade Máxima	15
2.3.4	VIL - Vertically Integrated Liquid-Water	16
2.3.5	Echo Top	16
2.3.6	VIL Density	17
2.4	Definições	17
2.4.1	Isoterma de Zero Grau	17
2.4.2	Banda Brilhante	18
2.4.3	Perfil Vertical de Refletividade	18
2.4.4	Alturas Relevantes	19
2.4.5	Zmax da Banda Brilhante	20
2.4.6	Ice Slope	20
2.5	Relação Z-R	20
2.6	Precipitações Convectivas e Estratiformes	21
2.6.1	Método de Steiner <i>et al.</i>	23
3	Support Vector Machine	26
3.1	Conceitos da Teoria da Otimização	26
3.1.1	Problema Primal	26
3.1.2	Teoria Lagrangeana	29
3.1.3	Dualidade	30
3.2	Reconhecimento de Padrões	31
3.2.1	Inteligência Artificial	32
3.2.2	Aprendizado de Máquina	32
3.3	Teoria do Aprendizado Estatístico	34
3.3.1	Minimização do Risco Empírico	35
3.3.2	Dimensão VC	36
3.3.3	Minimização do Risco Estrutural	37

3.4	O Hiperplano Ótimo	39
3.4.1	Hiperplano Ótimo para Classes Linearmente Separáveis - Margens Rígidas	39
3.4.2	Hiperplano Ótimo para Classes Não-Linearmente Separáveis - Margens Flexíveis .	45
3.5	Função Kernel	46
4	K-means	50
4.1	Análise Multivariada	50
4.2	Análise de Agrupamentos	50
4.2.1	O método <i>K-means</i>	53
5	Metodologia.....	55
5.1	Radar Meteorológico do SIMEPAR	55
5.1.1	Especificações Técnicas	55
5.2	Dados e Variáveis Utilizadas	56
5.2.1	Normalização	58
5.3	O <i>software</i> R	59
5.3.1	SVM	59
5.3.2	<i>K-means</i>	61
6	Aplicação e Resultados	62
6.1	SVM	62
6.2	K-means	79
6.3	Comparação entre os Métodos	83
7	Conclusão.....	89
7.1	Sugestões para trabalhos futuros	91
	Referências Bibliográficas.....	92

Lista de Figuras

Figura 2.1	(a) Ângulo de elevação da antena do radar (b) Ângulo azimute da antena do radar (c) Distância do alvo em relação ao radar	6
Figura 2.2	<i>Volume Covarage Patterns</i> (VCP)	7
Figura 2.3	Variação do volume analisado devido à abertura do feixe	7
Figura 2.4	Varredura PPI	8
Figura 2.5	Varredura RHI	8
Figura 2.6	Imagem de radar para a variável refletividade	10
Figura 2.7	Imagem de radar para a variável velocidade radial	11
Figura 2.8	Imagem de radar para a variável largura espectral	12
Figura 2.9	Espaçamento entre os bins	13
Figura 2.10	Produto PPI	14
Figura 2.11	Produto CAPPI	15
Figura 2.12	Produto Zmax	16

Figura 2.13 Produto <i>Echo Top</i>	17
Figura 2.14 Exemplos de PVR's	19
Figura 2.15 Ilustração do <i>Ice Slope</i>	20
Figura 2.16 Ilustração do raio convectivo e raio de fundo para um determinado pixel ...	24
Figura 2.17 Relação entre a média de refletividade ao redor do ponto e o raio de alcance convectivo	24
Figura 3.1 Exemplo de (a) conjunto convexo e (b) conjunto não-convexo	28
Figura 3.2 Função convexa	28
Figura 3.3 Indução de classificador em aprendizado supervisionado	33
Figura 3.4 (a) <i>Underfitting</i> ; (b) Hiperplano de separação mais adequado; (c) <i>Overfitting</i>	34
Figura 3.5 Possibilidades de rotulação de três amostras no \mathfrak{R}^2 e a classificação realizada por uma função linear	37
Figura 3.6 Princípio de minimização do risco estrutural	38
Figura 3.7 Classificação dos dados a partir do hiperplano separador	40
Figura 3.8 Hiperplano com máxima margem de separação	41
Figura 3.9 Possibilidades de separação das classes por diferentes hiperplanos	42

Figura 3.10 Cálculo da distância d entre os hiperplanos H_1 e H_2	43
Figura 3.11 Possibilidades de valores para a variável ξ	45
Figura 3.12 Mapeamento do espaço de entrada via função <i>kernel</i>	47
Figura 4.1 Dendograma ilustrando agrupamento hierárquico	52
Figura 4.2 Ilustração do algoritmo <i>K-means</i>	54
Figura 6.1 Índices de acerto para <i>kernel</i> linear	67
Figura 6.2 Índices de acerto para <i>kernel</i> radial	67
Figura 6.3 Índices de acerto para <i>kernel</i> sigmoidal	68
Figura 6.4 Índices de acerto para <i>kernel</i> polinomial	68
Figura 6.5 Índice de melhoria para <i>kernel</i> linear	73
Figura 6.6 Índice de melhoria para <i>kernel</i> radial	73
Figura 6.7 Índice de melhoria para <i>kernel</i> sigmoidal	74
Figura 6.8 Índice de melhoria para <i>kernel</i> polinomial	74
Figura 6.9 Resultados obtidos pelo método SVM e comparação com a técnica de Steiner <i>et al.</i> , para conjunto de dados 1	75

Figura 6.10 Resultados obtidos pelo método SVM e comparação com a técnica de Steiner <i>et al.</i> , para conjunto de dados 2	76
Figura 6.11 Resultados obtidos pelo método SVM e comparação com a técnica de Steiner <i>et al.</i> , para conjunto de dados 3	76
Figura 6.12 Resultados obtidos pelo método SVM e comparação com a técnica de Steiner <i>et al.</i> , para conjunto de dados 4	77
Figura 6.13 Resultados obtidos pelo método SVM e comparação com a técnica de Steiner <i>et al.</i> , para conjunto de dados 5	77
Figura 6.14 Resultados obtidos pelo método SVM e comparação com a técnica de Steiner <i>et al.</i> , para conjunto de dados 6	78
Figura 6.15 Resultados obtidos pelo método SVM e comparação com a técnica de Steiner <i>et al.</i> , para conjunto de dados 7	78
Figura 6.16 Resultados obtidos pelo método <i>K-means</i> e comparação com a técnica de Steiner <i>et al.</i> para conjunto de dados 1	80
Figura 6.17 Resultados obtidos pelo método <i>K-means</i> e comparação com a técnica de Steiner <i>et al.</i> para conjunto de dados 2	80
Figura 6.18 Resultados obtidos pelo método <i>K-means</i> e comparação com a técnica de Steiner <i>et al.</i> para conjunto de dados 3	81
Figura 6.19 Resultados obtidos pelo método <i>K-means</i> e comparação com a técnica de Steiner <i>et al.</i> para conjunto de dados 4	81
Figura 6.20 Resultados obtidos pelo método <i>K-means</i> e comparação com a técnica de Steiner	

<i>et al.</i> para conjunto de dados 5	82
Figura 6.21 Resultados obtidos pelo método <i>K-means</i> e comparação com a técnica de Steiner <i>et al.</i> para conjunto de dados 6	82
Figura 6.22 Resultados obtidos pelo método <i>K-means</i> e comparação com a técnica de Steiner <i>et al.</i> para conjunto de dados 7	83
Figura 6.23 Imagem comparativa entre a classificação feita por SVM e por <i>K-means</i> , para conjunto de dados 1	85
Figura 6.24 Imagem comparativa entre a classificação feita por SVM e por <i>K-means</i> , para conjunto de dados 2	85
Figura 6.25 Imagem comparativa entre a classificação feita por SVM e por <i>K-means</i> , para conjunto de dados 3	86
Figura 6.26 Imagem comparativa entre a classificação feita por SVM e por <i>K-means</i> , para conjunto de dados 4	86
Figura 6.27 Imagem comparativa entre a classificação feita por SVM e por <i>K-means</i> , para conjunto de dados 5	87
Figura 6.28 Imagem comparativa entre a classificação feita por SVM e por <i>K-means</i> , para conjunto de dados 6	87
Figura 6.29 Imagem comparativa entre a classificação feita por SVM e por <i>K-means</i> , para conjunto de dados 7	88

Lista de Tabelas

Tabela 2.1	Bandas de radar e seus correspondentes intervalos de frequência e comprimento de onda	5
Tabela 3.1	Funções <i>kernel</i>	49
Tabela 5.1	Data e hora associados a cada um dos conjuntos de dados	56
Tabela 5.2	Funções <i>kernel</i>	60
Tabela 6.1	Configurações de parâmetros para as funções <i>kernel</i>	63
Tabela 6.2	Tempo de execução e número de vetores suporte para cada <i>kernel</i> estabelecido	64
Tabela 6.3	Índices de acerto do método SVM para cada conjunto de dados	66
Tabela 6.4	Índices de acerto do método SVM para conjuntos de treinamento e de teste ..	70
Tabela 6.5	Índices de melhoria para conjunto de treinamento, conjunto de teste e conjunto total de dados	72
Tabela 6.6	Índice de semelhança entre os dados classificados por especialista e métodos Steiner <i>et al.</i> , SVM e <i>K-means</i>	84

Lista de Siglas

<i>AM</i>	Aprendizado de Máquina
<i>CAPPI</i>	<i>Constant Altitude of Plan Position Indicator</i>
<i>IA</i>	Inteligência Artificial
<i>KKT</i>	<i>Karush-Kuhn-Tucker</i>
<i>MVP</i>	Mapeamento por Vizinhos Próximos
<i>PPI</i>	<i>Plan Position Indicator</i>
<i>PRF</i>	<i>Pulse Repetition Frequency</i>
<i>PVR</i>	Perfil Vertical de Refletividade
<i>RHI</i>	<i>Range Height Indicator</i>
<i>SVM</i>	<i>Support Vector Machine</i>
<i>TAE</i>	Teoria do Aprendizado Estatístico
<i>TVVR</i>	Taxa Vertical de Variação de Refletividade
<i>VC</i>	<i>Vapnik & Chervonenkis</i>
<i>VCP</i>	<i>Volume Covarage Patterns</i>
<i>VIL</i>	<i>Vertically Integrated Liquid-Water</i>
<i>VIL</i>	<i>VIL Density</i>
<i>VS</i>	Vetores Suporte

Lista de Símbolos

$\ \cdot \ $	Norma
b	Bias
c	Velocidade da luz
c_m	Velocidade de propagação de ondas eletromagnéticas no meio
c_r	Constante do radar
C_i	Cluster i
C^1	Conjunto das funções contínuas em Ω
$c(f(x), y)$	Função custo
$d(A, B)$	Distância euclidiana entre objetos A e B
Δt	Intervalo de tempo entre transmissão e recepção do pulso
f	Frequência
F	Conjunto de possíveis funções para escolha de um hiperplano separador
$f(x)$	Função objetivo do problema primal
g_i	Restrição de desigualdade no problema primal
h_j	Restrição de igualdade no problema primal
h_{30}	Altura da refletividade de 30dBz
h_{45}	Altura da refletividade de 45dBz
h_{Zmax}	Altura da refletividade máxima
$I_{melhoria}$	Índice de melhoria
$indices_{SVM}$	Índice de acerto do método SVM

$indice_{st}$	Índice de acerto do método de Steiner <i>et al.</i> (1995)
$K(x_i, x_j)$	Função kernel
$L(w, \alpha, \beta)$	Função Lagrangeana
M_+	Margem positiva
M_-	Margem negativa
p_r	Energia refletida para o radar
$P(x, y)$	Função de probabilidade
r	Distância do alvo ao radar
R	Taxa de precipitação
$R(c)$	Risco esperado
$R_{emp}(f)$	Risco empírico
\mathfrak{R}^n	Espaço de dimensão n
V_r	Velocidade radial
x_i	Vetor de características
$(x_i \cdot x_j)$	Produto interno
w	Vetor de pesos
W	Largura espectral
y_i	Classe correspondente à entrada x_i
Z	Refletividade
z_{fundo}	Média de refletividade dos pontos dentro de uma determinada circunferência
Z_{max}	Refletividade máxima
$Z_{max\ BB}$	Refletividade máxima dentro do intervalo de banda brilhante
$\frac{\partial f}{\partial x}$	Derivada parcial da função f em relação à variável x
α_i	Multiplicador de Lagrange associado à restrição de desigualdade

β_j	Multiplicador de Lagrange associado à restrição de igualdade
θ	Ângulo de elevação
$\theta(\alpha, \beta)$	Função objetivo do problema dual
λ	Comprimento de onda
ξ_i	Variável de folga do ponto x_i
ρ	Margem de separação
ϕ	Ângulo azimute
$\phi : X \mapsto Z$	Mapeamento para o espaço característico
Ω	Subespaço em \mathfrak{R}^n

Resumo

A influência da chuva nas diversas áreas da atividade humana torna indispensável o seu monitoramento. A maioria das precipitações pode ser classificada em dois grupos: sistemas convectivos e estratiformes. Os sistemas estratiformes são caracterizados por chuvas leves e moderadas e grande área de cobertura. Já os convectivos estão associados com altas taxas de precipitação. Devido às acentuadas diferenças que existem entre os dois sistemas, ser capaz de classificá-los contribui para a melhora da estimativa de precipitação. Além disso, a classificação torna possível também a identificação, previsão e acompanhamento de tempestades severas (sistemas convectivos), resultados que podem ainda ser utilizados nos alertas de temporais, enchentes e inundações, com aplicações diretas em diversas áreas da atividade humana (energia, agricultura, turismo, defesa civil, entre outros). Tudo isso justifica a necessidade de um critério para separar e classificar os diferentes tipos de chuva que podem ocorrer na região de abrangência do radar. O objetivo deste trabalho é classificar os sistemas meteorológicos em convectivo ou estratiforme, de acordo com as características peculiares a cada evento. Para tanto, duas técnicas são aplicadas. A primeira é um método de reconhecimento de padrões, o chamado *Support Vector Machine* (SVM), técnica que, a partir de dados previamente classificados, cria um hiperplano separador entre as classes, através do qual novos dados podem ser classificados. Assumindo como padrão imagens de radar classificadas por especialista, os resultados obtidos pelo método apresentam um maior índice de acerto, quando comparados com as soluções obtidas por uma técnica de separação da chuva já existente. O segundo método abordado é uma técnica de análise de agrupamentos, o *K-means*, que classifica um conjunto de dados em um determinado número de grupos, baseado na similaridade ou dissimilaridade entre os dados. Em relação à classificação feita pelo especialista, o SVM apresenta melhores resultados. Porém, o *K-means* mostra resultados melhores quando a comparação é feita com a técnica de classificação já existente na literatura. Além disso, o *K-means* mostrou-se mais eficiente com relação a tempo de execução, e possui a vantagem de não necessitar de dados previamente classificados, já que a classificação, neste caso, é muito subjetiva. Quando a comparação é feita entre as duas técnicas propostas, observa-se que mais de 85% dos dados obtiveram a mesma classificação.

Palavras-chave: Reconhecimento de Padrões; *Support Vector Machine*; Análise de Agrupamentos; *K-means*; Meteorologia; Classificação Meteorológica; Radar Meteorológico.

Abstract

The influence of rainfall in several areas of human activity makes indispensable its monitoring. Most precipitation events can be classified in two groups: convective and stratiform systems. Stratiform systems are characterized by light and moderate rain and a large coverage area. Convective rain is associated to high rainfall rate. Due to differences which exist between the two systems, being capable of classification contributes to the improvement on precipitation estimation. Moreover, classification also allows for identification, forecasting and tracking of severe storms (convective systems), results that can yet be used on alerts for thunderstorms, increasing streamflow and flooding, with direct application to several areas of human activity (energy, agriculture, tourism, civil defense, among others). All these justify the needs for a criteria for classification and separation of different types of rain that may occur within weather radar coverage area. The objective of this work is to classify convective and stratiform meteorological systems, in accordance with the peculiar characteristics to each event. For that, two techniques were applied. The first one is a pattern recognition method called Support Vector Machine (SVM), a technique that, from previously classified data, creates a hyperplane separating between classes, through which new data can be recognized. Assuming radar images classified by a specialist, the results obtained with this method presented a better hit index, when compared with the results from a previous technique for rainfall separation. The second method evaluated is a cluster analysis, K-means, which classifies a data set in a number of determined groups or clusters, based on similarities or dissimilarities between the data. In relation to the classification made by the specialist, SVM presents better results. However, K-means showed better results when compared to a technique already existent in the literature. Moreover, K-means revealed more efficient with regard to execution time, and had the advantage of no need of data previously classified, since classification in this case is very subjective. When the comparison is made between the two proposed techniques, it was observed that more than 85% of the data got the same classification.

Key-words: Pattern Recognition; Support Vector Machine; Cluster Analysis; K-means; Meteorology; Meteorological Classification; Weather Radar.

1 Introdução

A influência da precipitação nas diversas áreas da atividade humana faz com que seu monitoramento seja uma tarefa indispensável. Entretanto, devido à grande aleatoriedade e variabilidade espacial e temporal dessa variável, torna-se difícil a sua identificação e quantificação. Para monitorar a precipitação, os principais instrumentos são: os pluviômetros, os pluviógrafos, os radares e os sensores que operam a bordo de satélites (MOL, 2005).

Estimativas de precipitação acuradas são de crucial importância para o sucesso de qualquer modelagem hidrológica (FILHO; CRAWFORD, 1999). Nesse aspecto, o radar meteorológico é uma das melhores ferramentas disponíveis atualmente para a elaboração de tais estimativas. As vantagens principais oferecidas por ele são o monitoramento em tempo real de uma ampla área com alta resolução espacial e temporal e a possibilidade de observar a estrutura tridimensional do sistema que deu origem à precipitação (SAUVAGEOT, 1994).

A maioria das precipitações pode ser classificada em dois grupos: sistemas convectivos e estratiformes. Sistemas convectivos estão associados com fortes campos verticais de vento e altas taxas pluviométricas; enquanto nos sistemas estratiformes predomina extensa área de cobertura, baixas velocidades verticais de vento, homogeneidade horizontal e taxas de precipitação menores (ANAGNOSTOU; KUMMEROW, 1997). Devido às diferenças existentes entre os dois sistemas, ser capaz de classificá-los contribui para a melhora da estimativa de precipitação obtida através dos dados medidos por sensores remotos tanto em terra quanto no espaço (SIMPSON; ADLER; NORTH, 1988).

Neste estudo, são apresentadas duas metodologias para classificação das precipitações em dois grupos: estratiforme ou convectivo, com base em dados de radar.

1.1 Importância

A classificação torna possível a identificação, previsão e acompanhamento de tempestades severas (precipitação convectiva), resultados que podem ainda ser utilizados nos alertas

de temporais, enchentes e inundações, com aplicações diretas em diversos setores da sociedade (defesa civil, energia, agricultura, turismo, entre outros).

Tudo isso justifica a necessidade de um critério para separar e classificar os diferentes eventos meteorológicos que podem ocorrer na região de abrangência do radar.

1.2 Objetivos

O objetivo geral desse estudo é classificar os eventos meteorológicos detectados por radar em convectivo ou estratiforme, utilizando um método de reconhecimento de padrões, o *Support Vector Machine* (SVM), e uma técnica de análise de agrupamentos, o *K-means*.

Como objetivos específicos, tem-se:

- pesquisar métodos de classificação de chuva já existentes na literatura;
- expor conceitos da área de Meteorologia, especialmente sobre o radar meteorológico;
- estudar o método *Support Vector Machine*, detalhando toda a teoria envolvida;
- apresentar a técnica de agrupamento *K-means*;
- aplicar as duas metodologias propostas utilizando dados reais obtidos por radar meteorológico;
- analisar os resultados obtidos por ambas as técnicas, comparando-os entre elas e também com alguma técnica de classificação de chuva já existente.

1.3 Limitações

Este trabalho limita-se a utilizar apenas dados de radar meteorológico.

1.4 Estrutura do Trabalho

O trabalho está dividido em 7 capítulos, incluindo este.

O Capítulo 1 apresenta uma introdução ao tema, na qual é mostrada a importância do trabalho e os seus objetivos.

O Capítulo 2 mostra alguns conceitos sobre radar aplicado à Meteorologia.

O Capítulo 3 apresenta todos os conceitos necessários para o entendimento do SVM, além de descrever com detalhes o método.

O Capítulo 4 apresenta a segunda técnica a ser utilizada no trabalho, o *K-means*.

O Capítulo 5 descreve a metodologia aplicada no estudo, assim como os dados utilizados.

O Capítulo 6 apresenta os resultados obtidos e análise de tais, incluindo comparação entre os resultados de cada técnica.

O Capítulo 7 apresenta a conclusão do trabalho e sugestões para trabalhos futuros.

2 Conceitos Meteorológicos

Este capítulo apresenta alguns fundamentos básicos sobre o radar aplicado à Meteorologia.

2.1 Radar Meteorológico

A palavra **radar** é o acrônimo do termo em inglês *RA*dio *DE*tectio*N* *AN*d *R*ang*IN*g (Rádio Detecção e Localização). Desenvolvido durante a Segunda Guerra Mundial com fins militares, o radar tornou-se indispensável nas décadas seguintes, devido ao seu constante aprimoramento e utilização, tanto por usuários militares quanto civis (RINEHART, 2004). Suas aplicações mais conhecidas atualmente são no controle de tráfego aéreo e marítimo, e no monitoramento de eventos meteorológicos.

Muito do que é conhecido da estrutura interna de sistemas de precipitação foi obtido através da utilização de radar meteorológico. A tecnologia atual permite que, além de detectar a existência de nuvens e precipitação, os radares possam estimar a velocidade das partículas presentes na atmosfera, se elas estão na forma líquida ou sólida e a turbulência existente na nuvem. Essas informações permitem estimar com mais precisão quais foram os eventos que deram origem ao fenômeno meteorológico e qual será seu comportamento no futuro.

A radiação eletromagnética é composta por oscilações de campos elétrico e magnético associados, e que se propagam através do espaço transportando energia à velocidade da luz. Duas características importantes da radiação eletromagnética são a sua frequência e o seu comprimento de onda, que estão relacionadas com a velocidade da luz através da equação:

$$f = \frac{c}{\lambda}, \quad (2.1)$$

onde f é a frequência em Hz , c é a velocidade da luz em m/s e λ é o comprimento de onda em m .

As frequências utilizadas pelos radares variam entre $100MHz$ e $100GHz$, classificadas

em bandas. A Tabela 2.1 lista as bandas do espectro eletromagnético que comumente são utilizadas pelos radares, com suas respectivas frequências e comprimentos de onda (RINEHART, 2004).

Tabela 2.1: Bandas de radar e seus correspondentes intervalos de frequência e comprimento de onda

Banda	Frequência	Comprimento de onda
HF	3 - 30 MHz	100 - 10 m
VHF	30 - 300 MHz	10 - 1 m
UHF	300 - 1000 MHz	1 - 0,3 m
L	1 - 2 GHz	30 - 15 cm
S	2 - 4 GHz	15 - 8 cm
C	4 - 8 GHz	8 - 4 cm
X	8 - 12 GHz	4 - 2,5 cm
Ku	12 - 18 GHz	2,5 - 1,7 cm
K	18 - 27 GHz	1,7 - 1,2 cm
Ka	27 - 40 GHz	1,2 - 0,75 cm
mm ou W	40 - 300 GHz	7,5 - 1 mm

Fonte: (RINEHART, 2004)

O radar funciona da seguinte forma: um transmissor emite um feixe de radiação eletromagnética com intensidade e frequência conhecidas. Quando a radiação atinge os hidrometeoros na atmosfera, parte dessa energia volta para o radar, sendo então detectada por um receptor e amplificada, uma vez que a intensidade da energia refletida é muito menor do que aquela que foi emitida (COLLIER, 1996).

Conhecendo-se o intervalo de tempo Δt entre a emissão e o retorno do eco, é possível determinar a distância r do alvo. Considerando a velocidade de propagação das ondas eletromagnéticas no meio igual a c_m , o intervalo de tempo Δt é relacionado à distância r pela seguinte equação:

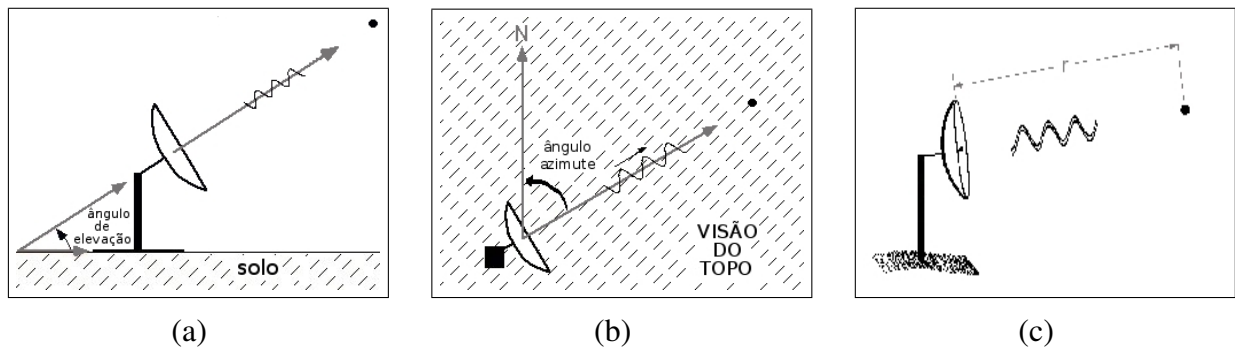
$$r = \frac{c_m \Delta t}{2}. \quad (2.2)$$

2.1.1 Coleta de Dados

Varredura é o movimento da antena do radar durante a coleta dos dados. A antena do radar normalmente é rotacionada sob um eixo vertical, varrendo o horizonte em todas as direções em torno do radar (RINEHART, 2004).

Existem três informações para determinar a localização de um alvo. A primeira é o ângulo do feixe em relação ao solo, chamado de *ângulo de elevação* (Figura 2.1-a). A segunda é o ângulo do feixe de radar em relação ao Norte Geográfico, chamado de *ângulo de azimute* (Figura 2.1-b). A terceira informação necessária é a *distância (r)* do radar ao alvo (Figura 2.1-c).

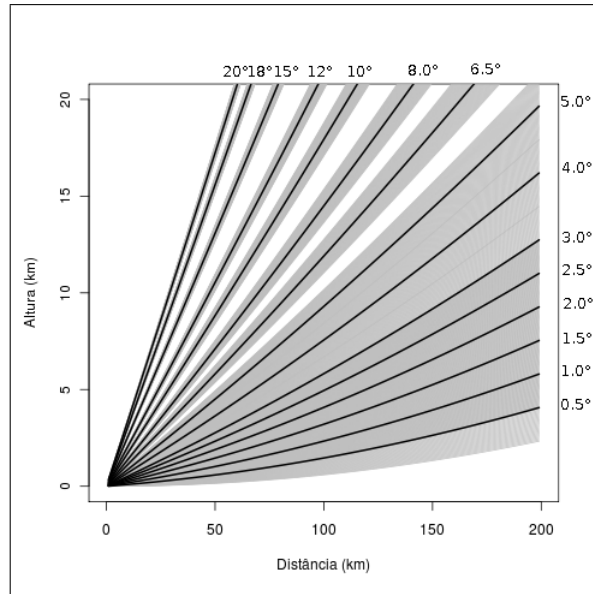
Figura 2.1: (a) Ângulo de elevação da antena do radar (b) Ângulo azimute da antena do radar (c) Distância do alvo em relação ao radar



Fonte: (ILLINOIS, 2010)

O conjunto de elevações é denominado *Volume Coverage Patterns (VCP)* e varia de acordo com a ocorrência de precipitação e sua intensidade. Um VCP de 15 elevações pode ser observado na Figura 2.2.

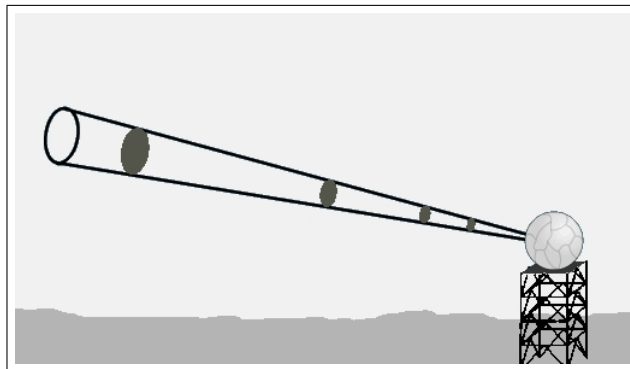
Figura 2.2: *Volume Covarage Patterns (VCP)*



Fonte: A autora (2011)

Devido à abertura do feixe (*beam width*), o feixe apresenta uma forma cônica e quanto maior a distância do radar, maior é o volume analisado, como pode ser observado na Figura 2.3.

Figura 2.3: Variação do volume analisado devido à abertura do feixe



Fonte: (METEOPT, 2011)

Quando o sinal refletido retorna, o *software* do radar se encarrega de separá-lo em intervalos com comprimento constante (*gate size*, usualmente entre 300 e 1000m) na direção radial. Cada um dos volumes criados por essa separação é denominado *bin* e recebe o valor da média ponderada dos alvos detectados naquela amostra. O conjunto final de todas as medições feitas nas elevações especificadas no VCP recebe o nome de Volume. Utilizando como exemplo um radar com raio de alcance de 200km, abertura do feixe de 1°, *gate size* de 0,5km e 7

elevações, seu volume será composto por $1008000 \left(360 \times \frac{200}{0,5} \times 7 \right)$ medições.

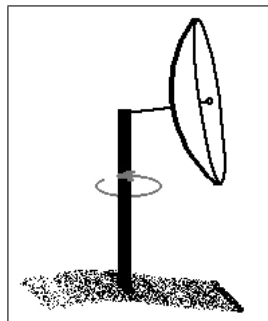
Formas de Coletas

A coleta de dados pode ser feita de duas formas diferentes:

1. PPI (*Plan Position Indicator*)

O radar é operado com antena com ângulo de elevação fixo e azimute variável, como mostrado na Figura 2.4. As variáveis coletadas (Seção 2.1.2) podem ser mapeadas em um plano horizontal.

Figura 2.4: Varredura PPI

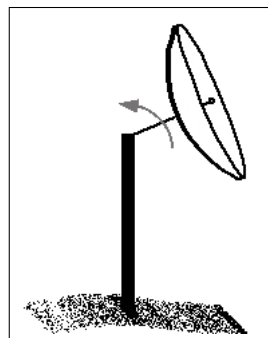


Fonte: (ILLINOIS, 2010)

2. RHI (*Range Height Indicator*)

O radar é operado com antena com ângulo de azimute constante e ângulo de elevação variável, como mostrado na Figura 2.5. As variáveis coletadas (Seção 2.1.2) podem ser mapeadas em um plano vertical.

Figura 2.5: Varredura RHI



Fonte: (ILLINOIS, 2010)

A taxa com a qual o radar emite os pulsos de energia é chamada de *Frequência de Repetição de Pulso* ou *PRF (Pulse Repetition Frequency)*, e é expressa em pulsos por segundo ou *Hertz*. A maioria dos radares convencionais opera com PRF entre 150 e 300Hz. Já os radares com capacidade Doppler - aqueles que medem a velocidade de deslocamento dos alvos, operam entre 700 e 3000Hz (RINEHART, 2004).

2.1.2 Variáveis Medidas

As variáveis coletadas pelo radar meteorológico em estudo são: refletividade (Z), velocidade radial (V_r) e largura espectral (W).

Refletividade (Z)

É possível distinguir as características dos diferentes hidrometeoros (e dos eventos meteorológicos por eles formados) com base na intensidade da energia refletida. A energia refletida por um alvo depende da sua *refletividade* (Z), que é uma medida de eficiência de um alvo em interceptar e retornar a energia originalmente emitida pelo radar. Tal característica depende do tamanho, formato, aspecto e propriedades dielétricas do alvo (MASSAMBANI, 2009). Ela é calculada a partir da energia refletida para o radar (p_r) utilizando-se a seguinte equação:

$$z = c_r r^2 p_r, \quad (2.3)$$

onde r é a distância do radar ao alvo, e c_r é chamado de constante do radar, que agrupa uma série de características do radar, como comprimento de onda, forma e largura do feixe, comprimento do pulso, potência transmitida, ganho da antena, entre outros.

Refletividade é um parâmetro meteorológico que é determinado pelo número e tamanho das partículas presentes em um volume de amostra, o que pode variar de pequenos valores (névoa) para valores muito grandes (granizo). Devido ao grande domínio em que os valores resultantes se encontram, é conveniente utilizar sua representação em uma escala logarítmica:

$$Z = 10 \log \left(\frac{z}{1 \text{mm}^6 / \text{m}^3} \right), \quad (2.4)$$

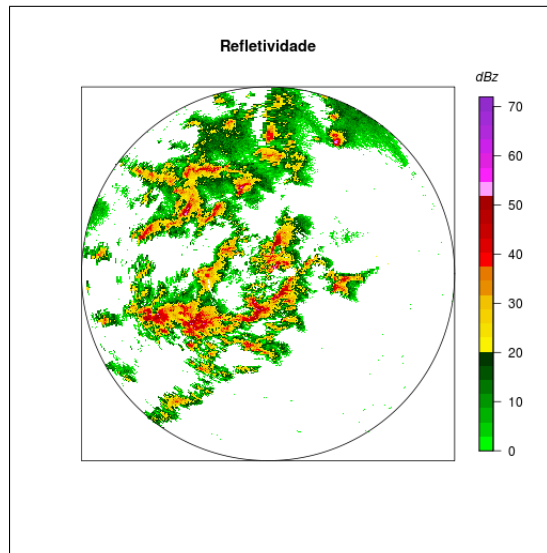
onde Z é valor de refletividade em escala logarítmica, medida em dBz , e z é o valor de refletividade em escala linear, em mm^6 / m^3 .

Normalmente, para visualização de imagens de refletividade, são utilizadas tonalidades

vermelhas para representar valores elevados (regiões convectivas) e tonalidades azuis e verdes para valores intermediários e baixos (estratiformes).

A Figura 2.6 mostra uma imagem da variável refletividade, com escala em dBz .

Figura 2.6: Imagem de radar para a variável refletividade



Fonte: A autora (2011)

Os valores elevados de refletividade estão associados à chuva intensa e quando ultrapassam cerca de $50dBz$, podem estar presentes hidrometeoros no estado sólido e há possibilidade de queda de granizo.

Velocidade Radial (V_r)

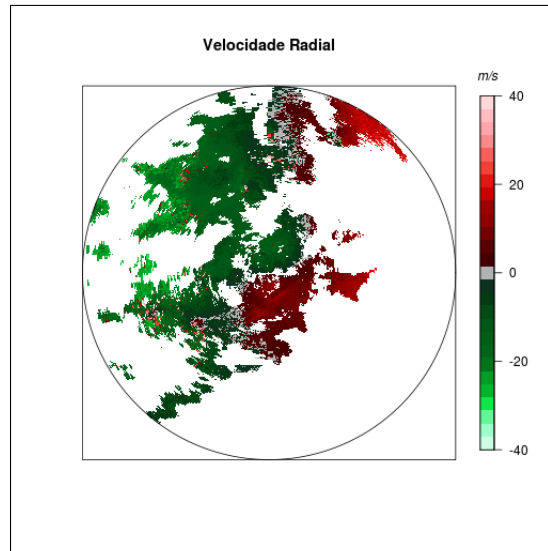
O grande diferencial do radar Doppler é que além de estimar a posição da tempestade, ele também é capaz de calcular sua *velocidade radial* (V_r), que é a velocidade de aproximação ou afastamento dos alvos (hidrometeoros) em relação ao radar na direção do feixe, o que facilita prever para onde o evento se deslocará.

Para isso, é utilizado o efeito Doppler, deduzido por Christian J. Doppler em 1842, e segundo o qual, a mudança na frequência causada por fontes móveis é diretamente proporcional à velocidade da fonte.

O mesmo princípio aplica-se à radiação eletromagnética do radar. Neste caso, o radar está parado e o alvo está se movendo. Se o alvo se move em direção ao radar, a frequência é aumentada; se ele está se afastando, a frequência é reduzida (RINEHART, 2004).

O vento radial é normalmente mostrado em m/s . A Figura 2.7 mostra a velocidade radial captada pelo radar, com valores positivos para os alvos que se afastam (em vermelho) e valores negativos para os alvos que se aproximam do radar (em verde).

Figura 2.7: Imagem de radar para a variável velocidade radial

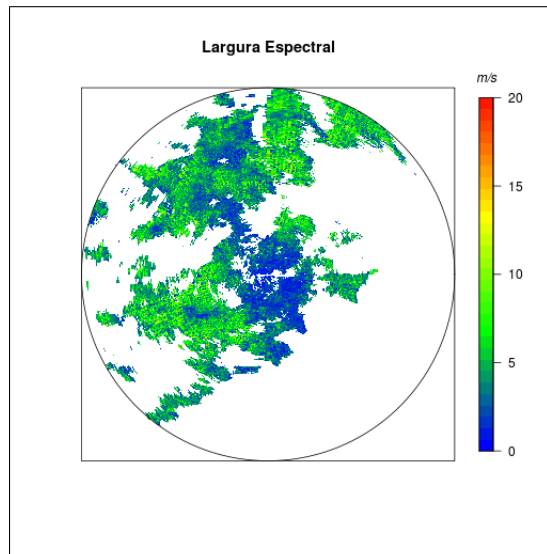


Fonte: A autora (2011)

Largura Espectral (W)

Uma outra variável medida pelo radar meteorológico é a *largura espectral* (W), que é o desvio padrão das medidas obtidas em cada bin e, meteorologicamente, indica a turbulência nesse volume conforme a variância dos alvos presentes. A variável é expressa em m/s e pode ser visualizada na Figura 2.8.

Figura 2.8: Imagem de radar para a variável largura espectral



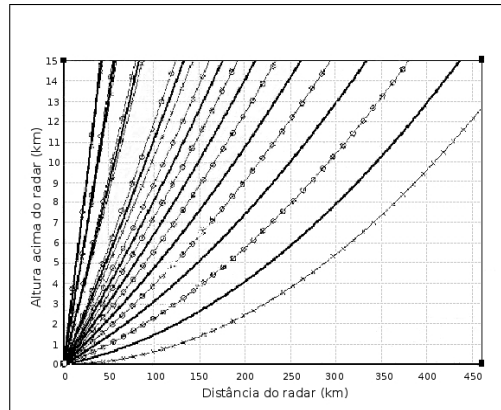
Fonte: A autora (2011)

2.2 Interpolação dos Dados

Devido às características físicas do radar, os dados por ele coletados encontram-se em um sistema de coordenadas esféricas (r, ϕ, θ) , onde r é a distância do alvo em relação ao radar, ϕ é o ângulo de azimute em relação ao Norte Geográfico e θ é o ângulo de elevação em relação ao horizonte.

Entretanto, em um sistema de coordenadas esféricas, os dados têm resolução espacial não-uniforme, com alta resolução na direção radial e baixa na azimutal. Como pode ser observado na Figura 2.9, próximo ao radar os centros dos *bins* estão mais próximos uns dos outros, enquanto em distâncias elevadas, estão mais espaçados. Devido a essa não-uniformidade, os dados são transformados para um sistema de coordenadas cartesianas, o que ainda facilita a visualização dos dados.

Figura 2.9: Espaçamento entre os bins



Fonte: (ZHANG et al., 2001)

Devido à refração da atmosfera, o feixe emitido não se propaga retilineamente, o que dificulta ainda mais as aproximações para o sistema cartesiano. Bean e Dutton (BEAN; DUTTON, 1966) propuseram uma solução para esse problema, com a aproximação do modelo de 4/3 Raio-efetivo-da-Terra. Supondo esse novo raio para o planeta, a curvatura do feixe (em condições normais de refração) desaparece.

Diversas técnicas de interpolação foram adaptadas ou desenvolvidas especificamente para a interpolação de dados de radar. A utilizada no presente trabalho foi a chamada Mapeamento por Vizinhos Próximos (MVP).

A técnica consiste em, primeiramente, criar uma grade cartesiana, que irá conter os dados interpolados. Esse esquema simplesmente preenche as células da grade cartesiana designando o valor do *bin* mais próximo a essa célula, sendo a distância avaliada usando-se a localização do centro do *bin* (ZHANG; HOWARD; GOURLEY, 2005).

Vale lembrar que quanto maior a distância do *bin* ao radar, maior é o número de pontos de grade “iluminados” pelo mesmo.

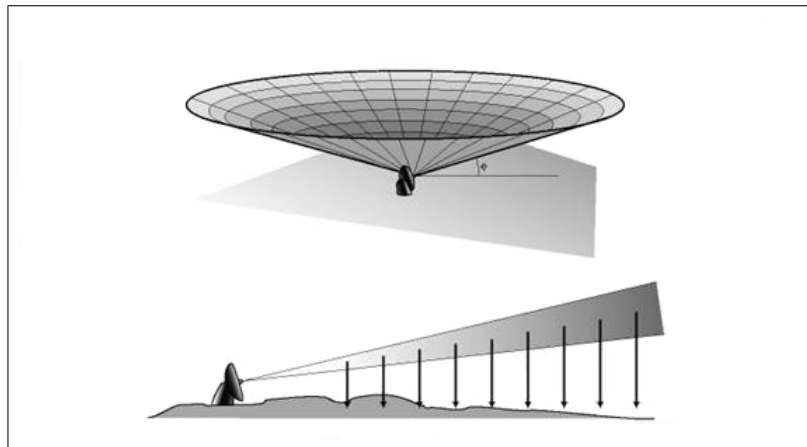
2.3 Formas de Visualização e Produtos Gerados

A seguir, são apresentadas as principais formas de visualização das variáveis coletadas pelo radar meteorológico, bem como alguns produtos gerados a partir das mesmas.

2.3.1 PPI - Plan Position Indicator

É a forma básica de visualização das variáveis medidas pelo radar. Normalmente, a variável é plotada em um plano onde o centro indica a posição do radar. O plano é fixo para um determinado ângulo de elevação. Portanto, deve-se observar que, conforme a distância em relação ao radar aumenta, a altura do feixe com relação ao solo também aumenta.

Figura 2.10: Produto PPI



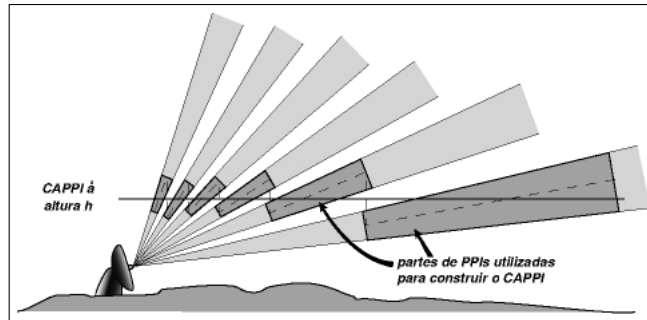
Fonte: (METEOPT, 2011)

2.3.2 CAPPI - Constant Altitude of Plan Position Indicator

O CAPPI é a visualização de uma variável a uma altura constante, ou seja, em um plano paralelo à superfície terrestre. Ao contrário do PPI, em uma imagem de CAPPI, a qualquer distância do radar os pontos estão na mesma altitude.

O cálculo do CAPPI é feito através da interpolação dos dados nos pontos superior e inferior mais próximo do ponto selecionado, como mostrado na Figura 2.11.

Figura 2.11: Produto CAPPI



Fonte: (METEOPT, 2011)

O valor para um ponto de grade (x, y) é interpolado pela seguinte equação:

$$CAPPI = x_b - \frac{x_a - x_b}{d_a - d_b} \cdot d_b, \quad (2.5)$$

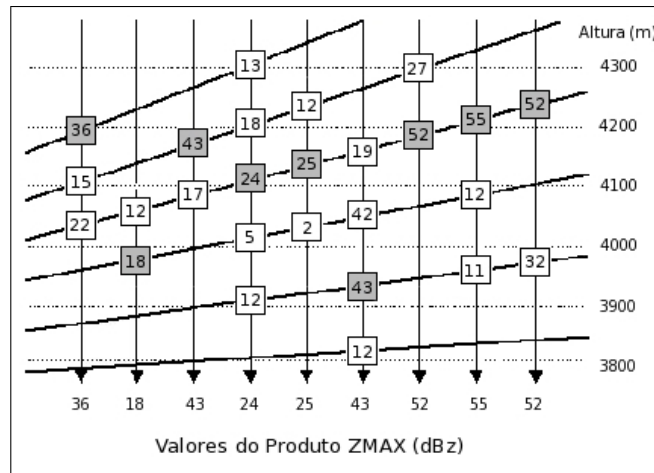
onde x_a e x_b são os valores nos pontos acima e abaixo do valor a ser interpolado; d_a e d_b são as respectivas distâncias.

Portanto, para gerar o CAPPI são necessárias ao menos duas varreduras com elevações diferentes. Para pontos distantes do radar, o intervalo entre os feixes é grande e o valor interpolado pode não representar adequadamente a variável no nível escolhido.

2.3.3 Zmax - Refletividade Máxima

Indica a máxima refletividade encontrada em uma coluna, como ilustrado na Figura 2.12.

Figura 2.12: Produto Zmax



Fonte: (MASSAMBANI, 2009)

2.3.4 VIL - Vertically Integrated Liquid-Water

O termo *VIL*, do inglês *Vertically Integrated Liquid-Water*, representa o conteúdo de água líquida integrado verticalmente. Assim, essa técnica mapeia as características tridimensionais das tempestades em um sistema bidimensional.

A integração é feita desde o ângulo elevacional mais baixo até o mais alto de toda a água líquida convertida a partir da refletividade Z pela relação $Z-R$ para cada ponto de grade.

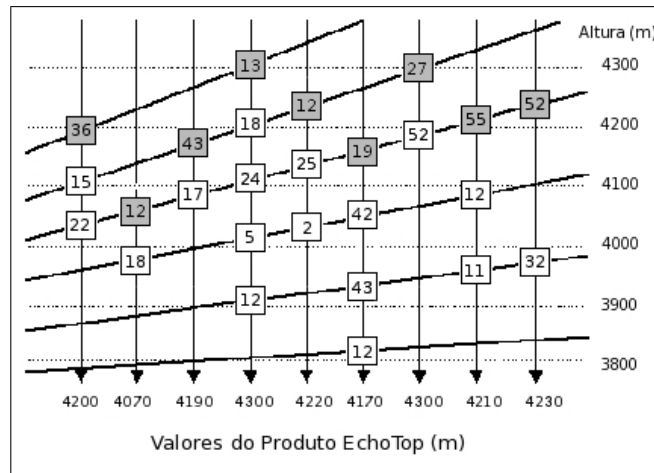
O VIL é definido por:

$$VIL = 3.44 \times 10^{-6} \int \left(\frac{z_i + z_{i+1}}{2} \right)^{4/7} \Delta h, \quad (2.6)$$

onde z_i e z_{i+1} são os valores de refletividade no limite inferior e superior, respectivamente, de uma camada de amostra e Δh é a diferença de altura entre os limites dessa camada. A variável é expressa em kg/m^2 .

2.3.5 Echo Top

É a altura máxima para cada ponto encontrado acima de um determinado valor de refletividade. O *Echo Top* indica o topo da nuvem, caracterizando assim, o estágio de desenvolvimento da célula.

Figura 2.13: Produto *Echo Top*

Fonte: (MASSAMBANI, 2009)

2.3.6 VIL Density

O *VIL Density* é simplesmente o *VIL* (Kg/m^2) dividido pelo *Echo Top* (m). O quociente é multiplicado por 1000 fornecer a unidade g/m^3 :

$$VILD = \frac{VIL}{EchoTop} \times 1000. \quad (2.7)$$

Quando o *VIL* é “normalizado” pelo *Echo Top*, o *VILD* resultante pode ser usado para identificar tempestades com alta refletividade em relação à sua altura. As tempestades intensas muitas vezes contêm núcleos de granizo. À medida que o valor do *VILD* aumenta, os núcleos de granizo tendem a ser mais profundos, intensos e maiores (AMBURN; WOLF, 1997).

2.4 Definições

Nesta seção, são descritas algumas variáveis utilizadas posteriormente no trabalho. Para tanto, é preciso apresentar, primeiramente, duas definições relevantes na área de Meteorologia: “isoterma de zero grau” e “banda brilhante”.

2.4.1 Isoterma de Zero Grau

A *isoterma de zero grau*, também chamada de *nível de congelamento*, representa a superfície na qual a temperatura do ar é constante e igual a $0^{\circ}C$. Essa temperatura varia de

acordo com o tempo, latitude, longitude e altura.

Esta superfície é a fronteira entre as temperaturas superiores e inferiores ao ponto de congelamento. Acima do nível de congelamento, a temperatura do ar está abaixo de zero. Abaixo, a temperatura está acima de zero.

2.4.2 Banda Brilhante

A *banda brilhante* é a representação, no radar, do “nível de derretimento”, uma estreita camada horizontal de elevada refletividade, próxima à isoterma de zero grau.

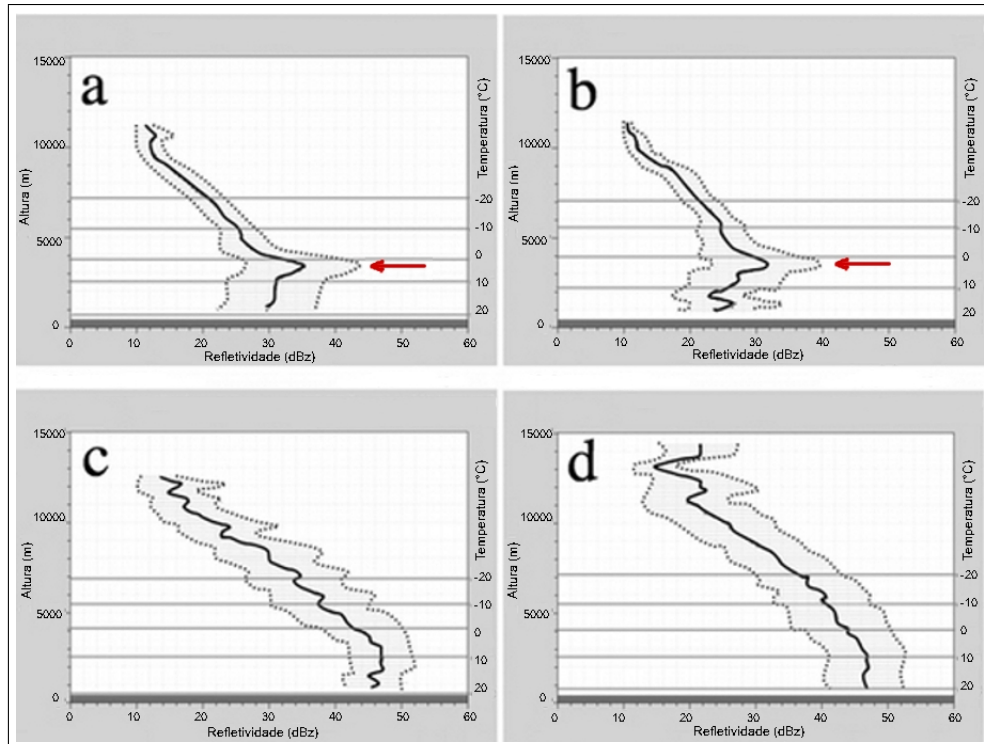
Quando uma camada de água no estado líquido se forma em uma partícula de gelo em processo de derretimento, sua refletividade pode aumentar em até $6.5dBz$. A refletividade diminui abaixo do nível de derretimento, porque quando o gelo converte-se em gotículas, o número de partículas por unidade de volume diminui.

A banda brilhante é observada em precipitações estratiformes. As fortes correntes de ar em precipitações convectivas não permitem a existência da estratificação horizontal essencial para a criação e sustentação de uma banda brilhante.

2.4.3 Perfil Vertical de Refletividade

O perfil vertical de refletividade (PVR) mostra o comportamento da variável refletividade com a variação de altura, para um determinado ponto de latitude e longitude.

Figura 2.14: Exemplos de PVR's



Fonte: (ZHANG; LANGSON; HOWARD, 2008)

Os gráficos 2.14(a)-(b) representam perfis verticais típicos de chuva estratiforme. Já as imagens em 2.14(c)-(d) mostram perfis verticais encontrados em chuva convectiva.

A parte indicada por flecha nos PVR's estratiformes representam presença de banda brilhante.

2.4.4 Alturas Relevantes

Representa a maior altura encontrada para um determinado valor de refletividade. Por exemplo, altura de $30dBz$ (h_{30}) e altura de $45dBz$ (h_{45}) representam a maior altitude encontrada para as refletividade de $30dBz$ e $45dBz$, respectivamente.

Se a altura é encontrada acima da isoterma de zero grau, significa que a tempestade em grandes altitudes é mais intensa, o que pode ser um indicativo de chuva convectiva.

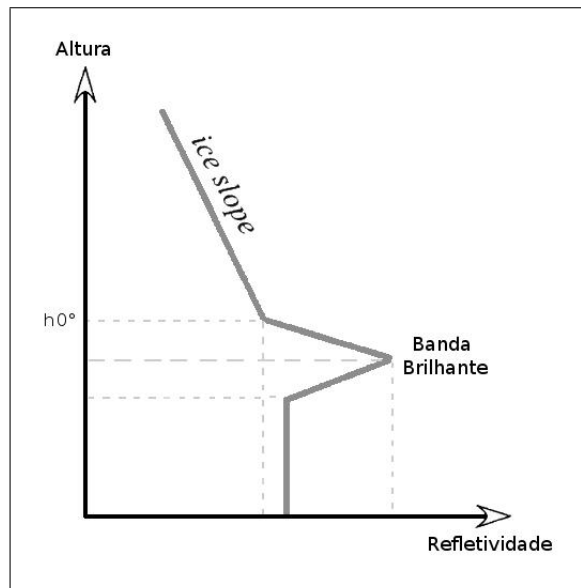
2.4.5 Z_{max} da Banda Brilhante

É a máxima refletividade encontrada na região de banda brilhante (isoterma de zero grau). Observa-se os dados em um intervalo de até 1km acima e 1km abaixo da altura da isoterma de zero grau, selecionando o maior valor de refletividade encontrado.

2.4.6 Ice Slope

É a variação dos valores de refletividade acima da isoterma de zero grau, medida em dBz/km .

Figura 2.15: Ilustração do *Ice Slope*



Fonte: (GOURLEY et al., 2009)

2.5 Relação Z-R

Da mesma forma que o radar não mede a velocidade radial diretamente, e sim a partir da diferença de fase entre o sinal emitido e refletido, ele também não é capaz de medir a precipitação (R). Esta é calculada a partir da intensidade do sinal que foi refletido pelos hidrometeoros (Z) presentes na atmosfera utilizando-se da relação Z-R, que se encarrega de transformar refletividade em precipitação. Uma relação empírica pode ser obtida plotando-se gráficos de taxa de precipitação que chega no solo contra a refletividade calculada pelo radar

(RINEHART, 1994). Sua forma mais conhecida e utilizada é

$$Z = aR^b, \quad (2.8)$$

onde Z é a refletividade em mm^6/m^3 , R é a taxa de precipitação em mm/h , e a e b são constantes empíricas, determinadas por meio da distribuição das partículas. Os valores dessas constantes variam de acordo com a região do globo e tipos de precipitações.

A relação mais frequentemente utilizada é a de Marshall e Palmer (MARSHALL; PALMER, 1948), que representa a chuva estratiforme:

$$Z = 200R^{1,6}. \quad (2.9)$$

Para a chuva convectiva (tempestades), Sekhon e Srivastava (SEKHON; SRIVASTAVA, 1971) encontraram, a partir de observações de radar, a seguinte relação:

$$Z = 300R^{1,35}. \quad (2.10)$$

Estes coeficientes são próximos aos encontrados para as chuvas convectivas em outras regiões da terra.

Para a neve, Sekhon e Srivastava (SEKHON; SRIVASTAVA, 1970) obtiveram os valores:

$$Z = 1780R^{2,21}. \quad (2.11)$$

2.6 Precipitações Convectivas e Estratiformes

Como pode ser observado na seção anterior, para cada tipo (ou classe) de precipitação, pode-se determinar uma relação Z-R. Este fato justifica a necessidade de um critério para separar e classificar os diferentes eventos meteorológicos que podem ocorrer na região onde o radar está instalado, em termos de precipitação convectiva ou estratiforme. Com a separação, crê-se que é possível alcançar resultados mais precisos na estimativa de precipitação com dados de radar, resultados esses que poderiam ser utilizados pela Defesa Civil no alerta de enchentes e em modelos hidrológicos, para citar algumas aplicações. Nesta seção, são apresentadas algumas das principais tentativas já feitas para a classificação de eventos de precipitação a partir de dados de radar e os resultados por elas alcançados.

Sistemas convectivos estão associados com fortes campos verticais de vento e altas taxas pluviométricas, enquanto nos sistemas estratiformes predomina extensa área coberta,

baixas velocidades verticais de vento, homogeneidade horizontal (gradiente suave) e taxas de precipitação menores (ANAGNOSTOU e KUMMEROW, 1997).

Steiner *et al.* (STEINER; HOUZE; YUTER, 1995) fizeram uma revisão das principais técnicas de separação desenvolvidas até então e uma nova proposta foi apresentada. Partindo da técnica elaborada por Churchill e Houze (CHURCHILL; HOUZE, 1984), que consistia em identificar um núcleo convectivo, a partir do qual todos os pontos dentro de uma distância fixa receberiam a mesma classificação, foi argumentado que um raio fixo como o proposto não era suficiente para uma correta identificação. A sugestão foi que um raio variável em função da média da intensidade do núcleo fosse adotado. Suas principais conclusões foram:

- a identificação de precipitação estratiforme através da banda brilhante é limitada; entretanto, pode ser usada para identificar classificação errônea de precipitação estratiforme como convectiva;
- técnicas baseadas no rastreamento do eco durante intervalos de tempo (coletas sucessivas de dados) são computacionalmente muito exigentes;
- critérios úteis na identificação de precipitação convectiva incluem intensidade da refletividade, valor mais elevado na região do núcleo e características da área no entorno.

No mesmo ano da apresentação desta técnica, DeMott *et al.* (DEMOTT; CIFFELI; RUTLEDGE, 1995) propuseram um novo aperfeiçoamento, dessa vez considerando não apenas o perfil horizontal, mas sim a estrutura tridimensional do campo analisado pelo radar. Isso foi feito aplicando-se a técnica de Steiner *et al.* (1995) para cada um dos níveis existentes no volume do radar e executando um pós-processamento nos resultados. Nessa etapa, as características físicas das precipitações foram utilizadas para garantir que não houvesse presença de precipitação convectiva nos níveis de derretimento e que, se sua ocorrência fosse detectada nos níveis médios, essa classificação fosse estendida até os níveis mais elevados. Segundo DeMott *et al.*, a utilização dos dados de vários níveis aumentou a acurácia da classificação.

Mais um aperfeiçoamento de Steiner *et al.* (1995) foi proposto, dessa vez por Biggerstaff e Listemaa (BIGGERSTAFF; LISTEMAA, 2000). Fazendo uma análise criteriosa do algoritmo de Steiner *et al.* quando aplicado em três situações (linha de instabilidade, extensas áreas estratiformes com centros convectivos e apenas precipitações convectivas espalhadas), concluiu-se que existem duas causas principais de erros: áreas de intensa precipitação estratiforme sendo classificadas como convectiva e bordas de convectivas sendo classificadas como estratiformes. Para corrigir o problema, um algoritmo que utiliza como entrada dados já classificados pela técnica de Steiner *et al.* foi desenvolvido. Além desses dados, o algoritmo

também necessita de um volume de radar interpolado em coordenadas cartesianas para que novos parâmetros possam ser calculados para cada ponto de grade.

A metodologia proposta por Steiner *et al.* (1995), que serviu de base para outras técnicas aperfeiçoadas, é descrita a seguir.

2.6.1 Método de Steiner *et al.*

Para utilização da técnica, os dados de refletividade coletados pelo radar são transformados em coordenadas cartesianas e, para cada ponto de grade, é feita a classificação em estratiforme ou convectiva, da seguinte maneira:

1. Só os dados de chuva acima de $15dBz$ são considerados;
2. Divide-se os dados em *clusters* e considera-se apenas os maiores que $20km^2$;
3. Os chamados centros convectivos são encontrados, através de dois testes:

Teste Absoluto: todo pixel com mais de $40dBz$ é convectivo;

Teste Relativo: calcula-se o z_{fundo} , que é a média de refletividade de todos os pontos (válidos) dentro de uma circunferência de raio de $11km$, centrada no ponto em estudo. Tal ponto será considerado um centro convectivo se a seguinte desigualdade for satisfeita:

$$z > z_{fundo} + \Delta z, \quad (2.12)$$

onde z é a refletividade no ponto e

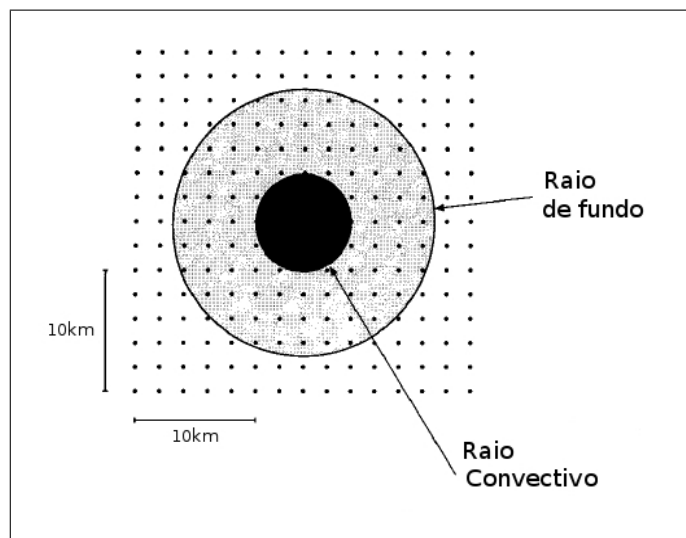
$$\Delta z = \begin{cases} 10 - \frac{z_{fundo}}{180}, & \text{se } 0 \leq z_{fundo} < 42.43 \\ 0, & \text{se } z_{fundo} \geq 42.43 ; \end{cases} \quad (2.13)$$

4. Para cada centro convectivo encontrado, pontos ao redor deste também são considerados chuva convectiva. São os pontos limitados por uma circunferência de raio definido de acordo com o z_{fundo} encontrado anteriormente, da seguinte forma:

$$R = \begin{cases} 1km, & \text{se } z_{fundo} \leq 25 \\ 2km, & \text{se } 25 < z_{fundo} \leq 30 \\ 3km, & \text{se } 30 < z_{fundo} \leq 35 \\ 4km, & \text{se } 35 < z_{fundo} \leq 40 \\ 5km, & \text{se } z_{fundo} > 40 . \end{cases} \quad (2.14)$$

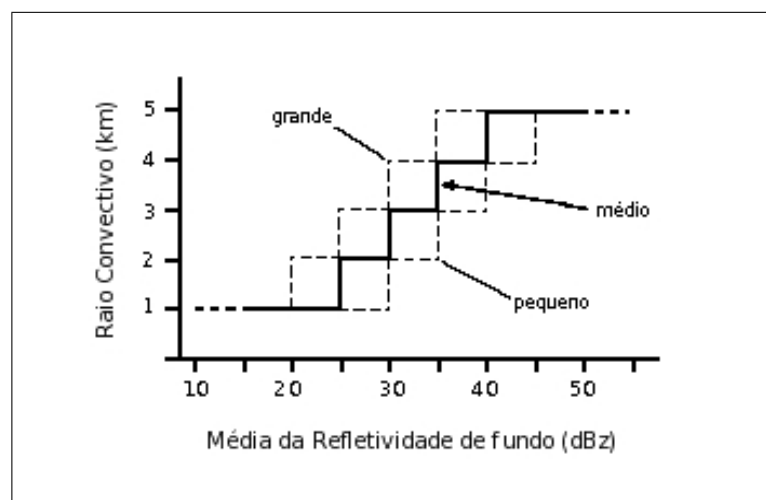
A Figura 2.16 esquematiza como pontos de chuva convectiva são identificados. O círculo levemente sombreado é a região de fundo para um dado ponto de grade (centro da imagem), dentro da qual uma média de refletividade é calculada. Neste caso, os pontos de grade estão espaçados 2km . O círculo mais escuro representa a região em torno do centro convectivo - se identificado como tal - incluída como zona convectiva. O raio do círculo de convecção é uma função da refletividade média do raio de fundo, mostrada na Figura 2.17. (STEINER; HOUZE; YUTER, 1995).

Figura 2.16: Ilustração do raio convectivo e raio de fundo para um determinado pixel



Fonte: (STEINER; HOUZE; YUTER, 1995)

Figura 2.17: Relação entre a média de refletividade ao redor do ponto e o raio de alcance convectivo



Fonte: (STEINER; HOUZE; YUTER, 1995)

Assim, através dos núcleos convectivos e raio de alcance da circunferência que classifica os pontos ao redor desses núcleos, todos os pontos de chuva convectiva são encontrados. Os demais pontos são identificados como estratiformes.

3 Support Vector Machine

O *Support Vector Machine* (SVM) é uma técnica de Aprendizado de Máquina que consiste em, a partir de dados previamente classificados, encontrar um hiperplano separador entre as classes, a partir do qual, dados desconhecidos até então podem ser classificados.

O método SVM requer a solução de um problema de otimização matemática. Sendo assim, a seguir são apresentados alguns teoremas e definições da Teoria da Otimização, necessários para fundamentar a formulação do método.

3.1 Conceitos da Teoria da Otimização

A teoria da otimização é o ramo da matemática que tem como objetivo determinar soluções para certa classe de problemas e desenvolver algoritmos eficazes que permitam encontrá-las (CRISTIANINI; SHAWE-TAYLOR, 2000).

Os conceitos apresentados a seguir são de fundamental importância para o entendimento do método SVM. Maiores detalhes podem ser encontrados em Cristianini e Shawe-Taylor (2000) e em Izmailov e Solodov (2005).

3.1.1 Problema Primal

Um problema de otimização é composto por uma função objetivo, que depende de variáveis sujeitas a determinadas restrições de igualdade ou desigualdade. Os possíveis valores das variáveis de decisão são delimitados pelas restrições, formando-se um conjunto de soluções factíveis do problema. Dependendo da natureza da função objetivo, tem-se um problema de maximização ou de minimização.

Definição 1 (Problema de Otimização Primal). *Um problema de otimização primal pode ser*

representado da seguinte forma:

$$\begin{aligned}
 & \text{Minimizar} && f(x), && x \in \Omega \subseteq \mathfrak{R}^n \\
 & \text{Sujeito a:} && g_i(x) \leq 0, && i = 1, \dots, k \\
 & && h_j(x) = 0, && j = 1, \dots, m
 \end{aligned} \tag{3.1}$$

onde

$f : \Omega \subseteq \mathfrak{R}^n \rightarrow \mathfrak{R}$ é a função objetivo;

$g_i : \Omega \subseteq \mathfrak{R}^n \rightarrow \mathfrak{R}, i = 1, \dots, k; \quad h_j : \Omega \subseteq \mathfrak{R}^n \rightarrow \mathfrak{R}, j = 1, \dots, m$ são as restrições.

É válido ressaltar que todo problema de maximização pode ser transformado em um problema de minimização equivalente, mudando-se o sinal da função objetivo. Assim, maximizar $f(x)$ é o mesmo que minimizar $-f(x)$.

A região factível do problema é denotada por qualquer x pertencente ao domínio do problema que satisfaça todas as restrições simultaneamente, ou seja:

$$F = \{x \in \Omega : g_i(x) \leq 0, i = 1, \dots, k; h_j(x) = 0, j = 1, \dots, m\}.$$

A solução do problema será o ponto $x^* \in F$ tal que $f(x^*) < f(x)$ para todo ponto $x \in F$. Este ponto é chamado de *mínimo global*. Se $\exists \varepsilon > 0$ tal que $f(x) \geq f(x^*) + \varepsilon, \forall x \in \Omega$ com $\|x - x^*\| < \varepsilon$, então o ponto x^* é chamado de *mínimo local*.

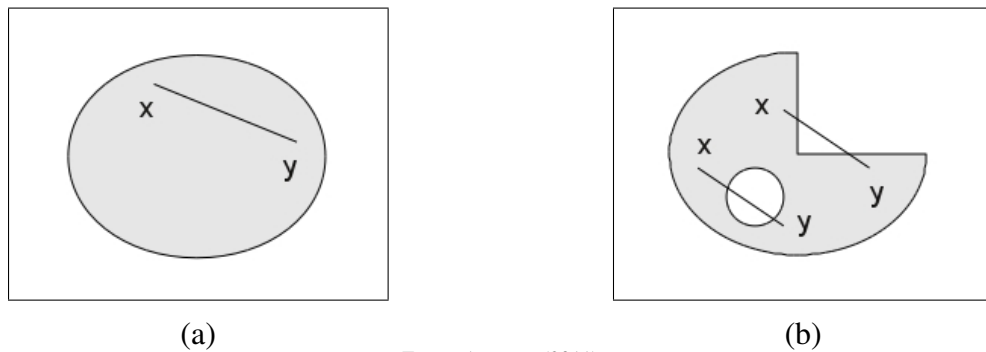
Quando a função objetivo é uma função quadrática e as restrições correspondentes são lineares ou quadráticas, tem-se um problema de otimização quadrática. Um caso específico deste problema ocorre quando a função objetivo e as restrições são funções convexas. Neste caso, tem-se um problema quadrático convexo.

As definições de função convexa e de conjunto convexo são mostradas a seguir.

Definição 2 (Conjunto Convexo). *Um conjunto $\Omega \in \mathfrak{R}^n$ é dito convexo quando o segmento de reta que une dois pontos quaisquer do conjunto está contido neste conjunto. Ou seja, para quaisquer $x \in \Omega, y \in \Omega$ e $\alpha \in [0, 1]$, a combinação convexa $(1 - \alpha)x + \alpha y \in \Omega$.*

A Figura 3.1 ilustra um conjunto convexo e um não-convexo.

Figura 3.1: Exemplo de (a) conjunto convexo e (b) conjunto não-convexo



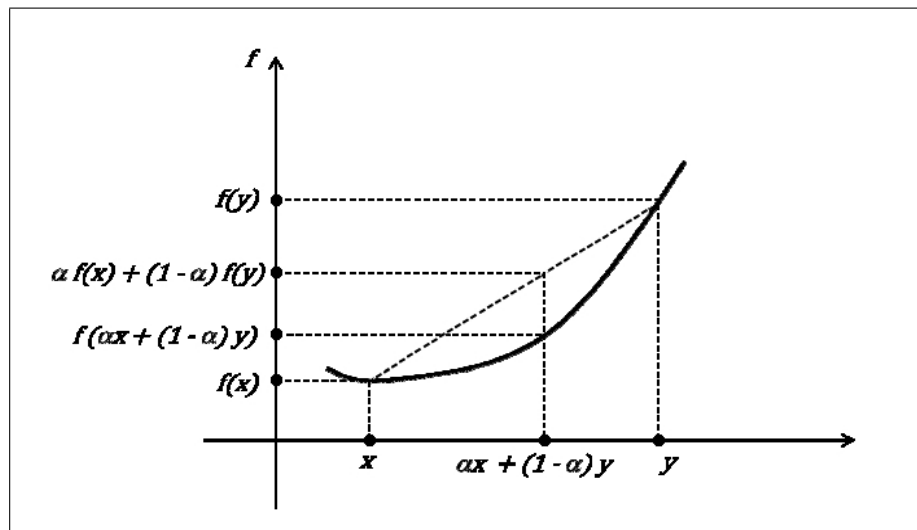
Fonte: A autora (2011)

Definição 3 (Função Convexa). *Uma função $f : \Omega \rightarrow \Re$ é considerada convexa se a região sobre o seu gráfico, ou seja, $\{(x,y) \in \Re^2 / y \geq f(x)\}$, for um conjunto convexo. Assim, para quaisquer $x \in \Omega$, $y \in \Omega$ e $\alpha \in [0, 1]$ é válida a equação:*

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

A Figura 3.2 ilustra a definição de função convexa, apresentada acima.

Figura 3.2: Função convexa



Fonte: (IZMAILOV; SOLODOV, 2005)

A noção de convexidade é muito importante na Teoria da Otimização, como pode ser observado pelo Teorema 1.

Teorema 1 (Teorema da Minimização Convexa). *Sejam $\Omega \in \mathfrak{R}^n$ um conjunto convexo e $f : \Omega \rightarrow \mathfrak{R}$ uma função convexa em Ω . Então todo minimizador local da função $f(x)$ em $x \in \Omega$ é minimizador global.*

O método SVM envolve a solução de um problema de minimização *quadrático convexo*. Sendo assim, pelo Teorema 1, tem-se a garantia da existência de um mínimo global, o que torna o método atrativo. Para resolver esse problema, utiliza-se a Teoria Lagrangeana.

3.1.2 Teoria Lagrangeana

O Teorema de Fermat foi desenvolvido em 1629, para problemas de minimização sem restrições. A Teoria Lagrangeana, desenvolvida por Lagrange em 1797, generalizou os resultados de Fermat para problemas com restrições de igualdade. Em 1951, Kuhn e Tucker estenderam o método de Lagrange para problemas com possíveis restrições de desigualdade, o que deu origem ao Teorema de Karush-Kuhn-Tucker. Esses teoremas são descritos a seguir.

Teorema 2 (Fermat). *Seja $f \in C^1$, onde C^1 é o conjunto das funções contínuas em Ω , a condição necessária para x^* ser um mínimo de $f(x)$ é*

$$\frac{\partial f(x^*)}{\partial x} = 0. \quad (3.2)$$

Essa condição, juntamente com a de convexidade de f , é também uma condição suficiente.

A Teoria Lagrangeana utiliza a chamada *função Lagrangeana*, que incorpora a função objetivo do problema com as restrições. É, portanto, uma combinação linear da função objetivo e cada uma das restrições de igualdade.

Definição 4 (Função Lagrangeana). *Seja um problema de otimização com função objetivo $f(x)$ e restrições de igualdade $h_j(x) = 0$, $j = 1, \dots, m$, define-se a função Lagrangeana como:*

$$L(x, \beta) = f(x) + \sum_{j=1}^m \beta_j h_j(x), \quad (3.3)$$

onde β_j é chamado multiplicador de Lagrange.

Teorema 3 (Lagrange). *A condição necessária para o ponto x^* ser um mínimo de $f(x)$, sujeito a $h_j(x) = 0$, $j = 1, \dots, m$ com $f, h_j \in C^1$, $j = 1, \dots, m$ é*

$$\frac{\partial L}{\partial x}(x^*, \beta^*) = 0; \quad (3.4)$$

$$\frac{\partial L}{\partial \beta}(x^*, \beta^*) = 0. \quad (3.5)$$

Se $L(x^*, \beta^*)$ é uma função convexa em x^* , então a condição acima é também suficiente.

Para problemas com restrições de igualdade e desigualdade, define-se a *função Lagrangeana generalizada*.

Definição 5 (Função Lagrangeana Generalizada). *Seja um problema de otimização com função objetivo $f(x)$ e restrições tanto de igualdade quanto de desigualdade, como o problema 3.1. Define-se a função Lagrangeana generalizada como:*

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^k \alpha_i g_i(x) + \sum_{j=1}^m \beta_j h_j(x), \quad (3.6)$$

onde α_i e β_j são os multiplicadores de Lagrange.

3.1.3 Dualidade

A teoria da dualidade baseia-se em associar ao problema original primal (Definição 1) um outro problema, chamado *dual*, que, sob certas condições, é equivalente ao primal e que pode ser mais fácil de se resolver (IZMAILOV; SOLODOV, 2005).

Definição 6 (Problema de Otimização Dual). *O problema dual Lagrangeano referente ao problema primal da Definição 1 é o seguinte:*

$$\begin{array}{ll} \text{Maximizar} & \theta(\alpha, \beta) \\ \text{Sujeito a:} & \alpha(x) \geq 0 \end{array} \quad (3.7)$$

onde $\theta(\alpha, \beta) = \inf_{x \in \Omega} L(x, \alpha, \beta)$.

Teorema 4 (Teorema Fraco da Dualidade). *Seja $x \in \Omega$ uma solução viável do problema primal (Definição 1), e seja (α, β) uma solução viável do problema dual (Definição 6). Então:*

$$f(x) \geq \theta(\alpha, \beta).$$

Este teorema afirma que o valor da solução dual é limitada superiormente pelo valor da solução primal.

Se $f(x^*) = \theta(\alpha^*, \beta^*)$, onde as restrições do problema primal e dual são satisfeitas, então x^* e (α^*, β^*) resolvem o problema primal e dual respectivamente.

A diferença entre os valores da função objetivo no problema primal e no dual é denominado de *gap de dualidade*. Assim, quanto mais próximo de zero está o valor do *gap de dualidade*, mais próximo se está da solução ótima do problema, já que esta ocorre quando o valor da função objetivo primal é igual ao da função objetivo dual.

Teorema 5 (Teorema Forte da Dualidade). *Dado o problema de otimização da Definição 1 com domínio convexo e $g_i, i = 1, \dots, k$, $h_j, j = 1, \dots, m$ sendo funções afins, ou seja, do tipo $h(x) = Ax - b$, o gap de dualidade é igual a zero.*

Teorema 6 (Condições de Karush-Kuhn-Tucker). *Dado o problema de otimização da Definição 1 com domínio convexo, $f \in C^1$ convexa, $g_i, i = 1, \dots, k$ e $h_j, j = 1, \dots, m$ sendo funções afins, a condição necessária e suficiente para o ponto x^* ser ótimo é a existência de α^*, β^* satisfazendo:*

$$\begin{aligned} \frac{\partial L}{\partial x}(x^*, \alpha^*, \beta^*) &= 0; \\ \frac{\partial L}{\partial \beta}(x^*, \alpha^*, \beta^*) &= 0; \\ \alpha_i^* g_i(x^*) &= 0, i = 1, \dots, k; \\ g_i(x^*) &\leq 0, i = 1, \dots, k; \\ \alpha_i^* &\geq 0, i = 1, \dots, k. \end{aligned}$$

O tratamento Lagrangeano para o problema de otimização convexo fornece a alternativa da descrição dual, que frequentemente tende a ser mais fácil de ser resolvida do que a descrição primal, a qual apresenta restrições de desigualdades mais complexas. Esta estratégia tornou-se padrão na teoria do *Support Vector Machine*, porque a representação dual permite trabalhar em um espaço de alta dimensão, devido ao número de parâmetros ajustados não depender do número de atributos sendo utilizados (dimensão dos dados de entrada). Posteriormente será visto que as amostras associadas aos multiplicadores de Lagrange maiores do que zero (restrições ativas) serão denominadas de *vetores suporte* (SEMOLINI, 2002).

3.2 Reconhecimento de Padrões

O Reconhecimento de Padrões, ou Classificação de Padrões, é um subtópico do Aprendizado de Máquina, o qual, por sua vez, é uma sub-área da Inteligência Artificial.

Um método de reconhecimento de padrões deve, baseado no conhecimento extraído dos exemplos de uma base, classificar um exemplo novo, desconhecido até então, ao padrão que mais reflete as suas características (CARVALHO, 2005).

3.2.1 Inteligência Artificial

Inúmeros são os problemas feitos de forma repetitiva e manualmente, através da ação humana. De forma a automatizar essas atividades, é possível realizar um estudo e posterior desenvolvimento de rotinas computacionais que auxiliem, ou até mesmo substituam, o trabalho humano em tais atividades (BERGAMINI; ARAUJO; MOTTER, 2006).

A Inteligência Artificial (IA) é uma área da Ciência da Computação responsável pelo estudo e desenvolvimento de tais rotinas computacionais, envolvendo pesquisas relacionadas à aprendizagem e evolução computacional, sistemas especialistas, sistemas de suporte à tomada de decisão, agentes computacionais inteligentes, entre outras áreas (BERGAMINI; ARAUJO; MOTTER, 2006).

Um dos ramos da IA é o Aprendizado de Máquina, técnica a ser abordada neste estudo.

3.2.2 Aprendizado de Máquina

As técnicas de Aprendizado de Máquina (AM) empregam um princípio de inferência denominado indução, no qual obtém-se conclusões genéricas a partir de um conjunto particular de exemplos (LORENA; CARVALHO, 2007). O aprendizado pode ser supervisionado ou não-supervisionado.

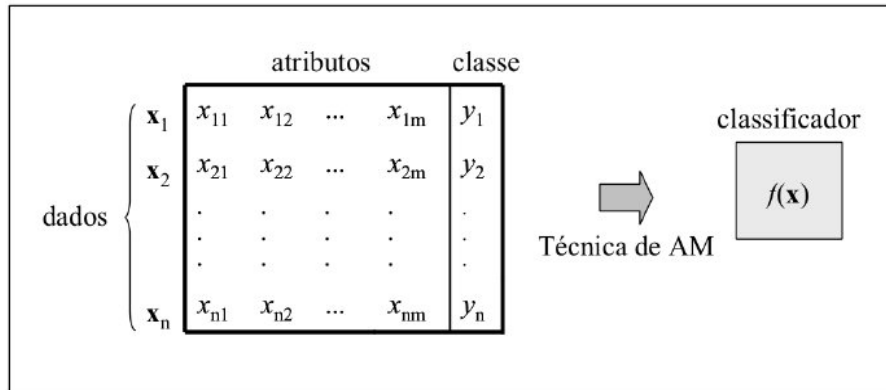
O aprendizado supervisionado é realizado por meio de um supervisor externo, que fornece ao sistema as entradas juntamente com os valores de saída desejados. A ideia é que, a partir da amostra apresentada, o sistema seja capaz de construir um classificador para rotular novos dados, desconhecidos até então.

No aprendizado não-supervisionado não existem exemplos já rotulados. O algoritmo de AM busca, a partir dos dados de entrada, criar alguma compreensão dos dados e gerar uma representação interna capaz de codificar as características de entrada em novas classes e agrupá-las corretamente (ALMEIDA, 2007).

A Figura 3.3 mostra um esquema do aprendizado supervisionado. Tem-se um conjunto com n dados; cada dado x_i , $i = 1, \dots, n$, é representado por um vetor com m características, ou seja, $x_i = (x_{i1}, \dots, x_{im})$. As variáveis y_i representam as classes. A partir dos exemplos e de

suas respectivas classes, o algoritmo de AM extrai um classificador (LORENA; CARVALHO, 2007).

Figura 3.3: Indução de classificador em aprendizado supervisionado



Fonte: (LORENA; CARVALHO, 2007)

A aprendizagem supervisionada é composta por duas fases. A primeira é a fase de *treinamento*, representada pela Figura 3.3, na qual exemplos são utilizados pelo sistema para aprendizagem e geração de um classificador. A segunda fase é a de *teste*, na qual novos exemplos são rotulados a partir do classificador existente.

O conjunto de treinamento deve ser estatisticamente representativo, para que a máquina consiga reconhecer os exemplos de teste, propriedade conhecida como *generalização* (CARVALHO, 2005).

Uma máquina de aprendizagem supervisionada deve ter a propriedade de, após a observação de vários pares de entrada e saída $\{x_i, y_i\}_{i=1}^n$, imitar o comportamento do sistema, gerando saídas próximas de y_i a partir de entradas próximas de x_i (VAPNIK, 1995).

Quando o número de padrões - saídas ou classes - é finito, assumindo valores discretos, a tarefa é denominada *classificação de padrões* (reconhecimento de padrões). Se houver apenas duas classes possíveis, dá-se o nome de *classificação binária*. Caso haja mais que dois padrões, tem-se uma *classificação multiclases*. Já quando existe um número infinito de padrões possíveis (valores reais), o problema é denominado de *regressão*.

Para classificar padrões existem vários métodos como: Redes Neurais Artificiais, Algoritmos Genéticos, Análise Multivariada, *Support Vector Machine* (SVM), entre outros.

O SVM, método utilizado no presente trabalho, é baseado na Teoria do Aprendizado Estatístico, que estabelece condições matemáticas para a escolha de um classificador.

3.3 Teoria do Aprendizado Estatístico

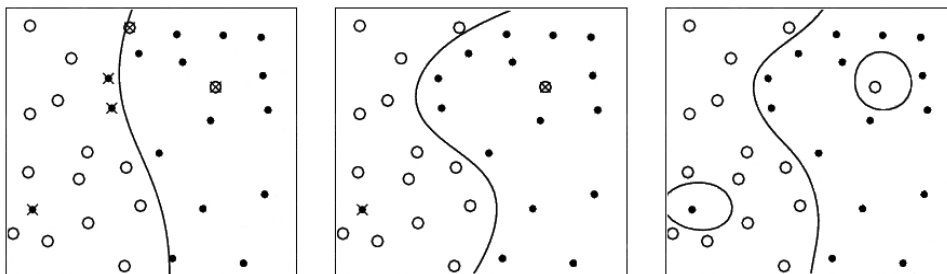
A Teoria do Aprendizado Estatístico (TAE), desenvolvida por Vapnik desde 1971 (VAPNIK; CHERVONENKIS, 1971), é também chamada de Teoria de Vapnik-Chervonenkis.

No contexto SVM, a Teoria do Aprendizado Estatístico objetiva controlar, em termos matemáticos, a habilidade de generalização, ou seja, a capacidade de classificação correta de padrões não treinados (SANTOS, 2002).

Sejam (x_i, y_i) pares de entrada e saída independentes e identicamente distribuídos, gerados de acordo com uma função de probabilidade $P(x, y)$, ou seja, cada amostra x_i é independente das demais amostras e segue $P(x, y)$. Dado um conjunto F de possíveis funções para escolha de um hiperplano separador, o problema do treinamento supervisionado consiste em encontrar uma função particular $f^* \in F$.

Considerando, por exemplo, o problema de classificação da Figura 3.4. O objetivo é encontrar uma função que separe os padrões da melhor forma possível.

Figura 3.4: (a) *Underfitting*; (b) Hiperplano de separação mais adequado; (c) *Overfitting*



Fonte: (SMOLA; SCHÖLKOPF, 2002)

A Figura 3.4(a) exemplifica o caso de uma função de complexidade inferior à necessidade do problema. Muitos dados do próprio conjunto de treinamento são classificados erroneamente. É o chamado *underfitting*. Já a função escolhida em 3.4(c) é de complexidade maior do que a necessária para a solução do problema. A função de classificação considera, inclusive, possíveis ruídos. Há, assim, um sobre-ajuste aos dados de treinamento, e a função acaba perdendo a capacidade de generalização. Esse problema é conhecido como *overfitting*. A Figura 3.4(b), enfim, apresenta uma função mais adequada ao problema. Tem uma complexidade intermediária e classifica corretamente grande parte dos dados.

A TAE estabelece condições matemáticas que auxiliam na escolha de um classificador particular f^* a partir de um conjunto de dados de treinamento. Essas condições levam em conta o desempenho do classificador no conjunto de treinamento e a sua complexidade, com o objetivo de obter um bom desempenho também para novos dados do mesmo domínio (LORENA; CARVALHO, 2007).

Para tal, é necessário definir uma função de risco que meça o erro médio ou o risco associado com o estimador e , então, encontrar o estimador com menor risco.

3.3.1 Minimização do Risco Empírico

Uma relação entre a saída desejada y_i com a previsão $f(x_i)$ pode ser feita através da *função custo*, ou *função erro*. Uma função custo comumente empregada em problemas de classificação binária (onde a saída é ± 1) é:

$$c(f(x), y) = \frac{1}{2} |f(x) - y|. \quad (3.8)$$

Essa função retorna 0 se a classificação é feita corretamente, e 1 caso contrário.

Uma escolha particular da função f resulta em uma máquina treinada. O *risco esperado* para o erro de teste é, então:

$$R(f) = \int c(f(x), y) dP(x, y). \quad (3.9)$$

Como a distribuição de probabilidade $P(x, y)$ é desconhecida, não se pode minimizar o risco diretamente. Normalmente, utiliza-se o princípio de indução para inferir uma função f^* que minimize o erro sobre esses dados e espera-se que esse procedimento leve também a um menor erro sobre os dados de teste (SMOLA; SCHÖLKOPF, 2002).

Assim, é utilizado o Princípio de *Minimização do Risco Empírico* que consiste em utilizar o conjunto de treinamento para construir uma aproximação do risco esperado. Essa aproximação é o chamado *risco empírico* e nada mais é do que o erro médio no conjunto de treinamento (BORGES, 2003):

$$R_{emp}(f) = \frac{1}{N} \sum_{i=1}^N c(f(x), y). \quad (3.10)$$

Este princípio baseia-se na intuição de que, pela lei dos grandes números, o risco

empírico converge para o risco esperado (DEFILIPPO, 2004), ou seja:

$$\lim_{N \rightarrow \infty} (|R(f) - R_{emp}(f)|) = 0. \quad (3.11)$$

Porém, minimizar o erro de treinamento não implica necessariamente em um pequeno erro de teste. Isso porque, ao escolher uma função f a partir de um conjunto de funções muito amplo, é sempre possível encontrar uma f com pequeno risco empírico. Entretanto, nesse caso os exemplos de treinamento podem se tornar pouco informativos para a tarefa de aprendizado, pois o classificador induzido pode se super-ajustar a eles (LORENA; CARVALHO, 2007).

Deve-se, portanto, restringir a classe de funções da qual f é extraída. A TAE lida com essa questão considerando a complexidade (ou capacidade) da classe de funções que o algoritmo é capaz de obter (SMOLA; SCHÖLKOPF, 2002). Para tal, é utilizado o conceito de *dimensão VC (dimensão Vapnik & Chervonenkis)*.

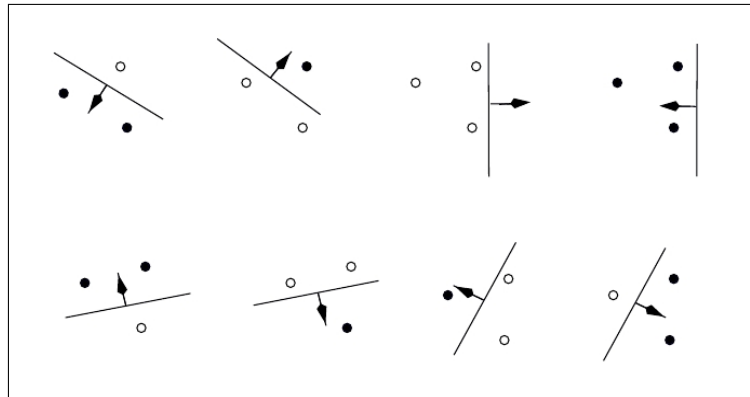
3.3.2 Dimensão VC

A *dimensão VC* é uma medida da capacidade de expressão de uma família de funções classificadoras obtidas por meio de um algoritmo de aprendizado. É dada pelo número máximo de amostras de treinamento que podem ser classificadas sem erro, usando-se funções da classe escolhida (LORENA; CARVALHO, 2007).

Definição 7 (Dimensão VC - Vapnik & Chervonenkis, 1971). *A dimensão VC de uma classe de funções é h se e somente se existe um conjunto de amostras $\{x_i\}_{i=1}^h$ tal que, para qualquer uma das 2^h configurações possíveis de rotulação binária, as amostras podem ser corretamente classificadas e não existe nenhum conjunto $\{x_i\}_{i=1}^q$ com $q > h$ satisfazendo esta propriedade.*

Na Figura 3.5, é apresentada uma forma de obtenção da dimensão VC para funções lineares no \mathcal{R}^2 , ou seja, retas. Conclui-se que a dimensão VC de retas no \mathcal{R}^2 é 3, pois 3 é o número máximo de amostras que podem ser corretamente classificadas por uma reta, para qualquer padrão de rotulação binária que as amostras podem admitir. Para 4 ou mais amostras, existem padrões de rotulação que não possibilitam uma classificação correta por intermédio de uma reta (LIMA, 2004).

Figura 3.5: Possibilidades de rotulação de três amostras no \mathfrak{R}^2 e a classificação realizada por uma função linear



Fonte: (BURGES, 1998)

Do Teorema 7, abaixo, tem-se um corolário que permite definir a dimensão de VC de uma família de hiperplanos em \mathfrak{R}^n .

Teorema 7 (Burges, 1998). *Considera-se um conjunto de m pontos em \mathfrak{R}^n . Escolhe-se qualquer um dos pontos como origem. Então os m pontos podem ser separados por hiperplanos orientados se e só se os restantes pontos são linearmente independentes.*

Corolário 1. *A dimensão VC de um conjunto de hiperplanos orientados em \mathfrak{R}^n é $n + 1$, uma vez que se pode sempre escolher $n + 1$ pontos, e escolher um dos pontos como origem, de forma que os restantes n pontos sejam linearmente independentes. No entanto é impossível escolher $n + 2$ pontos uma vez que é impossível ter-se $n + 1$ pontos linearmente independentes em \mathfrak{R}^n .*

3.3.3 Minimização do Risco Estrutural

Diversas técnicas de reconhecimento de padrões são baseadas na minimização do risco empírico, isto é, tenta-se otimizar o desempenho sobre o conjunto de treinamento. O método SVM minimiza o *risco estrutural*, ou seja, a probabilidade de classificar de forma errada padrões ainda não vistos por uma distribuição de probabilidade dos dados fixa e desconhecida (CRISTIANINI; SHAW-TAYLOR, 2000).

Uma relação entre o risco esperado e o risco empírico é dada por:

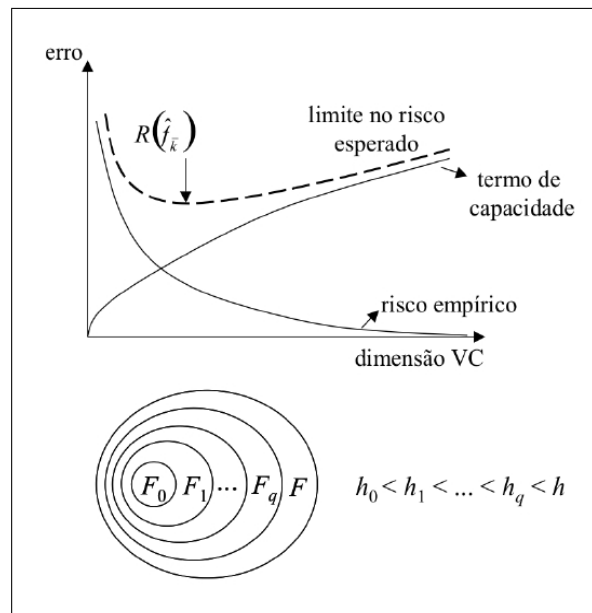
$$R(f) \leq R_{emp}(f) + \sqrt{\frac{h \ln\left(\frac{2N}{h} + 1\right) - \ln\left(\frac{\delta}{4}\right)}{N}}, \quad (3.12)$$

onde a parcela de raiz da soma é o termo de capacidade, h é a dimensão VC da classe de funções F à qual f pertence, N é a quantidade de exemplos no conjunto de treinamento e $\delta \in [0, 1]$.

O princípio de minimização do risco empírico visa otimizar o desempenho sobre o conjunto de treinamento. O princípio de minimização do risco estrutural agrega a essa minimização do risco esperado a minimização do termo de capacidade, ou seja, do erro na classificação de novos padrões, evitando assim o chamado *overfitting*.

Assim, pela Equação 3.12, tem-se que, para minimizar o risco esperado, é preciso associar a minimização do risco empírico com a minimização do termo de capacidade. Este último será pequeno quando a razão entre a dimensão VC e o número de amostras for pequeno.

Figura 3.6: Princípio de minimização do risco estrutural



Fonte: (SMOLA; SCHÖLKOPF, 2002)

A Figura 3.6 ilustra alguns conceitos da minimização do risco estrutural. Seja subconjuntos F_i de possíveis funções para a escolha de um classificador, de maneira que $F_0 \subset F_1 \subset \dots \subset F_q \subset F$. Como cada conjunto F_i é maior com o crescimento do índice i , a capacidade do conjunto de funções que ele representa também é maior, ou seja, $h_0 < h_1 < \dots < h_q < h$. Para um subconjunto particular F_k , seja $f_k \in F_k$ o classificador com menor risco empírico. A

medida que k cresce, o risco empírico de f_k diminui, uma vez que a complexidade do conjunto de classificadores é maior. Porém, o termo de capacidade aumenta com k . Como resultado, deve haver um valor ótimo k^* em que se obtém uma soma mínima do risco empírico e do termo de capacidade, minimizando assim o limite sobre o risco esperado. A escolha da função f_{k^*} constitui o *Princípio da Minimização do Risco Estrutural* (LORENA; CARVALHO, 2007).

3.4 O Hiperplano Ótimo

O objetivo do SVM é encontrar um hiperplano que separe as duas classes de dados da melhor forma possível, o chamado *hiperplano ótimo*.

3.4.1 Hiperplano Ótimo para Classes Linearmente Separáveis - Margens Rígidas

Para um problema de classificação binária, seja um conjunto de treinamento com n exemplos, representado por $\{(x_i, y_i)\}_{i=1}^n$, onde $x_i \in \mathfrak{R}^m$ é o vetor de características e $y_i \in \{-1, +1\}$ é a saída correspondente.

Para classificações binárias, uma função $f : X \subseteq \mathfrak{R}^m \rightarrow \mathfrak{R}$ é utilizada, de tal forma que a classificação seja representada pelo sinal de $f(x)$. Ou seja, se $f(x) \geq 0$ então a entrada pertence à classe positiva; caso contrário, pertence à classe negativa.

Considerando, inicialmente, conjuntos linearmente separáveis, a função $f(x)$ é dada por:

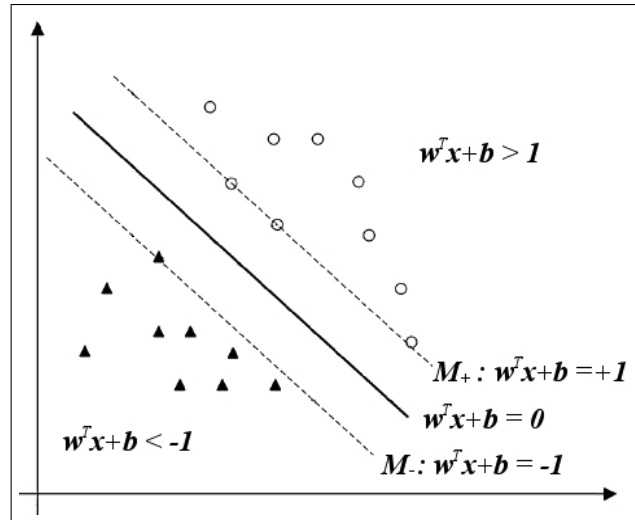
$$f(x) = w^T x + b, \quad (3.13)$$

onde $w \in \mathfrak{R}^m$ é o vetor pesos e $b \in \mathfrak{R}$ é o chamado *bias*.

O hiperplano definido pela equação $w^T x + b = 0$ divide o espaço em duas partes, correspondentes a dados de duas classes distintas. O vetor w define uma direção perpendicular ao hiperplano, enquanto a variação do valor de b (bias) move o hiperplano paralelamente a ele mesmo.

A classificação de cada padrão x do conjunto de treinamento é dada conforme a proximidade em relação às margens do hiperplano separador. Ou seja, é classificado como pertencente à classe -1 se estiver mais próximo da margem negativa $w^T x + b = -1$, e é pertencente à classe $+1$ se estiver mais próximo da margem positiva $w^T x + b = +1$ (ALES, 2008).

Figura 3.7: Classificação dos dados a partir do hiperplano separador



Fonte: A autora (2011)

A Figura 3.7 ilustra um conjunto de dados dividido por um hiperplano separador $w^T x + b = 0$, representado pela linha contínua. As amostras mais próximas da margem positiva $M_+ : w^T x + b = +1$ são classificadas como pertencentes à classe +1 (círculos); enquanto as amostras representadas por triângulos pertencem à classe -1, pois estão mais próximas da margem negativa $M_- : w^T x + b = -1$.

Um padrão é classificado corretamente se estiver fora da margem de separação de sua classe, ou seja:

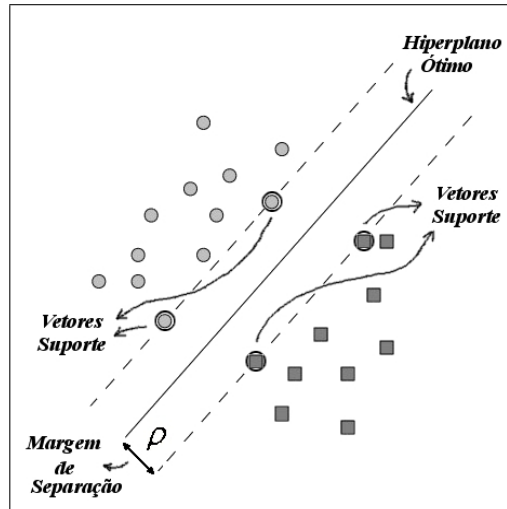
$$\begin{aligned} w^T x_i + b &\geq 1 && \text{se } y_i = +1; \\ w^T x_i + b &\leq -1 && \text{se } y_i = -1. \end{aligned} \quad (3.14)$$

De forma simplificada, temos:

$$y_i(w^T x_i + b) \geq 1, \quad \forall i = 1, \dots, n. \quad (3.15)$$

Assim, no caso representado na Figura 3.7, todas as amostras são classificadas corretamente, pois não há nenhum padrão entre as duas margens de separação.

Figura 3.8: Hiperplano com máxima margem de separação

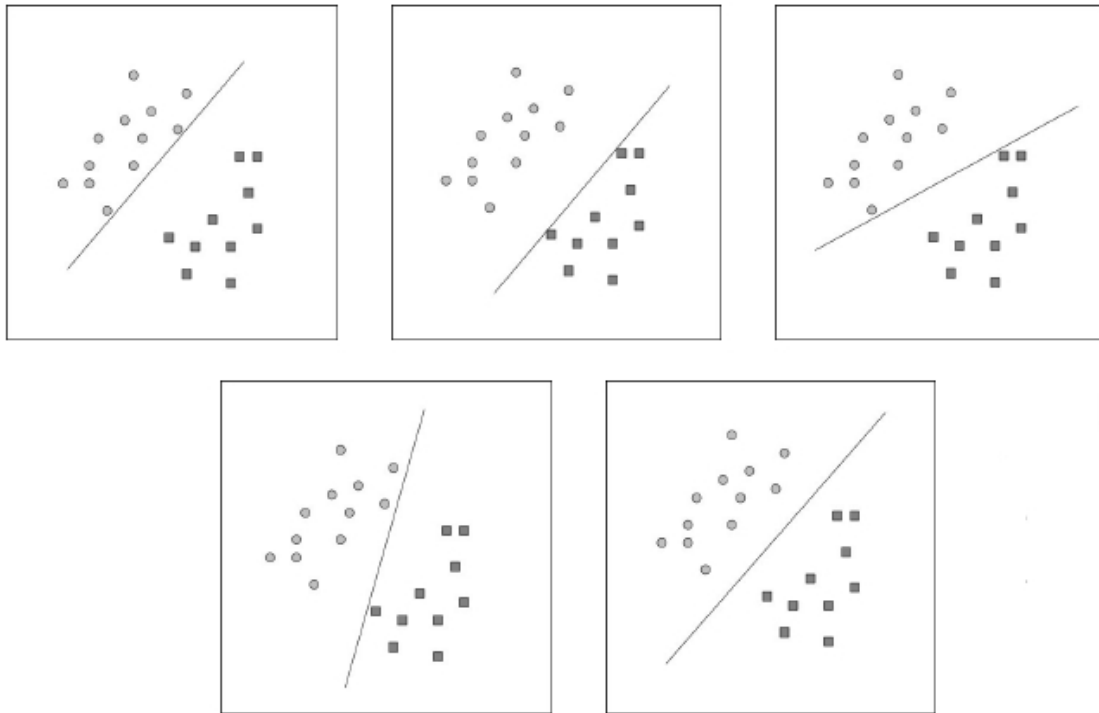


Fonte: (SEMOLINI, 2002)

A Figura 3.8 ilustra alguns conceitos definidos para a técnica SVM. A distância entre o hiperplano de separação $f(x)$ e o padrão mais próximo a este é denominada *margem de separação*, denominada por ρ . Sempre que for possível obter um $\rho > 0$, existirão infinitos hiperplanos (ilustrado na Figura 3.9), dentre os quais busca-se um hiperplano em particular, em que a margem de separação ρ é maximizada. De acordo com esta condição, a superfície de decisão é dita ser o *hiperplano ótimo* e a técnica de aprendizado de máquina utilizada para a obtenção deste hiperplano é denominada *Support Vector Machine* (SEMOLINI, 2002).

Os pontos que satisfazem a igualdade na Equação 3.15 encontram-se nas margens do hiperplano separador ($w^T x + b = +1$ e $w^T x + b = -1$), e são denominados *vetores suporte*. Como são os pontos mais próximos da superfície separadora, são os mais difíceis de serem classificados.

Figura 3.9: Possibilidades de separação das classes por diferentes hiperplanos



Fonte: A autora (2011)

A Figura 3.10 ilustra dois conjuntos de padrões separados pelo hiperplano ótimo. As margens deste hiperplano são $H_1 : w^T x + b = +1$ e $H_2 : w^T x + b = -1$. O vetor w é perpendicular às margens. Sendo $x_1 \in H_1$ e $x_2 \in H_2$, projetando $x_1 - x_2$ em w , pode-se obter a distância entre as margens. Essa projeção é dada por:

$$(x_1 - x_2) \left(\frac{w}{\|w\|} \frac{(x_1 - x_2)}{\|x_1 - x_2\|} \right). \quad (3.16)$$

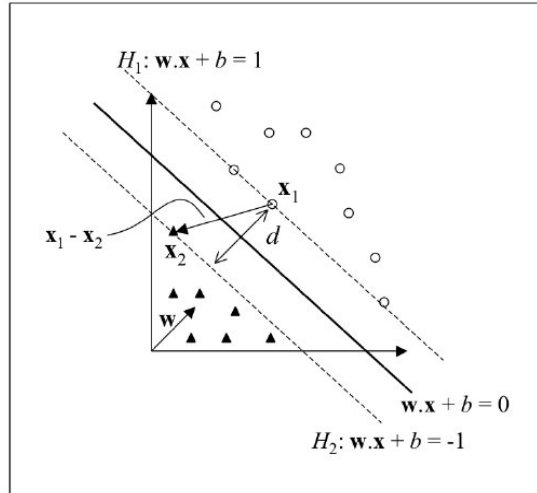
Subtraindo $w x_1 + b = +1$ de $w x_2 + b = -1$, encontra-se $w(x_1 - x_2) = 2$. Substituindo na Equação 3.16, tem-se:

$$\frac{2(x_1 - x_2)}{\|w\| \|x_1 - x_2\|}. \quad (3.17)$$

Tomando-se a norma da equação 3.17, tem-se o comprimento do vetor projetado, ou seja, a distância d entre as margens de separação:

$$d = \frac{2}{\|w\|}. \quad (3.18)$$

Figura 3.10: Cálculo da distância d entre os hiperplanos H_1 e H_2



Fonte: (LORENA; CARVALHO, 2007)

Assim, para encontrar o hiperplano ótimo, deve-se maximizar a distância d , o que é equivalente a minimizar $\|w\|$. Desse modo, formula-se o problema de otimização primal como:

$$\begin{aligned} \text{Minimizar} \quad & \frac{1}{2} \|w\|^2 \\ \text{Sujeito a:} \quad & y_i(w^T x_i + b) \geq 1, \quad \forall i = 1, \dots, n \end{aligned} \quad (3.19)$$

onde $w \in \mathfrak{R}^m$ e $b \in \mathfrak{R}$ são as incógnitas do problema.

A função Lagrangeana referente ao problema 3.19 apresentado acima é a que segue:

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i [y_i (w^T x_i + b) - 1], \quad (3.20)$$

onde $\alpha_i \geq 0$ são os multiplicadores de Lagrange associados a i -ésima restrição de desigualdade do primal.

A solução do problema de otimização em questão é determinada minimizando-se a função 3.20 em relação às variáveis primais e maximizando-a em relação aos multiplicadores de Lagrange. Para minimizar 3.20 em relação às variáveis primais, calculam-se as derivadas parciais primeiras dessa função em relação à w e b , e em seguida igualam-nas a zero (BELTRAMI, 2009).

$$\frac{\partial L}{\partial w}(w, b, \alpha) = w - \sum_{i=1}^n y_i \alpha_i x_i = 0; \quad (3.21)$$

$$\frac{\partial L}{\partial b}(w, b, \alpha) = \sum_{i=1}^n y_i \alpha_i = 0.$$

Substituindo-se as equações obtidas

$$w = \sum_{i=1}^n y_i \alpha_i x_i \quad (3.22)$$

e

$$\sum_{i=1}^n y_i \alpha_i = 0 \quad (3.23)$$

na função Lagrangeana 3.20, obtém-se:

$$\begin{aligned} L(w, b, \alpha) &= \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j (x_i^T x_j) - \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j (x_i^T x_j) + \sum_{i=1}^n \alpha_i \\ &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j (x_i^T x_j). \end{aligned} \quad (3.24)$$

Assim, formula-se o problema de otimização dual correspondente ao problema 3.19 como:

$$\begin{aligned} \text{Maximizar} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j (x_i^T x_j) \\ \text{Sujeito a:} \quad & \sum_{i=1}^n y_i \alpha_i = 0 \\ & \alpha_i \geq 0 \quad \forall i = 1, \dots, n \end{aligned} \quad (3.25)$$

onde α_i são os multiplicadores de Lagrange.

3.4.2 Hiperplano Ótimo para Classes Não-Linearmente Separáveis - Margens Flexíveis

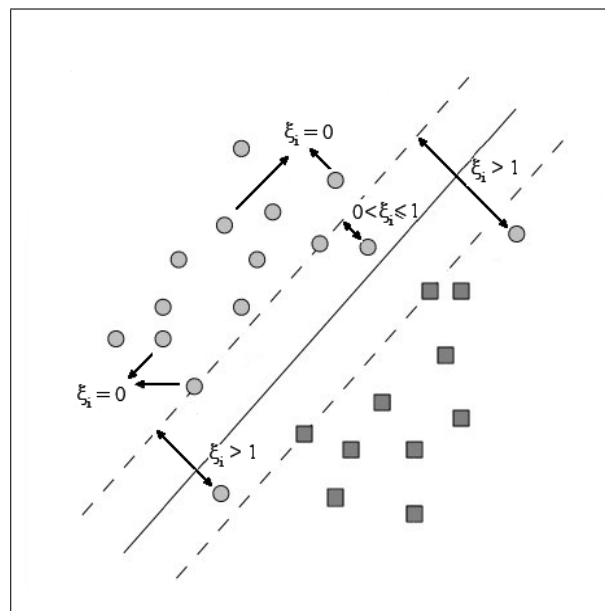
Para o caso de classes não linearmente separáveis, não é possível construir um hiperplano separador sem encontrar erros de classificação. Todavia, é possível encontrar um hiperplano que minimiza a probabilidade de erro de classificação junto às amostras de treinamento (SEMOLINI, 2002).

Para tal, são utilizadas *variáveis de folga* $\xi_i \geq 0$ associadas a cada vetor de treinamento x_i . Assim, a restrição 3.15 fica da forma:

$$y_i(w^T x_i + b) \geq 1 + \xi_i, \quad \forall i = 1, \dots, n. \quad (3.26)$$

Pontos classificados fora das margens de separação têm $\xi = 0$. Quando um ponto está localizado entre a margem de separação de sua classe e o hiperplano separador, ou seja, ainda do lado correto, tem-se que $0 < \xi_i \leq 1$. Caso o ponto esteja localizado no lado incorreto do hiperplano separador, tem-se que $\xi_i > 1$.

Figura 3.11: Possibilidades de valores para a variável ξ



Fonte: (SEMOLINI, 2002)

Assim, para encontrar o hiperplano ótimo com margens flexíveis formula-se o seguinte

problema de otimização primal:

$$\begin{aligned}
 \text{Minimizar} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\
 \text{Sujeito a:} \quad & y_i(w^T x_i + b) \geq 1 + \xi_i \\
 & \xi_i \geq 0 \quad \forall i = 1, \dots, n
 \end{aligned} \tag{3.27}$$

onde $w \in \mathfrak{R}^n$, $b \in \mathfrak{R}$ são as incógnitas do problema e C é chamada *constante de regularização*.

O primeiro termo da função objetivo ($\frac{1}{2} \|w\|^2$) visa maximizar a margem, enquanto o segundo termo ($C \sum_{i=1}^n \xi_i$) objetiva minimizar o valor das variáveis de folga ξ . Assim, o parâmetro C determina maior ou menor importância às variáveis de folga, possibilitando que o modelo do SVM seja menos sensível à presença de pontos “mal comportados” no conjunto de treinamento (ALES, 2008).

De maneira semelhante à descrita anteriormente pode-se transformar o problema primal 3.27 em um problema dual correspondente:

$$\begin{aligned}
 \text{Maximizar} \quad & \sum_{i=1}^n \alpha_i - \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j (x_i^T x_j) \\
 \text{Sujeito a:} \quad & \sum_{i=1}^n y_i \alpha_i = 0 \\
 & 0 \leq \alpha_i \leq C \quad \forall i = 1, \dots, n
 \end{aligned} \tag{3.28}$$

onde α_i são os multiplicadores de Lagrange e C é o parâmetro que limita o valor dos multiplicadores de Lagrange no treinamento SVM.

De acordo com as condições de KKT (Teorema 6), as soluções do problema 3.25 devem satisfazer a seguinte relação:

$$\alpha_i^* [y_i(w^{*T} x_i) - 1 + \xi_i] = 0 \quad \forall i = 1, \dots, n. \tag{3.29}$$

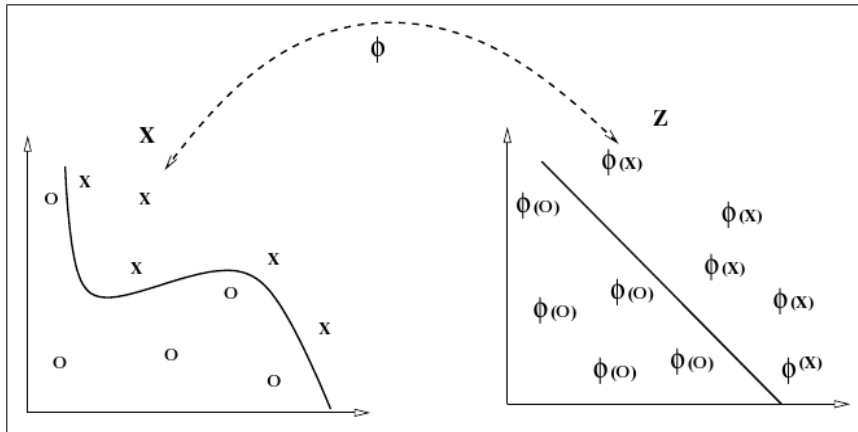
Isso implica que os α_i^* não zeros estão na margem funcional, chamados de *vetores suporte*, e os demais são nulos (CRISTIANINI; SHAWE-TAYLOR, 2000).

3.5 Função Kernel

O SVM implementa a seguinte ideia: mapeia os vetores de entrada x em um espaço de características de alta dimensão Z através de um mapeamento não-linear. Neste espaço, um

hiperplano ótimo separador é construído (VAPNIK, 1995), como pode ser observado na Figura 3.12.

Figura 3.12: Mapeamento do espaço de entrada via função *kernel*



Fonte: (SANTOS, 2002)

Na prática, este mapeamento é feito para um espaço de dimensão muito maior do que a original (SEMOLINI, 2002).

Embora a dimensão do espaço aumente em Z , a complexidade diminui, porque a classificação, que no espaço de entrada só era possível utilizando superfícies de decisão não lineares, no espaço de características, pode ser feita apenas com um simples hiperplano (superfície de decisão linear) (SANTOS, 2002).

Uma estratégia de pré-processamento em algoritmos de aprendizado envolve a mudança de representação dos dados na forma:

$$x = (x_1, \dots, x_m) \mapsto \phi(x) = (\phi_1(x), \dots, \phi_M(x)), \quad \text{onde } M \gg m.$$

Este passo é equivalente ao mapeamento não-linear dos dados do espaço de entrada X em um novo espaço $Z = \{\phi(x) | x \in X\}$, chamado de *espaço característico*. As medidas originais de representação dos dados são chamadas de *atributos* e as medidas no espaço Z são chamadas de *características* (SEMOLINI, 2002).

Este mapeamento é realizado através da chamada *função kernel*.

Definição 8 (Função Kernel). *Uma função kernel K é tal que, para todo $x_i, x_j \in X$ tem-se:*

$$K(x_i, x_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle, \quad (3.30)$$

onde $\phi : X \mapsto Z$ é uma função não linear do espaço original X no espaço final Z onde está definido um produto interno.

Para que uma função K seja denominada *kernel*, esta deve satisfazer as seguintes propriedades:

1. Ser simétrica:

$$K(x_i, x_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle = \langle \phi(x_j) \cdot \phi(x_i) \rangle = K(x_j, x_i);$$

2. Satisfazer a desigualdade de Cauchy-Schwarz:

$$\begin{aligned} K(x_i, x_j) &= \langle \phi(x_i) \cdot \phi(x_j) \rangle^2 \leq \|\phi(x_i)\|^2 \|\phi(x_j)\|^2 \\ &= \langle \phi(x_i) \cdot \phi(x_i) \rangle \langle \phi(x_j) \cdot \phi(x_j) \rangle = K(x_i, x_i) K(x_j, x_j). \end{aligned}$$

Na função objetivo do problema dual no modelo SVM (3.25), dada por

$$\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j),$$

o mapeamento é realizado aplicando-se ϕ aos dados presentes:

$$\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (\phi(x_i) \cdot \phi(x_j)), \quad (3.31)$$

ou seja,

$$\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j). \quad (3.32)$$

Assim, pela equação 3.32, pode-se observar que a função *kernel* pode ser utilizada para construir o hiperplano ótimo no espaço característico sem ter que considerar este espaço de forma explícita.

Teorema 8 (Mercer). *Seja $K(x, x')$ uma função kernel contínua e simétrica definida na região fechada $[a, b] \times [a, b]$. A função kernel $K(x, x')$ pode ser expandida em uma série:*

$$K(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \cdot \phi_i(x')$$

com coeficientes $\lambda_i > 0$. Para esta expansão ser válida e convergir absolutamente e uniforme-

mente, é uma condição necessária e suficiente que:

$$\int_a^b \int_a^b K(x, x') \psi(x) \psi(x') dx dx' \geq 0,$$

para todo $\psi(\cdot)$ no qual

$$\int_a^b \psi^2(x) dx < \infty.$$

A estrutura de produto interno no espaço de Hilbert¹ que conduz à construção do SVM implica que, para qualquer função *kernel* satisfazendo a condição de Mercer, existe um espaço característico onde as funções geram um produto interno (VAPNIK, 1998).

De forma simplificada, um *kernel* que satisfaz as condições de Mercer é caracterizado por dar origem a matrizes positivas semi-definidas K , em que cada elemento K_{ij} é definido por $K_{ij} = K(x_i, x_j)$, para todo $i, j = 1, \dots, n$ (HERBRICH, 2002).

O uso de diferentes funções *kernel* possibilita a construção de máquinas de aprendizagem com diferentes tipos de superfícies de decisão não-linear no espaço de entrada (SANTOS, 2002).

Alguns exemplos de função *kernel* são listados na Tabela 5.2:

Tabela 3.1: Funções *kernel*

Kernel	Função $K(x_i, x_j)$
Linear	$x_i \cdot x_j$
Polinomial	$\gamma((x_i \cdot x_j) + c_0)^d$
Gaussiano	$\exp(-\gamma \ x_i - x_j\ ^2)$
Sigmoidal	$\tanh(\delta(x_i \cdot x_j) + k)$

Fonte: (MEYER, 2011)

¹Espaço de Hilbert é um espaço vetorial de dimensão finita ou infinita.

4 K-means

O *K-means* é uma técnica de Análise de Agrupamentos, que permite classificar um conjunto de dados em um determinado número de grupos, baseado na similaridade ou dissimilaridade entre os dados. A Análise de Agrupamentos, também chamada de análise de *clusters*, é uma sub-área da Análise Multivariada.

4.1 Análise Multivariada

A análise multivariada é baseada em técnicas estatísticas que analisam simultaneamente múltiplas medidas de objetos sob análise.

As técnicas de análise multivariadas podem ser de *dependência* ou de *interdependência*. Uma técnica de dependência pode ser definida como uma na qual uma variável é identificada como a variável dependente a ser prevista ou explicada por outras variáveis conhecidas como independentes. Um exemplo de técnica de dependência é a análise de regressão múltipla. Uma técnica de interdependência é aquela em que nenhuma variável ou grupo de variáveis é definida como sendo dependente ou independente. Ao invés disso, o procedimento envolve a análise simultânea de todas as variáveis do conjunto. Análise de agrupamentos é um exemplo de técnica de interdependência (HAIR et al., 2009).

4.2 Análise de Agrupamentos

A análise de agrupamentos tem como objetivo agrupar objetos de acordo com suas características. A técnica reúne os dados em grupos, de forma que objetos de um mesmo grupo possuem maior similaridade uns com os outros do que com objetos de outros grupos. Em outras palavras, busca-se obter uma grande homogeneidade interna (entre elementos do mesmo grupo) e grande heterogeneidade externa (entre os grupos).

A aplicação desta técnica se baseia em medidas de semelhança ou de distância entre

objetos e na escolha de critérios de agregação (FERREIRA, 2000).

Dado um conjunto de N elementos $X = \{X_1, X_2, \dots, X_N\}$, o problema de agrupamento consiste na obtenção de um conjunto de k grupos, ou *clusters*, $C = \{C_1, C_2, \dots, C_k\}$, tal que os elementos contidos no grupo C_i possuam uma maior similaridade entre si do que com os elementos de qualquer um dos demais grupos do conjunto C . O conjunto C é considerado um agrupamento com k *clusters* caso as seguintes condições sejam satisfeitas (RODRIGUES, 2009):

$$\bigcup_{i=1}^k C_i = X; \quad (4.1)$$

$$C_i \neq \emptyset, \quad \text{para } 1 \leq i \leq k; \quad (4.2)$$

$$C_i \cap C_j = \emptyset, \quad \text{para } 1 \leq i, j \leq k \text{ e } i \neq j. \quad (4.3)$$

Segundo Hair *et al.* (2009), o objetivo principal da análise de agrupamentos é definir a estrutura dos dados colocando as observações mais parecidas em grupos. Para conseguir isso, deve-se tratar de três questões básicas:

1. Medição de similaridade.

É necessário desenvolver alguma medida de similaridade entre os objetos a serem usados ao longo do processo de agrupamento. Similaridade representa o grau de correspondência entre os objetos ao longo de todas as características usadas na análise.

2. Formação de agrupamentos.

Com medidas de similaridade já calculadas, o próximo passo é a formação dos agrupamentos com base na medida de similaridade de cada par de observações.

3. Determinação do número de agrupamentos na solução final.

É necessário definir a quantidade de grupos na qual os dados serão divididos, de forma que se obtenha as maiores homogeneidade interna e heterogeneidade externa possíveis.

A análise de agrupamentos é usada para descobrir estruturas em dados sem fornecer uma explicação ou interpretação. Em outras palavras, a técnica simplesmente descobre as estruturas de dados sem explicar por que eles existem (SILVA, 2010).

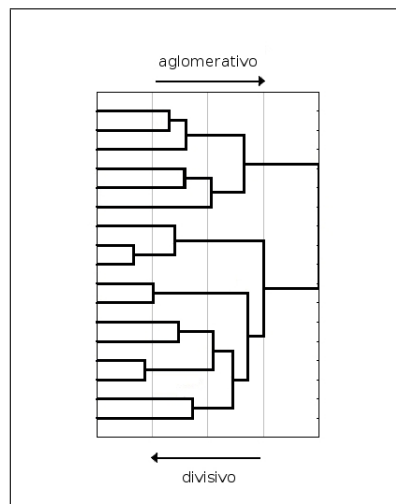
Ainda segundo Hair *et al.* (2009), os algoritmos de agrupamento podem ser classificados como hierárquico ou não-hierárquico:

Hierárquicos Os procedimentos hierárquicos envolvem uma série de $n - 1$ decisões de agrupamento (sendo n o número de observações) que combinam observações em uma estrutura de hierarquia ou do tipo de árvore.

O processo pode ser do tipo *aglomerativo*, no qual cada objeto ou observação começa com seu próprio agrupamento, e reúne-se os dois agrupamentos mais parecidos até que todas as observações estejam contidas em um único agrupamento; ou *divisivo*, no qual todas as observações iniciam em um único agrupamento e são sucessivamente divididas até que cada observação seja um agrupamento unitário.

A representação desse processo é dada pelo chamado *dendograma* ou gráfico em árvore, ilustrado na Figura 4.1.

Figura 4.1: Dendograma ilustrando agrupamento hierárquico



Fonte: (HAIR et al., 2009)

Não-Hierárquicos Diferentemente dos métodos hierárquicos, os métodos não-hierárquicos não envolvem o processo de construção em árvore. Em vez disso, designam objetos a agrupamentos assim que o número de agregados a serem formados tenha sido especificado.

O processo essencialmente tem dois passos:

1. Especificar sementes de agrupamento.

A primeira tarefa é identificar pontos de partida, conhecidos como sementes de agrupamento, para cada agregado. Uma semente de agrupamento pode ser pré-especificada pelo pesquisador ou observações podem ser escolhidas aleatoriamente.

2. Designação.

Com as sementes de agrupamento definidas, basta designar cada observação a uma das sementes de agrupamento com base em similaridade. O objetivo básico é designar cada observação à semente mais parecida.

4.2.1 O método *K-means*

O *K-means* é um método não-hierárquico de análise de agrupamentos, que busca reunir os dados em k grupos, baseando-se na medida de distância entre os objetos.

A técnica exige a definição prévia do número de grupos e do posicionamento do centro de cada grupo no espaço de atributos. Esse centro é chamado de *centróide*, que é o ponto médio do grupo (RODRIGUES, 2009).

O método é baseado pelos seguintes passos:

1. escolher k objetos para centros iniciais de k *clusters*;
2. atribuir cada amostra para o centro mais próximo;
3. recalculando os centros dos grupos, utilizando as associações de *cluster* atual;
4. retornar à etapa 2, até que os centros dos grupos se estabilizem.

O número de *clusters* k é previamente determinado pelo usuário e, portanto, o algoritmo agrupará quantos grupos forem desejados. Porém, nem todos os valores de k levam a grupos “naturais”(RODRIGUES, 2009), o que requer uma atenção com a escolha.

Sendo os objetos caracterizados através de m variáveis quantitativas, a medida de similaridade entre dois objetos A e B definida pela *distância euclidiana* é:

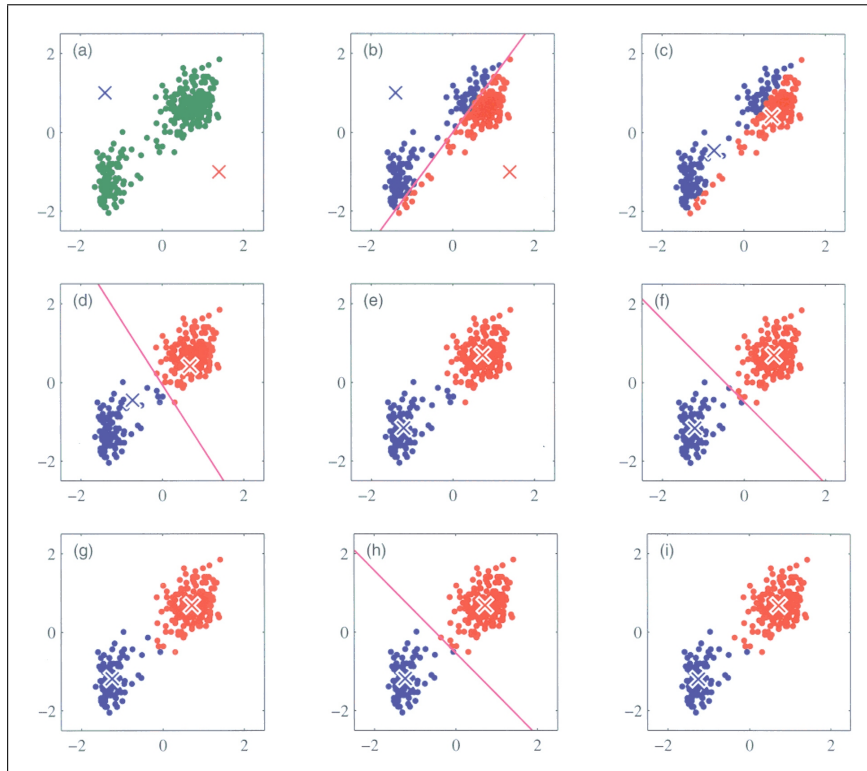
$$d(A, B) = \sqrt{(x_{1A} - x_{1B})^2 + (x_{2A} - x_{2B})^2 + \dots + (x_{mA} - x_{mB})^2}. \quad (4.4)$$

É importante salientar que todas as variáveis devem ter a mesma ordem de grandeza para o cálculo da distância. Assim, é necessário normalizar os dados.

A Figura 4.2 ilustra a aplicação do algoritmo *K-means* para agrupamento dos dados em dois grupos. O conjunto de dados está representado em (a) pelos pontos em verde, e os dois centros iniciais, escolhidos aleatoriamente, são representados pelos pontos azul e vermelho, marcados em x . Em (b) cada um dos dados é associado ao seu centro mais próximo, formando-se assim dois grupos distintos, um em vermelho, e outro em azul. Os novos centros dos *clusters* são recalculados em (c), através da média dos pontos pertencentes a cada grupo. O processo se

repete até que haja convergência do método, quando não há mais mudança (ou há mínima) nos centros dos *clusters*.

Figura 4.2: Ilustração do algoritmo *K-means*



Fonte: (QUEENSLAND, 2011)

5 Metodologia

A fim de classificar os dados de chuva de uma imagem de radar em convectivo ou estratiforme, é aplicado o método SVM, detalhado no Capítulo 3, e o *K-means*, descrito no Capítulo 4.

Os dados utilizados no estudo são provenientes do radar meteorológico do Instituto Tecnológico SIMEPAR.

5.1 Radar Meteorológico do SIMEPAR

O radar meteorológico Doppler do SIMEPAR está em operação desde outubro de 1998, sendo o primeiro do estado do Paraná. As medições do radar são realizadas numa área de até $480km$ de raio, cobrindo o estado do Paraná, Santa Catarina, parte do centro-sul de São Paulo e norte do Rio Grande do Sul. As informações disponibilizadas pelo radar permitem um melhor desempenho na vigilância em condições meteorológicas severas e na previsão a curto prazo de precipitação (BENETI; NOZU; SARAIVA, 1998).

5.1.1 Especificações Técnicas

O radar meteorológico do SIMEPAR é um radar Banda S Doppler, modelo DWSR-95S, e está em operação na região central do estado (município de Teixeira Soares, latitude $-25,51^\circ$ e longitude $-50,36^\circ$). A antena do radar DWSR-95S, de $8,2m$ de diâmetro, gera um feixe de ondas eletromagnéticas com aproximadamente 1° de abertura e monitora continuamente a atmosfera com uma sequência pré-programada de varreduras azimutais em 360° . Um volume de varredura corresponde a uma sequência completa de várias varreduras azimutais com diferentes elevações da antena.

As variáveis coletadas pelo radar meteorológico do SIMEPAR são: refletividade, velocidade radial e largura espectral (citadas na Seção 2.1.2).

O sistemas de aquisição dos dados está configurado de forma a permitir uma avaliação de alta resolução espacial e temporal das informações de precipitação e vento (BENETI; NOZU; SARAIVA, 1998).

5.2 Dados e Variáveis Utilizadas

Os dados utilizados são provenientes da variável refletividade captada pelo radar, em um raio de alcance de $200km$. Esta variável é usada para o cálculo de novos produtos e variáveis (descritas no Capítulo 2), sempre considerando a estrutura tridimensional da varredura. A classificação do *pixel* como convectivo ou estratiforme é mostrada em uma imagem de radar de um PPI de elevação $0,5^\circ$.

A interpolação utilizada foi a chamada Mapeamento por Vizinhos Próximos (MVP), citada em 2.2, em uma grade cartesiana de resolução $1km$. Ou seja, cada *pixel* representa uma área de $1km \times 1km$.

Os dados são proveniente de 7 imagens de radar, ou **conjunto de dados**, totalizando uma amostra de 352557 *pixels*. Das 7 imagens, 3 foram utilizadas como conjunto de treinamento, com um total de 101063 amostras. O restante, 251494 *pixels*, foram utilizados na validação, como conjunto de teste.

A Tabela 5.1 mostra a data e hora de cada um dos conjuntos de dados (imagens de radar) utilizados no trabalho.

Tabela 5.1: Data e hora associados a cada um dos conjuntos de dados

Imagens	Data e Hora (UTC)
Conjunto de dados 1 (CJ1)	29/04/2008 08:31
Conjunto de dados 2 (CJ2)	14/04/2009 20:01
Conjunto de dados 3 (CJ3)	24/02/2009 17:01
Conjunto de dados 4 (CJ4)	23/10/2008 23:01
Conjunto de dados 5 (CJ5)	02/11/2008 04:38
Conjunto de dados 6 (CJ6)	14/05/2009 10:48
Conjunto de dados 7 (CJ7)	19/08/2009 18:01

Fonte: A autora (2011)

Todos as imagens foram analisadas por um especialista na área, ou seja, um meteoro-

logista, que fez a especificação do que representa chuva convectiva e o que representa chuva estratiforme. Essa classificação foi feita através de análise de imagens de radar anteriores e posteriores ao horário. Os dados classificados como chuva convectiva foram identificados na própria imagem. Dessa forma, uma leitura posterior dessa imagem permitiu localizar os *pixels* marcados, e estes receberam classificação convectiva. Os dados de chuva não marcados pelo especialista, receberam, assim, classificação estratiforme.

A tarefa da máquina é aprender o mapeamento $x_i \mapsto y_i$, onde x_i é um vetor de variáveis meteorológicas que representa um *pixel*, e y_i é a classificação correspondente (no caso, 1 para amostra convectiva e -1 para estratiforme).

As variáveis escolhidas para representar cada *pixel* de chuva estão associadas a características do perfil vertical neste ponto. Assim, cada vetor x_i é formado pelas seguintes variáveis, detalhadas no Capítulo 2:

1. *Z* (refletividade):

Valor da variável refletividade no *pixel* analisado, em *dBz*. Quanto maior a refletividade, maior a quantidade de gotas de chuva, ou maior o diâmetro das gotas, e, portanto, maior a precipitação.

No entanto, o radar não detecta apenas chuva, o que pode interferir no processo.

2. *Zmax* (refletividade máxima):

A variável *Zmax*, como visto anteriormente, indica o valor da máxima refletividade vertical. Esse valor, quando acompanhado da variável *h_Zmax* (próximo item), pode ser um bom indicativo para a classificação, pois valores de *Zmax* próximos à altura da isoterma de zero grau estão relacionados à existência de banda brilhante e, portanto, chuva estratiforme. Já quando *Zmax* é encontrado em alturas mais elevadas, significa que a chuva é convectiva.

3. *h_Zmax* (altura da refletividade máxima):

Indica a altura correspondente ao valor de máxima refletividade vertical. Deve acompanhar o valor de *Zmax*, pelos motivos citados acima.

4. *Zmax_BB* (refletividade máxima dentro do intervalo de banda brilhante - isoterma de zero grau):

Valor máximo de refletividade encontrado dentro do intervalo de *1km* acima ou abaixo da altura da isoterma de zero grau, em *dBz*.

5. *h₃₀* (altura da refletividade de 30dBz):

Alturas elevadas mostram que a tempestade em grandes altitudes é mais intensa, o que está associado à existência de chuva convectiva.

6. *h₄₅* (altura da refletividade de 45dBz):

Seguindo o mesmo raciocínio, alturas elevadas estão relacionadas à existência de chuva convectiva.

7. *IceSlope*:

Representa a variação dos valores de refletividade acima da isoterma de zero grau. Como pode ser observado nos perfis verticais representativos de chuva estratiforme e de chuva convectiva (Seção 2.4.3), uma maior variação da refletividade no topo da nuvem é um indicativo de chuva convectiva.

8. *VIL*:

Como esta variável representa a integração vertical de toda água líquida, valores maiores estão relacionados a maior quantidade de precipitação, ou seja, precipitação convectiva.

9. *EchoTop*:

Indica a altura do topo da nuvem. Nuvens de chuva convectiva geralmente possuem maior extensão vertical.

10. *VILD*:

Como comentado anteriormente, o *VILD* pode ser usado para identificar tempestades com alta refletividade em relação à sua altura (ou seja, chuva convectiva).

É importante salientar que o especialista não utilizou essas variáveis na classificação dos dados, apenas imagens de refletividade.

5.2.1 Normalização

Como as variáveis possuem grandezas de dimensões diferentes, é conveniente normalizar os dados. Isso é feito atribuindo uma escala no intervalo $[-1, 1]$ para cada uma das variáveis.

Dado um vetor de variáveis, a sua normalização é feita dividindo-se todos os valores pela norma desse vetor.

O objetivo da normalização é evitar que variáveis com escalas significativamente grandes dominem as variáveis de menor escala.

5.3 O software R

R é um *software* livre para análise de dados em geral. O programa está disponível em <http://www.r-project.org> (R-PROJECT, 2009). Apresenta código fonte aberto, podendo ser modificado ou implementado com novos procedimentos por qualquer usuário. Além disso, o R possui um grande número de colaboradores das mais diversas áreas do conhecimento.

5.3.1 SVM

Empregando-se o *software* do R-project (R-PROJECT, 2009), é usada a implementação LIBSVM de Chang e Lin (CHANG; LIN, 2001) através da função *svm()*, após a instalação do pacote *e1071*. Este pacote pode ser utilizado tanto para classificação quanto para regressão.

As entradas para a função *svm()* são os dados de treinamento, definidos por uma matriz de dados X e um correspondente vetor de respostas y , e uma função *kernel*, com seus respectivos parâmetros. Assim, a chamada da função dá-se da forma:

$$svm(X, y, type, kernel, ...),$$

onde:

- X é a matriz com os dados de treinamento já normalizados, na qual cada linha representa uma amostra e cada coluna, uma variável definida na Seção 5.2.
- y é o vetor de respostas correspondente à matriz de dados. Cada linha i da matriz X tem um padrão definido em y_i . Para este caso, definiu-se 1 para convectiva e -1 para estratiforme, conforme classificação do especialista.
- *type* determina se o svm será usado para classificação (C) ou regressão (R).
- *kernel* é a função *kernel* escolhida para realizar o mapeamento. Cada uma delas possui parâmetros livres que precisam ser definidos pelo usuário. A Tabela 5.2 especifica as possíveis funções *kernel* e seus respectivos parâmetros de entrada:

Tabela 5.2: Funções *kernel*

Kernel	Função $K(x_i, x_j)$	Parâmetros
linear	$x_i \cdot x_j$	-
radial (gaussiano)	$\exp(-\text{gamma}\ x_i - x_j\ ^2)$	<i>gamma</i>
sigmoid	$\tanh(\text{gamma}(x_i \cdot x_j) + c_0)$	<i>gamma, c₀</i>
polynomial	$\text{gamma}((x_i \cdot x_j) + c_0)^{\text{degree}}$	<i>gamma, degree, c₀</i>

Fonte: (MEYER, 2011)

Além disso, um parâmetro comum a todas as possibilidades de *kernel* é a variável C , presente na função objetivo do modelo 3.27, que determina menor ou maior importância às variáveis de folga do problema. Na chamada da função *svm()* esse parâmetro é representado por *cost*.

A seguir, é exemplificada a chamada da função *svm()* para cada uma das possíveis funções *kernel*.

1. Linear:

$$\text{svm}(X, y, \text{type} = 'C', \text{kernel} = 'linear', \text{cost} =).$$

2. Radial (Gaussiana):

$$\text{svm}(X, y, \text{type} = 'C', \text{kernel} = 'radial', \text{gamma} = , \text{cost} =).$$

3. Sigmoidal:

$$\text{svm}(X, y, \text{type} = 'C', \text{kernel} = 'sigmoid', \text{gamma} = , c_0 = , \text{cost} =).$$

4. Polinomial:

$$\text{svm}(X, y, \text{type} = 'C', \text{kernel} = 'polynomial', \text{gamma} = , c_0 = , \\ \text{degree} = , \text{cost} =).$$

A função retorna um objeto da classe SVM (SVM_{class}) que contém o modelo ajustado. Novas amostras podem ser classificadas de acordo com esse modelo através da função *predict()*. A chamada dessa função é feita da seguinte forma:

$$\text{predict}(SVM_{class}, N),$$

onde SVM_{class} é o objeto da classe SVM obtido pela função $svm()$ e N é uma matriz com os novos dados a serem classificados.

5.3.2 *K-means*

Para aplicação do método *K-means* no R, é utilizada a função $kmeans()$ da seguinte forma:

$$kmeans(X, centers),$$

onde $X \in \mathfrak{R}^{n \times m}$ é a matriz de dados, sendo n o número de amostras e m o número de características de cada amostra; e $centers$ é o número de *clusters*.

Esta função retorna um objeto da classe *kmeans*, que é uma lista com os seguintes componentes:

- *cluster*: Um vetor de inteiros indicando o *cluster* para qual cada ponto é atribuído;
- *centers*: A matriz contendo os centros de cada *cluster*;
- *withinss*: A soma dos quadrados das distâncias internas para cada *cluster*;
- *size*: O número de amostras atribuídas a cada *cluster*.

6 Aplicação e Resultados

Este capítulo apresenta a aplicação dos métodos mostrados anteriormente, SVM e *K-means*, no problema apresentado no Capítulo 2.

Ambas as técnicas são aplicadas utilizando o *software* R, e o detalhamento dessa aplicação, bem como os resultados, são mostrados a seguir. Uma comparação entre os dois métodos também é apresentada.

6.1 SVM

Os parâmetros usados na formulação do SVM são: a variável C , e as variáveis da própria função *kernel*. Não é possível saber de antemão qual a função *kernel* mais adequada, bem como seus melhores parâmetros, para a melhor classificação dos dados.

O número de vetores suporte (VS) não é definido, pois depende da dimensão do espaço onde ocorre a separação linear dos dados, ou seja, o espaço característico. Não se tem conhecimento da dimensão que a função *kernel* gera neste espaço, sendo impossível determinar o número mínimo de vetores suporte para o reconhecimento de padrões. Caso a maioria das amostras seja vetor suporte, os dados estão classificados exageradamente, ocorrendo *overfitting*. Portanto, quanto menor o número de vetores suporte, melhor é a generalização na classificação (ALES, 2008).

Dessa forma, diversos treinamentos foram realizados, com diferentes funções *kernel* e variação em seus parâmetros, bem como na variável C , com o objetivo de encontrar uma configuração que melhor se adapte ao problema. Algumas das configurações de parâmetros testadas são apresentadas na Tabela 6.1, totalizando 25 configurações diferentes. Cada uma delas é representada por uma sigla, para posterior análise dos resultados.

Tabela 6.1: Configurações de parâmetros para as funções *kernel*

Sigla	Kernel	Parâmetros do <i>kernel</i>	C
L1	Linear	-	0,01
L2	Linear	-	0,1
L3	Linear	-	1
L4	Linear	-	10
L5	Linear	-	100
R1	Radial	radial($gamma = 0,01$)	0,1
R2	Radial	radial($gamma = 0,1$)	0,1
R3	Radial	radial($gamma = 0,1$)	0,01
R4	Radial	radial($gamma = 0,1$)	1
R5	Radial	radial($gamma = 0,5$)	0,1
R6	Radial	radial($gamma = 1$)	0,1
S1	Sigmoidal	sigmoid($gamma = 0,1; c_0 = 0$)	1
S2	Sigmoidal	sigmoid($gamma = 0,1; c_0 = 20$)	1
S3	Sigmoidal	sigmoid($gamma = 0,5; c_0 = 0$)	1
S4	Sigmoidal	sigmoid($gamma = 0,5; c_0 = 20$)	1
S5	Sigmoidal	sigmoid($gamma = 0,1; c_0 = 0$)	10
S6	Sigmoidal	sigmoid($gamma = 0,5; c_0 = 0$)	10
P1	Polinomial	polynomial($gamma = 0,1; c_0 = 0; degree = 3$)	1
P2	Polinomial	polynomial($gamma = 0,1; c_0 = 0; degree = 5$)	1
P3	Polinomial	polynomial($gamma = 1, c_0 = 0; degree = 3$)	0,1
P4	Polinomial	polynomial($gamma = 1, c_0 = 0; degree = 5$)	0,1
P5	Polinomial	polynomial($gamma = 0,1; c_0 = 0; degree = 3$)	10
P6	Polinomial	polynomial($gamma = 0,1; c_0 = 0; degree = 5$)	10
P7	Polinomial	polynomial($gamma = 0,01; c_0 = 0; degree = 3$)	1
P8	Polinomial	polynomial($gamma = 0,01; c_0 = 0; degree = 2$)	1

Fonte: A autora (2011)

Os primeiros resultados são apresentados na Tabela 6.2, e dizem respeito ao tempo de execução e número de vetores suporte (VS) obtidos, mostrado também em porcentagem do total de dados.

Tabela 6.2: Tempo de execução e número de vetores suporte para cada *kernel* estabelecido

Sigla	Tempo (minutos)	Número VS	Porcentagem VS
L1	14	35881	35,50%
L2	11	35140	34,77%
L3	16	35009	34,64%
L4	48	34995	34,63%
L5	317	34996	34,63%
R1	39	34879	34,51%
R2	31	31150	30,82%
R3	30	29100	28,79%
R4	30	29100	28,79%
R5	35	34087	33,73%
R6	41	38296	37,89%
S1	40	27705	27,41%
S2	54	60116	59,48%
S3	45	27672	27,38%
S4	46	53708	53,14%
S5	39	27696	27,40%
S6	40	27672	27,38%
P1	29	29922	29,61%
P2	166	28950	28,65%
P3	18	30780	30,46%
P4	45	30276	29,96%
P5	128	29492	29,18%
P6	1139	27955	27,66%
P7	26	37371	36,98%
P8	21	39511	39,10%

Fonte: A autora (2011)

Para o *kernel* linear, a variação do parâmetro C pouco influenciou no número de VS, porém o tempo de execução ficou bastante elevado na configuração L5, na qual $C = 100$.

Na configuração para *kernel* radial, observa-se um aumento no tempo de execução, com relação ao *kernel* linear, porém alguns casos apresentam redução no número de VS, vari-

ando de uma média de 34% no *kernel* linear para até 28%.

O *kernel* sigmoidal apresentou, em geral, maior tempo computacional. Apesar de algumas configurações apresentarem maioria dos dados como VS (S2 e S4), as demais obtiveram uma redução no número de VS, variando em torno de 27% das amostras.

Analisando, finalmente, as configurações de *kernel* polinomial, observa-se que algumas apresentam tempo de execução mais elevado (P2, P5 e P6). Porém, com relação ao número de VS, a maioria apresentou resultado abaixo de 30%.

A análise feita a seguir é a respeito da porcentagem de dados classificados corretamente em relação à classificação feita pelo especialista. Os resultados obtidos pelo SVM são comparados com um método de classificação de chuva já existente na literatura, o método de Steiner *et al.* (STEINER; HOUZE; YUTER, 1995), cuja descrição é mostrada na Seção 2.6.1. A Tabela 6.3 mostra o **índice de acerto**, que é a porcentagem de *pixels* classificados corretamente em cada um dos 7 conjuntos de dados. Valores em negrito indicam índices de acerto melhores que o obtido pelo método de Steiner *et al.* (1995).

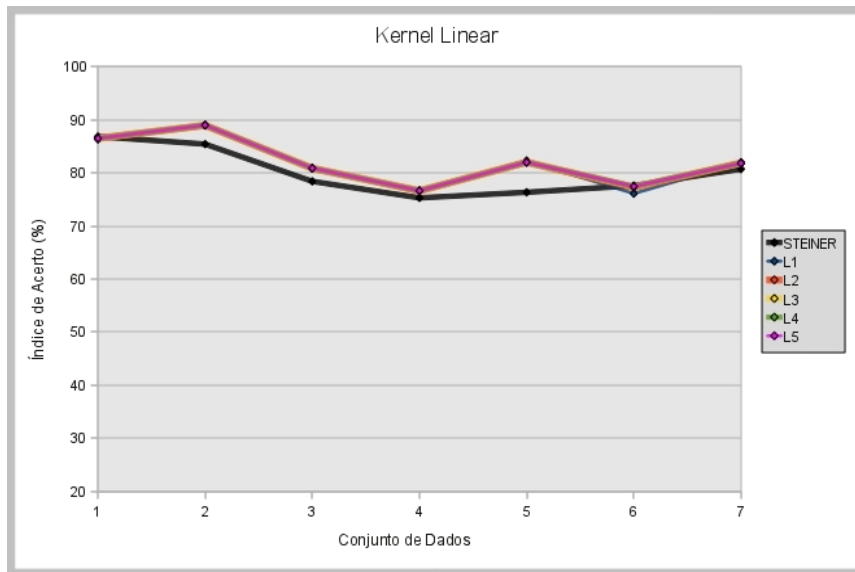
Tabela 6.3: Índices de acerto do método SVM para cada conjunto de dados

Sigla	CJ1	CJ2	CJ3	CJ4	CJ5	CJ6	CJ7
Steiner	86,82%	85,41%	78,39%	75,28%	76,33%	77,58%	80,72%
L1	86,30%	89,00%	80,88%	76,57%	82,31%	76,09%	81,97%
L2	86,50%	88,94%	80,86%	76,60%	82,00%	77,22%	81,85%
L3	86,49%	88,98%	80,88%	76,62%	81,97%	77,39%	81,80%
L4	86,48%	88,98%	80,87%	76,62%	81,97%	77,43%	81,80%
L5	86,48%	88,98%	80,88%	76,62%	81,98%	77,43%	81,79%
R1	87,02%	89,20%	82,96%	77,19%	80,95%	78,81%	81,80%
R2	87,82%	89,14%	84,58%	77,69%	77,53%	80,43%	80,90%
R3	87,08%	88,41%	83,65%	77,13%	78,54%	79,51%	79,96%
R4	88,44%	89,71%	85,14%	77,68%	77,93%	80,00%	82,01%
R5	88,26%	89,66%	85,31%	77,15%	78,22%	76,49%	79,51%
R6	88,37%	89,65%	85,97%	76,68%	78,91%	73,49%	78,47%
S1	71,26%	76,61%	74,42%	71,72%	79,20%	61,19%	73,48%
S2	74,22%	71,79%	58,80%	62,54%	61,94%	91,40%	60,06%
S3	72,68%	71,75%	72,86%	72,64%	64,19%	59,83%	69,85%
S4	70,07%	24,10%	29,52%	40,44%	29,00%	66,93%	34,15%
S5	71,26%	76,59%	74,41%	71,72%	79,18%	61,18%	73,47%
S6	72,68%	71,75%	72,85%	72,63%	64,18%	59,82%	69,84%
P1	88,06%	88,82%	83,99%	77,15%	77,01%	82,73%	81,22%
P2	88,98%	88,98%	83,85%	76,57%	76,00%	84,64%	81,23%
P3	87,87%	88,55%	83,56%	76,87%	76,65%	83,60%	80,88%
P4	88,60%	88,46%	82,76%	76,00%	75,36%	86,52%	80,28%
P5	88,22%	88,93%	84,15%	77,36%	77,12%	82,49%	81,14%
P6	89,20%	89,43%	84,70%	77,06%	76,78%	83,65%	81,68%
P7	87,22%	87,19%	78,10%	73,86%	75,20%	88,69%	77,36%
P8	85,59%	87,15%	80,66%	74,36%	76,21%	87,65%	76,75%

Fonte: A autora (2011)

Os valores da Tabela 6.3 são ilustrados pelos gráficos abaixo, separados por função *kernel*. A curva em preto representa os índices de acerto para o método de Steiner *et al.* (1995), enquanto as coloridas representam as diferentes configurações de *kernel*.

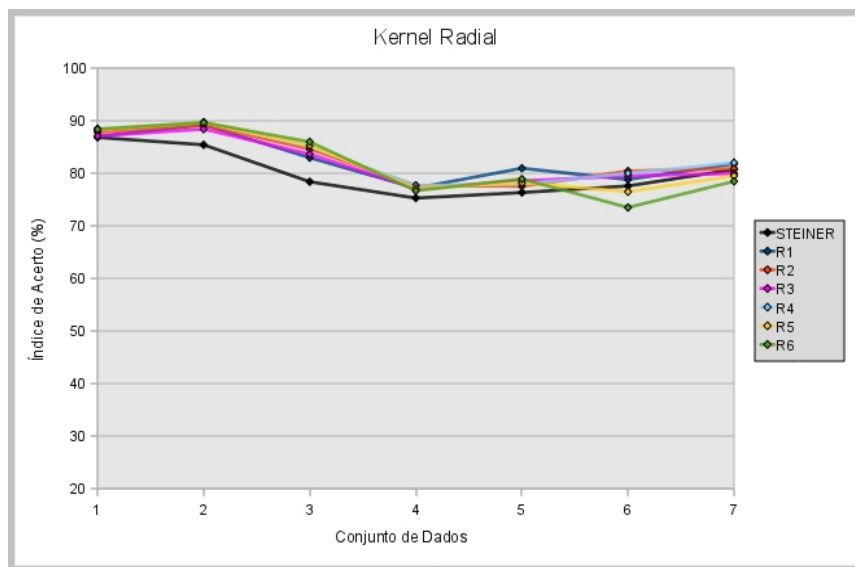
Figura 6.1: Índices de acerto para *kernel linear*



Fonte: A autora (2011)

Para *kernel linear*, observa-se que as diferentes configurações apresentam resultados praticamente iguais. A maioria dos resultados é superior ao de Steiner *et al.* (1995), com exceção do conjunto de dados 1 (CJ1) e 6 (CJ6).

Figura 6.2: Índices de acerto para *kernel radial*

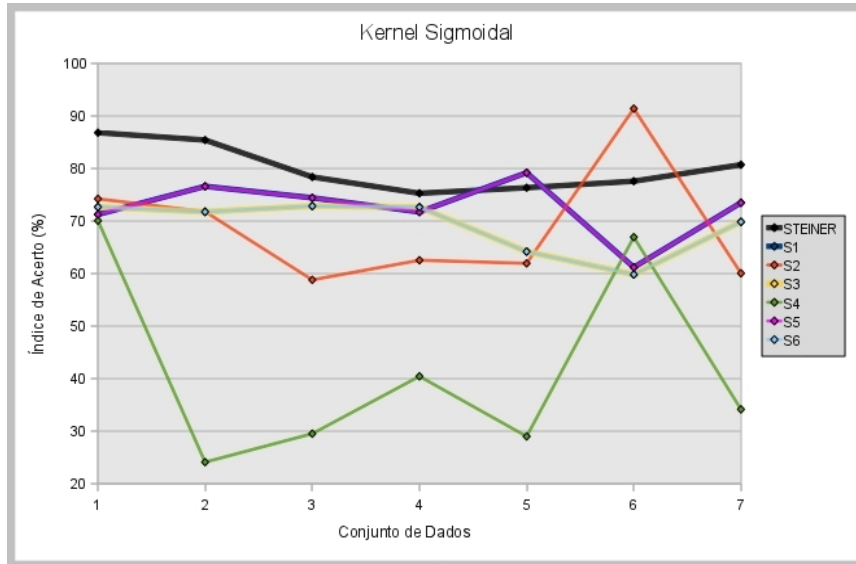


Fonte: A autora (2011)

Para *kernel radial*, algumas configurações tiveram índices inferiores ao de Steiner *et*

al. (1995) somente para o conjunto de dados 6 e 7.

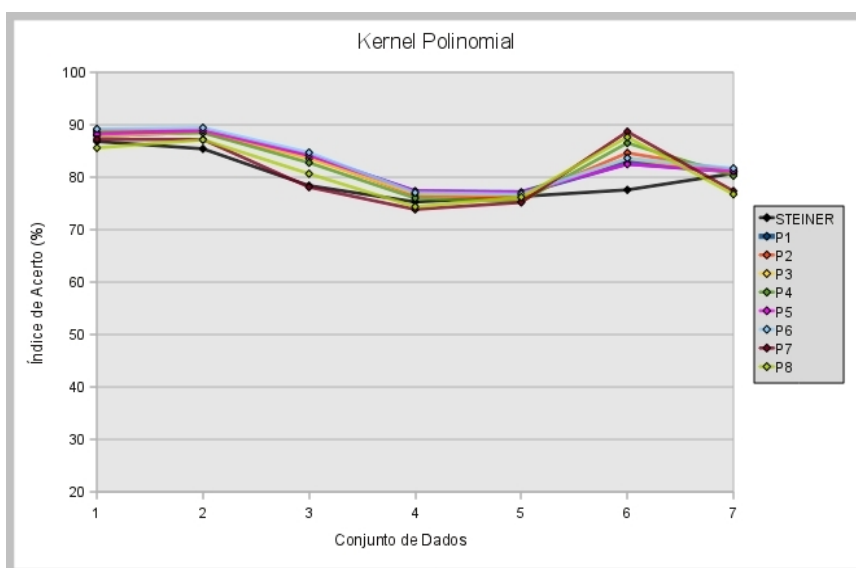
Figura 6.3: Índices de acerto para *kernel* sigmoidal



Fonte: A autora (2011)

O *kernel* sigmoidal não apresentou bons resultados. Quase todas as configurações apresentaram índices muito inferiores ao de Steiner *et al.* (1995), com exceção apenas da configuração S2 para o conjunto de dados 6, e S1 e S5 para conjunto de dados 5.

Figura 6.4: Índices de acerto para *kernel* polinomial



Fonte: A autora (2011)

O *kernel* polinomial apresentou, em geral, bons resultados. Apenas alguns índices de acerto obtidos foram inferiores ao índice obtido por Steiner *et al.* (1995).

A Tabela 6.4 apresenta uma média das taxas de acerto para dois conjunto distintos: o conjunto de treinamento, formado pelos três primeiros conjuntos de dados (ou imagens), o conjunto de teste, formado pelos dados restantes; e também a média geral de acerto para todo o conjunto de dados (352557 pontos).

Tabela 6.4: Índices de acerto do método SVM para conjuntos de treinamento e de teste

Sigla	Conjunto de Treinamento	Conjunto de Teste	Média Geral
L1	85,40%	79,24%	82,32%
L2	85,43%	79,42%	82,43%
L3	85,45%	79,45%	82,45%
L4	85,44%	79,46%	82,45%
L5	85,45%	79,46%	82,45%
R1	86,39%	79,69%	83,04%
R2	87,18%	79,14%	83,16%
R3	86,38%	78,79%	82,58%
R4	87,76%	79,41%	83,58%
R5	87,74%	77,84%	82,79%
R6	88,00%	76,89%	82,44%
S1	74,10%	71,40%	72,75%
S2	68,27%	68,99%	68,63%
S3	72,43%	66,63%	69,53%
S4	41,23%	42,63%	41,93%
S5	74,09%	71,39%	72,74%
S6	72,43%	66,62%	69,52%
P1	86,96%	79,53%	83,24%
P2	87,27%	79,61%	83,44%
P3	86,66%	79,50%	83,08%
P4	86,61%	79,54%	83,07%
P5	87,10%	79,53%	83,31%
P6	87,78%	79,79%	83,78%
P7	84,17%	78,78%	81,46%
P8	84,47%	78,74%	81,60%

Fonte: A autora (2011)

Pode-se observar que a configuração P6 apresenta os valores mais elevados tanto para conjunto de treinamento como conjunto de teste e, conseqüentemente, possui maior média geral.

Com o objetivo de analisar melhor o resultado do método proposto, é calculada uma taxa que compara o índice de acerto do método SVM ($indice_{svm}$) com o do método de Steiner

et al. ($indice_{st}$). Este índice, denominado $I_{melhoria}$, calcula a melhoria na taxa de amostras classificadas corretamente pelo método SVM com relação ao método de Steiner *et al.* (1995), e é calculado da seguinte forma:

$$I_{melhoria} = \frac{indice_{svm}}{indice_{st}}. \quad (6.1)$$

Valores maiores que 1 indicam uma melhoria na porcentagem de amostras classificadas corretamente pelo método SVM, com relação ao método de Steiner *et al.* (1995). Analogamente, valores abaixo de 1 indicam que a técnica SVM não conseguiu superar o índice de acerto do método de Steiner *et al.* (1995).

Supondo $indice_{svm} = 90\%$, por exemplo. Se $indice_{st} = 89\%$, então $I_{melhoria} = 1,011235$, ou seja, o método SVM apresentou um resultado 1,12% melhor que o método de Steiner *et al.* Porém, supondo $indice_{st} = 70\%$, então $I_{melhoria} = 1,285714$. Ou seja, a performance do SVM foi 28,57% melhor. De forma análoga, valores abaixo de 1, indicam uma redução no índice de acerto do método SVM em relação ao método de Steiner *et al.* (1995).

Os valores calculados para cada configuração são mostrados na Tabela 6.5:

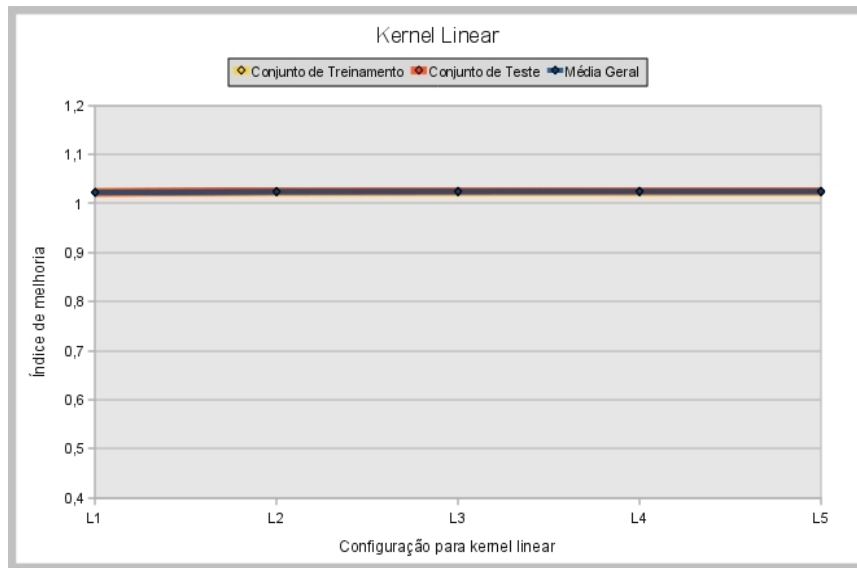
Tabela 6.5: Índices de melhoria para conjunto de treinamento, conjunto de teste e conjunto total de dados

Sigla	Conjunto de Treinamento	Conjunto de Teste	Média Geral
L1	1,022718	1,022689	1,022701
L2	1,023051	1,025043	1,024189
L3	1,023254	1,025404	1,024483
L4	1,023173	1,025533	1,024522
L5	1,023215	1,025535	1,024541
R1	1,034992	1,028532	1,031301
R2	1,044718	1,021427	1,031409
R3	1,035073	1,017002	1,024747
R4	1,051704	1,024753	1,036303
R5	1,051541	1,004897	1,024887
R6	1,054731	0,992710	1,019290
S1	0,889034	0,922112	0,907936
S2	0,815167	0,890924	0,858457
S3	0,868885	0,860392	0,864032
S4	0,488606	0,550624	0,524045
S5	0,888913	0,921984	0,907811
S6	0,868843	0,860263	0,863940
P1	1,041882	1,026333	1,032997
P2	1,045443	1,027284	1,035067
P3	1,038270	1,025975	1,031244
P4	1,037320	1,026414	1,031088
P5	1,043606	1,026369	1,033756
P6	1,051658	1,029668	1,039093
P7	1,007249	1,016471	1,012673
P8	1,011721	1,016471	1,014435

Fonte: A autora (2011)

Os valores da Tabela 6.5 são ilustrados nos gráficos abaixo. A curva em amarelo representa os valores obtidos para o conjunto de treinamento, a vermelha para o conjunto de teste, e a azul para a média geral de todo o conjunto de dados.

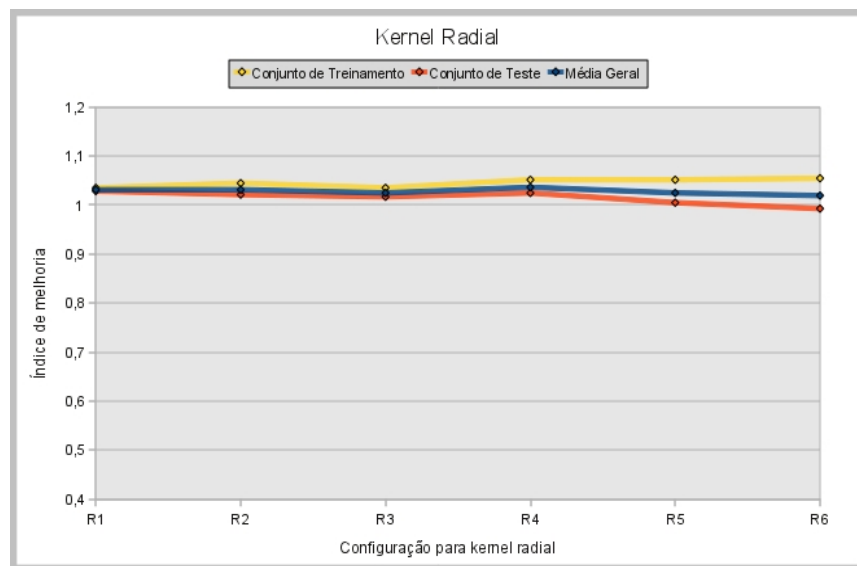
Figura 6.5: Índice de melhoria para *kernel* linear



Fonte: A autora (2011)

Observa-se que o *kernel* linear obteve índice de melhoria acima de 1 para todas as configurações, tanto para conjunto de treinamento, quanto para conjunto de teste. Os valores obtidos para conjunto de treinamento e de teste são muito semelhantes.

Figura 6.6: Índice de melhoria para *kernel* radial

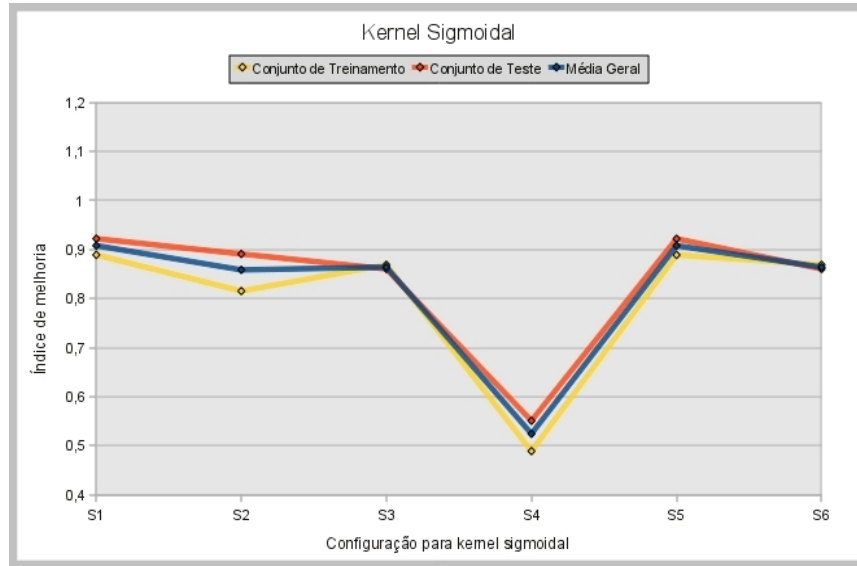


Fonte: A autora (2011)

O *kernel* radial também obteve bons resultados, sendo o único valor abaixo de 1 obtido

na configuração R6, para o conjunto de teste.

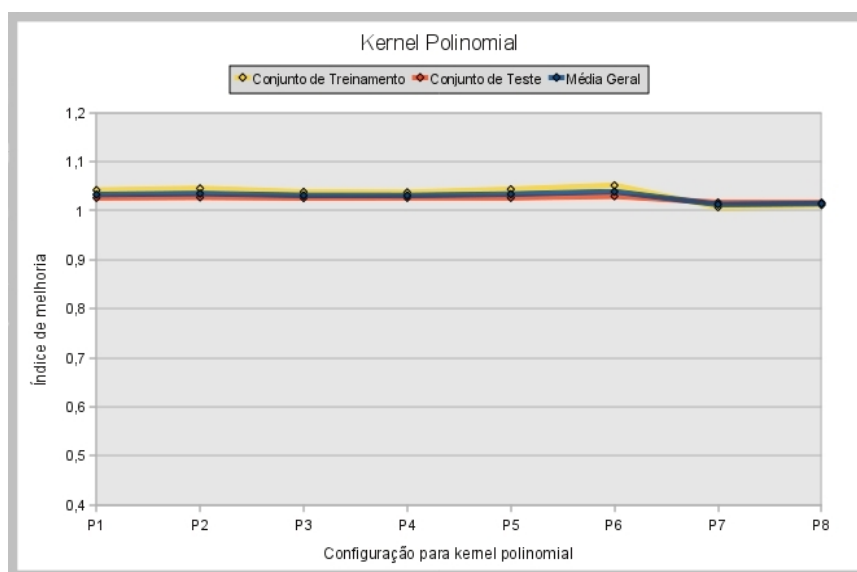
Figura 6.7: Índice de melhoria para *kernel* sigmoidal



Fonte: A autora (2011)

O *kernel* sigmoidal apresentou resultados muito ruins, visto que os índices de melhoria são todos abaixo de 1, ou seja, não houve resultados melhores que o método de Steiner *et al.* (1995) em nenhum caso.

Figura 6.8: Índice de melhoria para *kernel* polinomial



Fonte: A autora (2011)

Todas as configurações de *kernel* polinomial apresentaram índice de melhoria superior a 1, tanto no conjunto de treinamento, quanto no de teste.

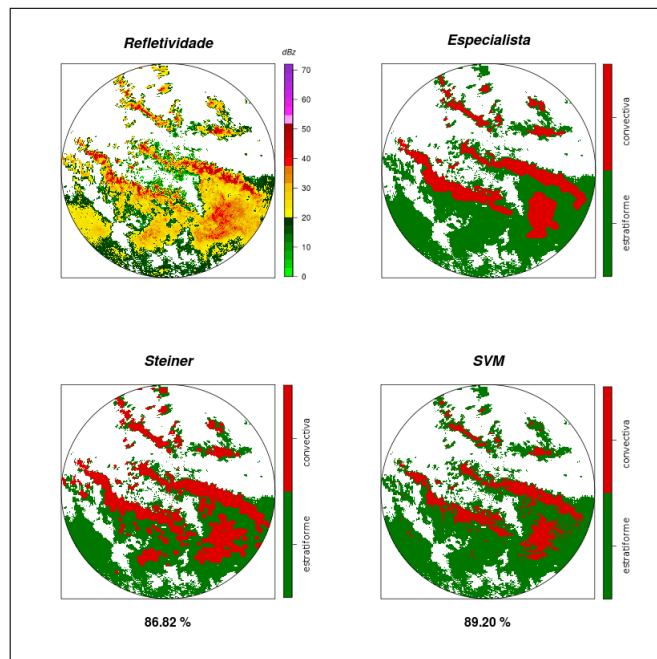
Assim, como P6 obteve os melhores resultados, tanto para índice de acerto, quanto para índice de melhoria, os resultados obtidos por P6 são ilustrados a seguir.

São apresentadas figuras contendo: a imagem de refletividade original, e mais três imagens de classificação em convectiva ou estratiforme: uma classificada pelo especialista, outra com os resultados obtidos pela técnica de Steiner *et al.* (STEINER; HOUZE; YUTER, 1995), e uma terceira com os resultados obtidos pelo método SVM proposto.

As porcentagens indicadas são referentes à taxa de amostras identificadas corretamente, de acordo com a imagem classificada pelo meteorologista.

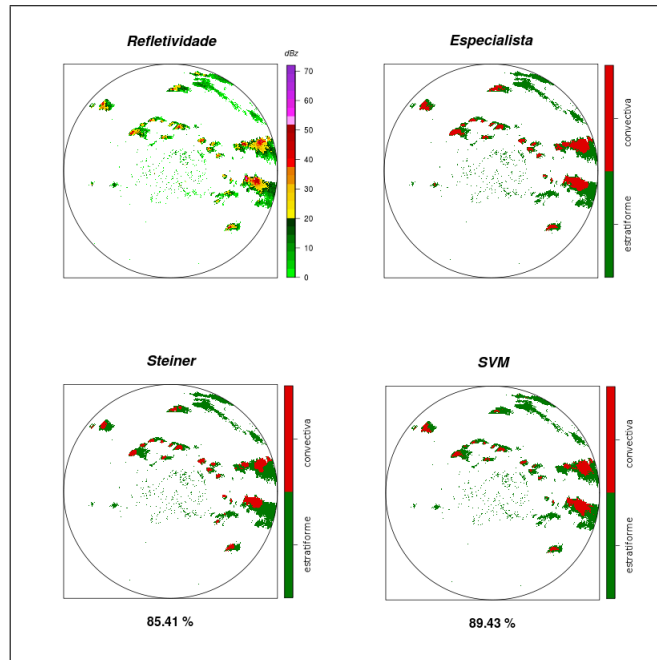
As três primeiras figuras são referentes aos dados utilizados como conjunto de treinamento, enquanto as demais, contêm dados do conjunto de teste.

Figura 6.9: Resultados obtidos pelo método SVM e comparação com a técnica de Steiner *et al.*, para conjunto de dados 1



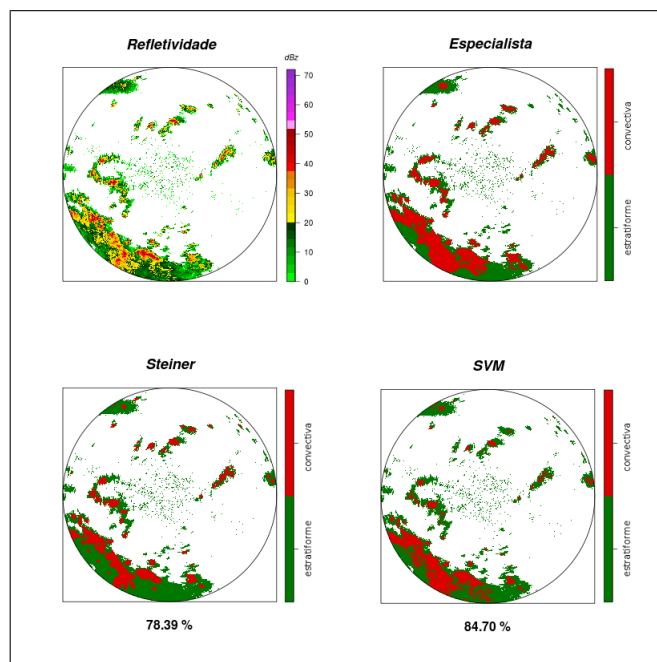
Fonte: A autora (2011)

Figura 6.10: Resultados obtidos pelo método SVM e comparação com a técnica de Steiner *et al.*, para conjunto de dados 2



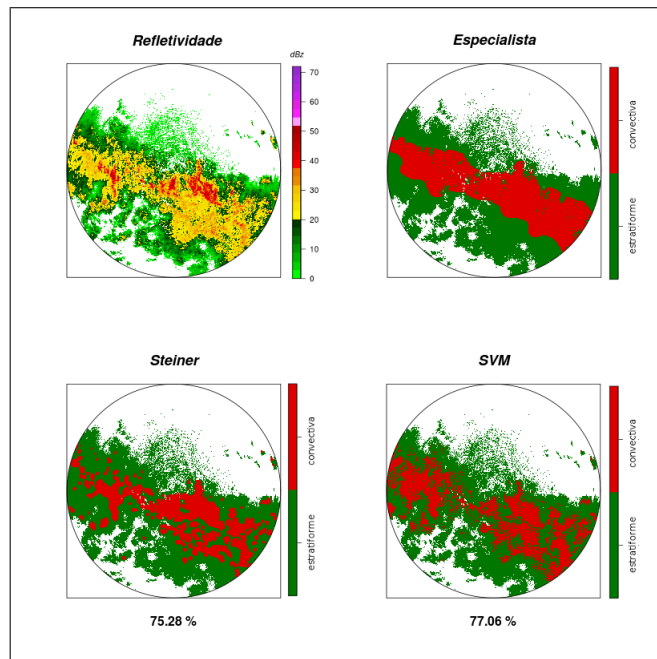
Fonte: A autora (2011)

Figura 6.11: Resultados obtidos pelo método SVM e comparação com a técnica de Steiner *et al.*, para conjunto de dados 3



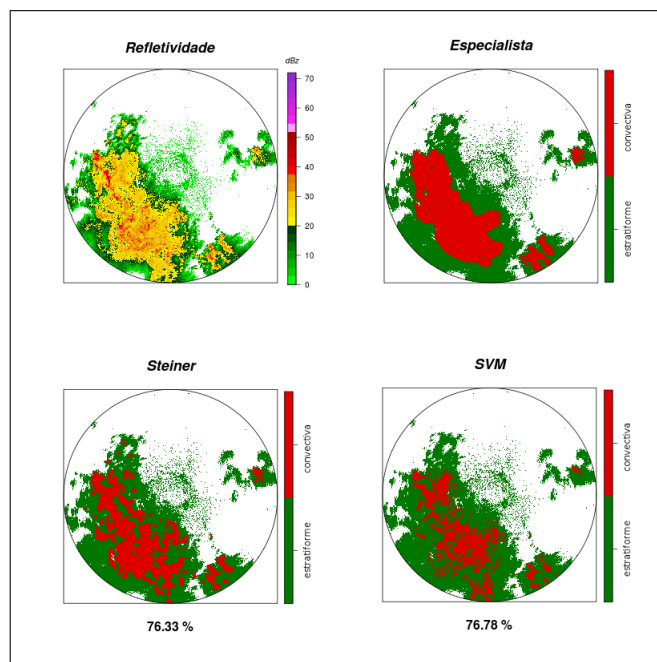
Fonte: A autora (2011)

Figura 6.12: Resultados obtidos pelo método SVM e comparação com a técnica de Steiner *et al.*, para conjunto de dados 4



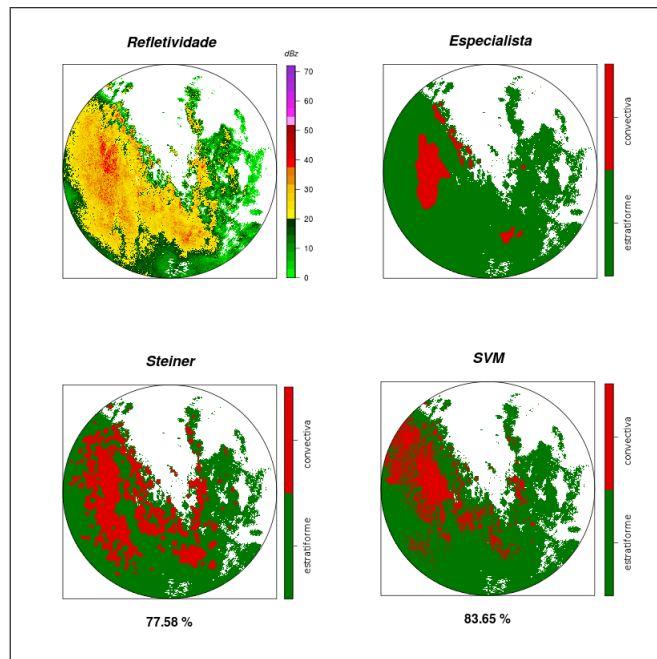
Fonte: A autora (2011)

Figura 6.13: Resultados obtidos pelo método SVM e comparação com a técnica de Steiner *et al.*, para conjunto de dados 5



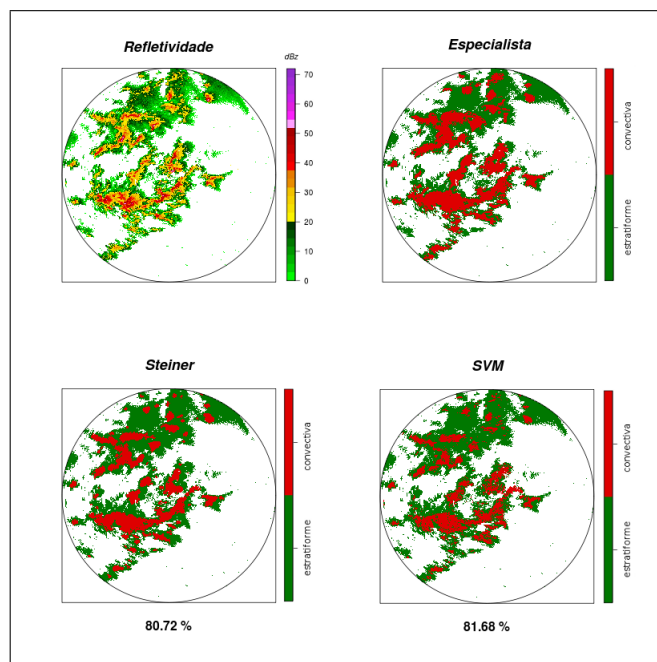
Fonte: A autora (2011)

Figura 6.14: Resultados obtidos pelo método SVM e comparação com a técnica de Steiner *et al.*, para conjunto de dados 6



Fonte: A autora (2011)

Figura 6.15: Resultados obtidos pelo método SVM e comparação com a técnica de Steiner *et al.*, para conjunto de dados 7



Fonte: A autora (2011)

6.2 K-means

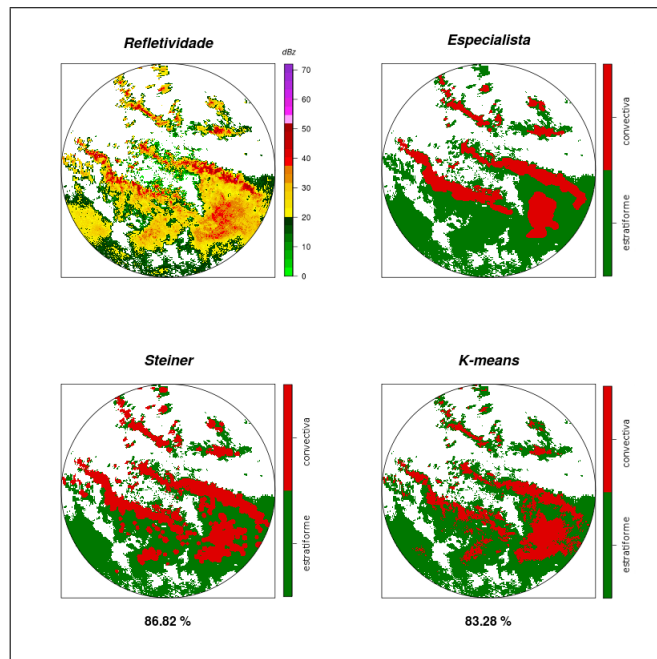
O *K-means* é utilizado com o objetivo de reunir as amostras em dois grupos, ou *clusters*, um para dados que representam chuva estratiforme, e outro para chuva convectiva. Como esses dois eventos meteorológicos possuem características distintas, espera-se que as variáveis escolhidas para representar cada um dos dados (Seção 5.2) possam caracterizar bem essa diferença, de forma que seja possível agrupar os objetos e que cada um dos grupos encontrados represente bem cada um dos eventos.

O *K-means* é aplicado com todo o conjunto de dados, ou seja, 352557 amostras.

O tempo de execução para esta técnica é menor do que 1 segundo. Os dois grupos obtidos representam bem o que se era esperado, ou seja, um com características de chuva estratiforme, e outro com características de precipitação convectiva. O *cluster* de dados estratiformes agrupou 278057 amostras; enquanto o *cluster* de dados convectivos teve 74500 amostras.

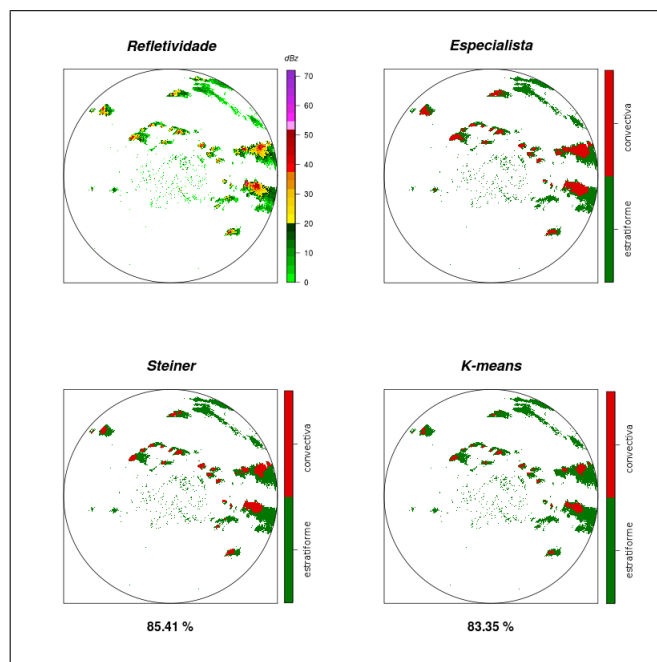
As imagens com os resultados da classificação obtida pelo método são mostradas a seguir. Optou-se por ilustrar os resultados de forma semelhante à feita na seção anterior, para o método SVM, ou seja, com a imagem da classificação feita pelo especialista, a classificação feita pelo método Steiner *et al.* (STEINER; HOUZE; YUTER, 1995), e a obtida pelo método em estudo, o *K-means*. Isso porque considera-se a ideia de que a classificação do especialista é o padrão a ser alcançado, portanto, o *índice de acerto*, ou seja, porcentagem de dados classificados corretamente, é feita de acordo com essa classificação. A imagem do método de Steiner *et al.* também é colocada para comparação, por ser uma técnica já existente na literatura e bastante usada. O índice de acerto mostrado para essa técnica também é em comparação ao especialista.

Figura 6.16: Resultados obtidos pelo método *K-means* e comparação com a técnica de Steiner *et al.* para conjunto de dados 1



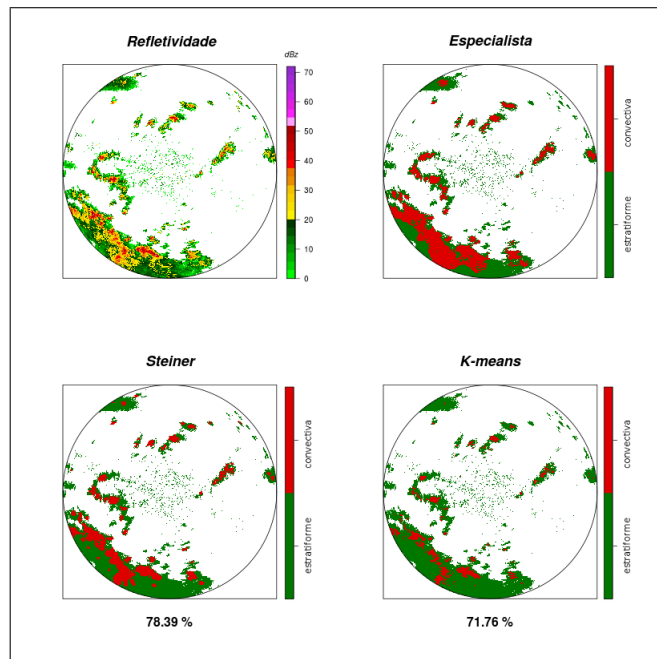
Fonte: A autora (2011)

Figura 6.17: Resultados obtidos pelo método *K-means* e comparação com a técnica de Steiner *et al.* para conjunto de dados 2



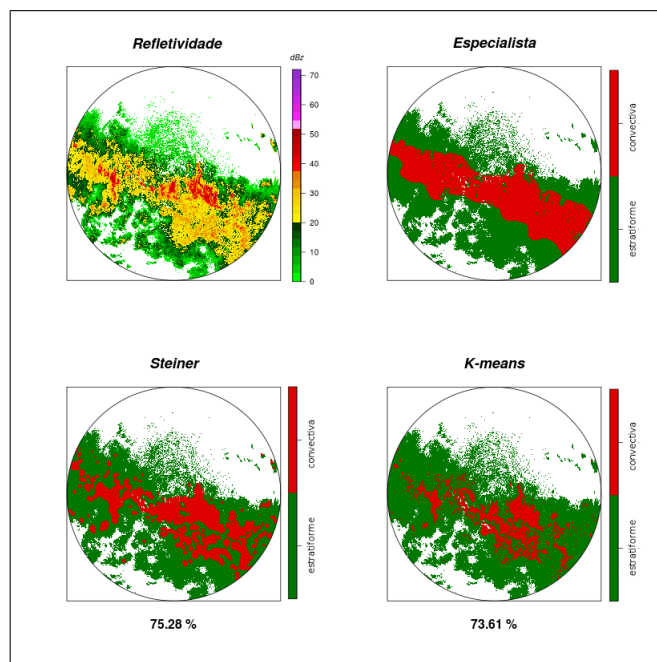
Fonte: A autora (2011)

Figura 6.18: Resultados obtidos pelo método *K-means* e comparação com a técnica de Steiner *et al.* para conjunto de dados 3



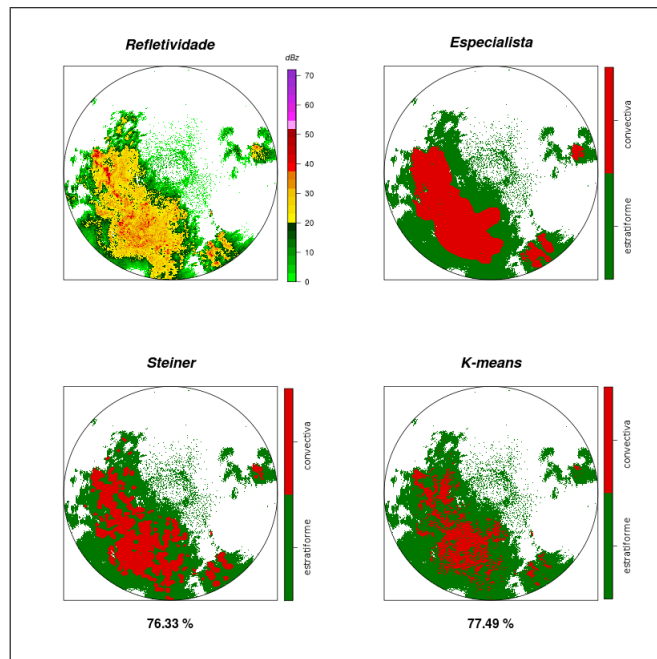
Fonte: A autora (2011)

Figura 6.19: Resultados obtidos pelo método *K-means* e comparação com a técnica de Steiner *et al.* para conjunto de dados 4



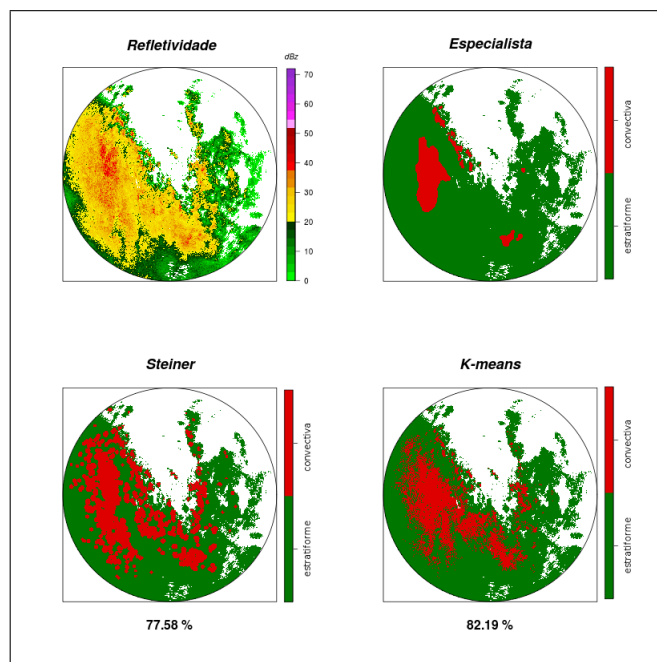
Fonte: A autora (2011)

Figura 6.20: Resultados obtidos pelo método *K-means* e comparação com a técnica de Steiner *et al.* para conjunto de dados 5



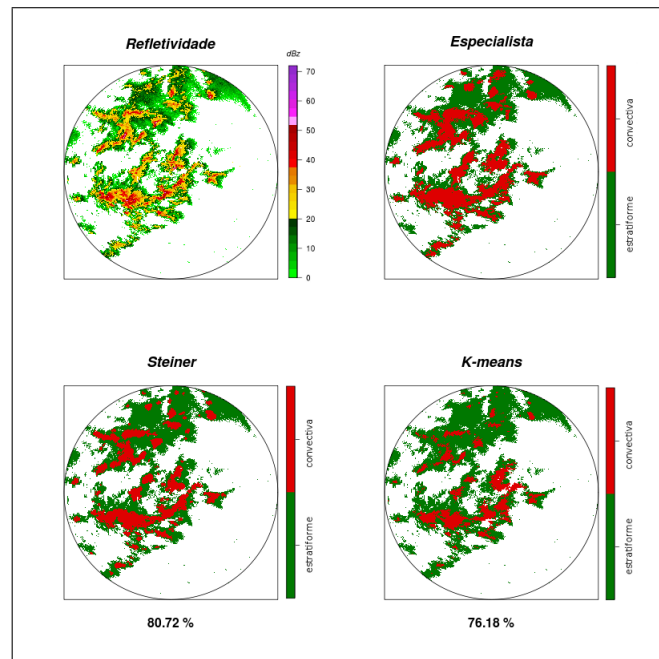
Fonte: A autora (2011)

Figura 6.21: Resultados obtidos pelo método *K-means* e comparação com a técnica de Steiner *et al.* para conjunto de dados 6



Fonte: A autora (2011)

Figura 6.22: Resultados obtidos pelo método *K-means* e comparação com a técnica de Steiner *et al.* para conjunto de dados 7



Fonte: A autora (2011)

Considerando a classificação do especialista como padrão, o índice de acerto obtido pelo método *K-means* é inferior ao obtido pelo método de Steiner *et al.* para todas as imagens, exceto para a imagem 5.

Novos dados podem ser designados a cada um dos grupos sem que haja a necessidade de rodar novamente o algoritmo. Basta designar cada um dos pontos ao grupo cujo centróide seja mais próximo.

Na seção a seguir, é feita uma comparação entre as duas técnicas aplicadas no trabalho, assim como entre as classificações obtidas pelo especialista e pelo método de Steiner *et al.* (STEINER; HOUZE; YUTER, 1995).

6.3 Comparação entre os Métodos

Quando comparados com a classificação feita pelo especialista, os resultados obtidos pelo *K-means* possuem menores índices de acerto. O que é compreensível, já que o SVM foi treinado com esses dados pré-classificados, e, portanto, espera-se que obtenha maior semelhança com essa identificação feita pelo especialista.

Foram calculados também os índices de acerto com relação ao método de Steiner *et al.* (STEINER; HOUZE; YUTER, 1995), bem como entre os dois métodos apresentados (SVM e *K-means*). A Tabela 6.6 apresenta, assim, todos esses índices de acerto calculados. A coluna Steiner - Especialista apresenta o índice de acerto do método Steiner *et al.* com relação ao especialista; a coluna *K-means* - Steiner representa o índice de acerto do método *K-means* em relação ao método Steiner *et al.*; e assim por diante.

Tabela 6.6: Índice de semelhança entre os dados classificados por especialista e métodos Steiner *et al.*, SVM e *K-means*

Conjunto de Dados	Steiner - Especialista	SVM - Especialista	K-means - Especialista	SVM - Steiner	K-means - Steiner	SVM - K-means
CJ1	86.82%	89.20%	83.28%	80.64%	81.25%	87.17%
CJ2	85.41%	89.43%	83.35%	87.59%	90.14%	89.63%
CJ3	78.39%	84.70%	71.76%	83.47%	84.25%	81.39%
CJ4	75.28%	77.06%	73.61%	80.12%	82.33%	80.46%
CJ5	76.33%	76.78%	77.49%	80.01%	81.33%	88.35%
CJ6	77.58%	83.65%	82.19%	76.96%	80.81%	83.88%
CJ7	80.72%	81.68%	76.18%	84.20%	82.87%	87.59%
TOTAL	80.08%	83.21%	78.27%	81.86%	83.28%	85.50%

Fonte: A autora (2011)

Pode-se observar que, embora o *K-means* tenha obtido resultados inferiores ao SVM quando comparados com o especialista, a técnica obtém resultados melhores que o SVM quando comparados com Steiner *et al.*

Quando a comparação dos resultados é feita entre ambas as técnicas (última coluna da Tabela 6.6), tem-se um índice de semelhança de 85.5%. Ou seja, mais de 85% dos dados foram classificados igualmente pelas duas metodologias.

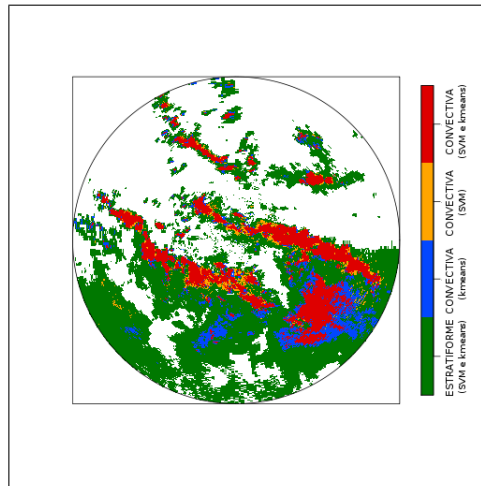
Essa semelhança entre as duas classificações, e as suas diferenças podem ser observadas nas imagens a seguir.

Para cada amostra foram observadas as seguintes possibilidades:

1. ambos os métodos classificaram o dado como convectivo (vermelho);
2. apenas o SVM classificou o dado como convectivo (laranja);

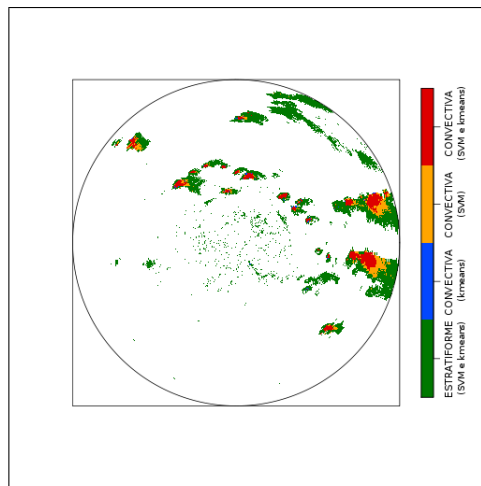
3. apenas o *K-means* classificou o dado como convectivo (azul);
4. ambos os métodos classificaram o dado como estratiforme (verde).

Figura 6.23: Imagem comparativa entre a classificação feita por SVM e por *K-means*, para conjunto de dados 1



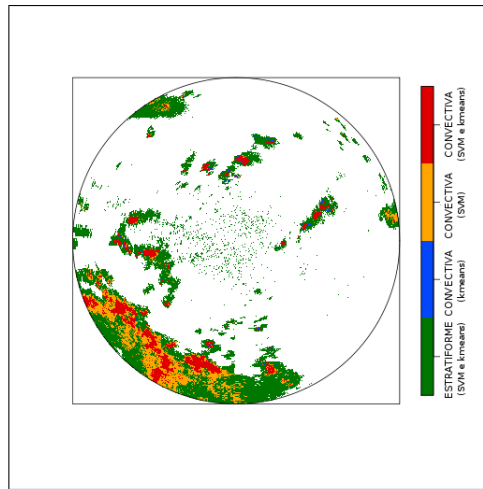
Fonte: A autora (2011)

Figura 6.24: Imagem comparativa entre a classificação feita por SVM e por *K-means*, para conjunto de dados 2



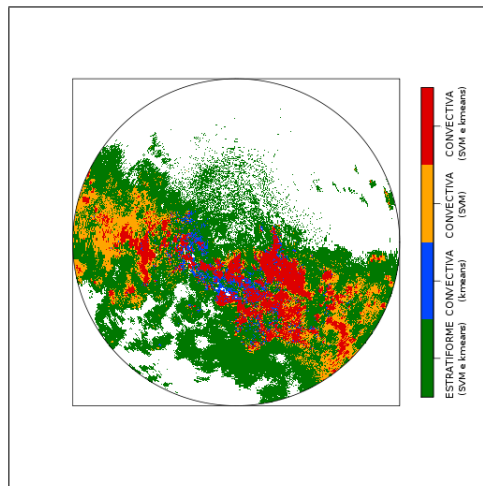
Fonte: A autora (2011)

Figura 6.25: Imagem comparativa entre a classificação feita por SVM e por *K-means*, para conjunto de dados 3



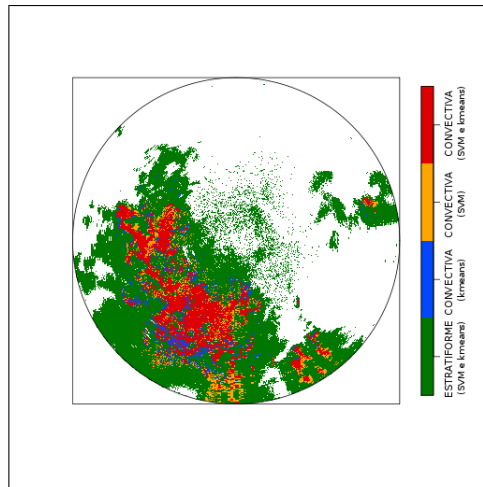
Fonte: A autora (2011)

Figura 6.26: Imagem comparativa entre a classificação feita por SVM e por *K-means*, para conjunto de dados 4



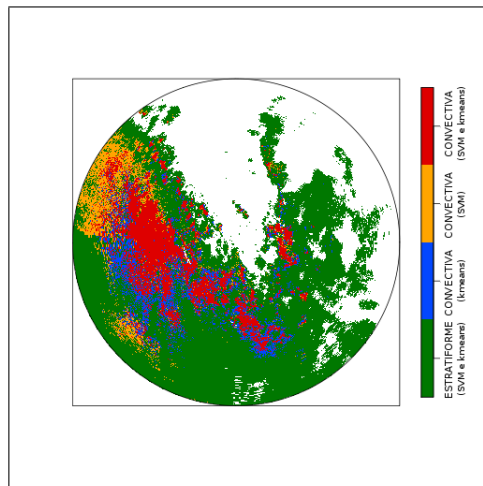
Fonte: A autora (2011)

Figura 6.27: Imagem comparativa entre a classificação feita por SVM e por *K-means*, para conjunto de dados 5



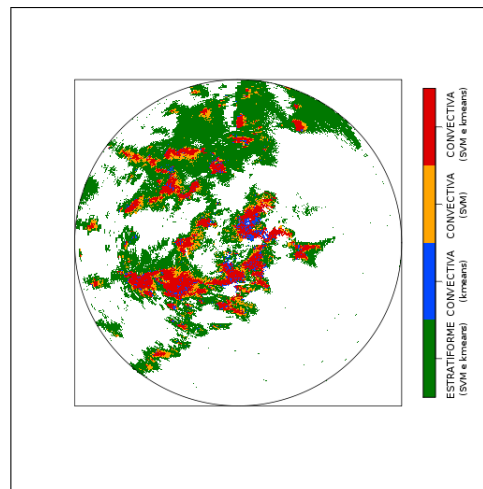
Fonte: A autora (2011)

Figura 6.28: Imagem comparativa entre a classificação feita por SVM e por *K-means*, para conjunto de dados 6



Fonte: A autora (2011)

Figura 6.29: Imagem comparativa entre a classificação feita por SVM e por *K-means*, para conjunto de dados 7



Fonte: A autora (2011)

Nota-se que, em geral, a diferença de identificação de chuva convectiva entre os dois métodos ocorre quando o método SVM classifica o *pixel* em chuva convectiva e o *K-means*, não. Uma explicação para esse fato pode ser dada pelo fato de o SVM ter feito treinamento da rede com dados classificados por um meteorologista. Além dessa classificação ser muito subjetiva, ela é feita na imagem como um todo, ou seja, não há muita precisão nos limites das regiões classificadas como convectivas. Já a técnica *K-means* analisa *pixel* por *pixel*, observando toda as características do dados, seu perfil vertical, e, portanto, tem uma maior precisão.

7 Conclusão

O monitoramento da chuva é fundamental para diversos setores econômicos, como agricultura, turismo, energia, dentre outros. As precipitações podem ser classificadas em estratiformes ou convectivas. Os sistemas convectivos estão associados a tempestades severas, e a sua identificação torna possível o seu acompanhamento e os alertas de temporais e enchentes.

Baseado nessa ideia, o presente trabalho apresenta duas metodologias para classificação da chuva: o SVM e o *K-means*. Ambos foram aplicados e testados no *software* R.

O método SVM é uma técnica de reconhecimento de padrões, cuja teoria é bem fundamentada. Possui boa capacidade de generalização e é robusta diante de dados de grande dimensão. Além disso, a convexidade do modelo garante a existência de uma solução global. Dados não linearmente separáveis podem ser mapeados em um espaço de maior dimensão, onde, então, podem ser separados linearmente. Isso é feito através da função *kernel*. Porém, a escolha dessa função, bem como de seus parâmetros, não é uma tarefa trivial, o que pode tornar difícil a escolha de um *kernel* adequado à aplicação.

Outra característica do método SVM é que dados previamente classificados são utilizados para treinamento da máquina. Como as técnicas de Inteligência Artificial (na qual inclui-se o SVM) visam automatizar tarefas realizadas pela ação humana, optou-se por usar como padrão imagens classificadas por um especialista na área, ou seja, um meteorologista. Deve-se, entretanto, salientar que essa classificação manual é muito subjetiva. E, portanto, isso acaba sendo, para o problema apresentado, uma desvantagem do método.

Mesmo com essas limitações, o método SVM atendeu bem aos objetivos do trabalho, ao conseguir identificar padrões meteorológicos em imagens de radar, e assim, classificar a chuva representada nestas imagens em estratiforme ou convectiva de forma satisfatória, de acordo com o padrão a ser alcançado (classificação do especialista).

Para observar se a técnica sugerida apresentou bons resultados, estes foram comparados com os obtidos por uma técnica de classificação de chuva já existente na literatura, o método de Steiner *et al.* (STEINER; HOUZE; YUTER, 1995). Comparando-se a porcentagem de

amostras classificadas corretamente de acordo com a classificação feita por um especialista, o método SVM apresentou taxas superiores para todas as imagens.

O segundo método aplicado foi o *K-means*, uma técnica de análise de agrupamentos, sub-área da análise multivariada. A grande vantagem em relação ao SVM é que não precisa de dados previamente classificados. A técnica procura um agrupamento “natural” dos dados, colocando amostras com características semelhantes em um mesmo grupo. Outra vantagem é em relação ao tempo de execução, que não passa de 1 segundo, enquanto o SVM leva muitos minutos para gerar um classificador.

Para o problema apresentado, o *K-means* foi aplicado para agrupar os dados em dois grupos, esperando-se obter um grupo de dados com características de chuva estratiforme, e outro, com características de precipitação convectiva.

O método atendeu bem esse objetivo, e, apesar de ter apresentado índices menores que os resultados obtidos com o SVM, nota-se que as imagens obtidas são muito parecidas com a técnica de Steiner *et al.* Vale lembrar que esse índice de acerto é em relação à classificação do especialista e, como o SVM foi treinado com esses dados, é de se esperar que apresente resultados mais semelhantes à esta classificação do que o *K-means*.

Uma terceira etapa do trabalho mostra uma comparação entre as duas metodologias. As possibilidades apresentadas são quatro:

- ambos os métodos classificaram o dado como convectivo;
- apenas o SVM classificou o dado como convectivo;
- apenas o *K-means* classificou o dado como convectivo;
- ambos os métodos classificaram o dado como estratiforme.

Observou-se que, em geral, a diferença entre as duas metodologias ocorre quando o SVM classifica o dado como convectivo e o *K-means*, não. Isso pode ser explicado pelo fato de o SVM ser treinado com dados classificados manualmente por um especialista. Dessa forma, pode não haver tanta precisão no espaço delimitador de uma região convectiva, e dados a mais são considerados convectivos também. O *K-means* já é baseado na análise pixel a pixel, e portanto, pode ter uma precisão maior.

As duas técnicas apresentaram a mesma classificação para mais de 85% dos dados.

Além disso, ao relacionar modelos matemáticos de Otimização, técnica de Análise

Multivariada e Meteorologia, o trabalho contribui para o desenvolvimento de outras áreas do conhecimento.

7.1 Sugestões para trabalhos futuros

A seguir são descritas algumas sugestões para possíveis trabalhos futuros:

- Levar em conta a classificação feita por diversos especialistas na área (e não apenas um), e encontrar assim, uma classificação “média”, talvez mais ideal, já que essa classificação é muito subjetiva;
- Aplicar as relações Z-R para as novas classificações, e verificar se os resultados são compatíveis com a taxa de precipitação medida.
- Utilizar outras técnicas para classificar os dados, como por exemplo, uma Rede Neural.

Referências Bibliográficas

- ALES, V. T. *O Algoritmo Sequential Minimal Optimisation para Resolução do Problema de Support Vector Machine: Uma Técnica Para Reconhecimento de Padrões*. Dissertação (Mestrado) — Universidade Federal do Paraná, Curitiba - PR, 2008.
- ALMEIDA, F. F. M. Support Vector Machine. *Universidade Federal de Campina Grande - Centro de Ciências e Tecnologia - Departamento de Sistemas e Computação*, 2007.
- AMBURN, S. A.; WOLF, P. L. VIL Density as a Hail Indicator. *Weather and Forecasting*, v. 12, p. 473 – 478, 1997.
- ANAGNOSTOU, E. N.; KUMMEROW, C. Stratiform and Convective Classification of Rainfall using SSM/I-GHZ Brightness Temperature Observations. *Journal of Atmospheric and Oceanic Technology*, v. 14, p. 570–575, 1997.
- BEAN, B. R.; DUTTON, E. J. *Radio Meteorology*. [S.l.]: Dover Publications, 1966. 424 p.
- BELTRAMI, M. *Precificação de Opções Sobre Ações por Modelos de Support Vector Vector Regression*. Dissertação (Mestrado) — Universidade Federal do Paraná, Curitiba - PR, 2009.
- BENETI, C. A. A.; NOZU, I.; SARAIVA, E. A. Monitoramento da Precipitação e de Eventos de Tempo Severo com Radar Meteorológico no Estado do Paraná. Brasília, 1998. Congresso Brasileiro de Meteorologia.
- BERGAMINI, C. M.; ARAUJO, P. V.; MOTTER, G. Modelos de Aprendizagem de Máquina na Classificação de Caracteres Manuscritos. *Synergismus Scyentifica UTFPR - Pato Branco*, p. 338–348, 2006.
- BIGGERSTAFF, M. I.; LISTEMAA, S. A. An Improved Scheme for Convective/Stratiform Echo Classification Using Radar Reflectivity. *Journal of Applied Meteorology*, v. 39, p. 2129–2150, 2000.
- BORGES, J. da S. *Classificadores e Máquinas de Suporte Vectorial em Imagens Multi-Espectrais*. Dissertação (Mestrado) — Faculdade de Ciências da Universidade do Porto, 2003.
- BURGES, C. J. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 1998.
- CARVALHO, B. P. R. de. O Estado da Arte em Métodos para o Reconhecimento de Padrões: Support Vector Machine. *Congresso Nacional de Tecnologia da Informação e Comunicação*, Belo Horizonte-MG, 2005.
- CHANG, C. C.; LIN, C. L. LIBSVM: a Library for Support Vector Machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. 2001.

- CHURCHILL, D. D.; HOUZE, R. A. Development and Structure of Winter Mosoon Cloud Clusters on 10 December 1978. *Journal of Atmospheric Sciences*, v. 41, p. 933–960, 1984.
- COLLIER, C. G. *Applications of Weather Radar Systems: a Guide to Uses of Radar Data in Meteorology and Hydrology*. 2th. ed. [S.l.]: John Wiley & Sons, 1996.
- CRISTIANINI, N.; SHAWE-TAYLOR, J. *An introduction to Support Vector Machines and other kernel-based learning methods*. [S.l.]: Cambridge University Press, 2000.
- DEFILIPPO, S. B. *Máquinas de Vetores Suporte*. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, Juiz de Fora - MG, 2004.
- DEMOTT, C. A.; CIFFELI, R.; RUTLEDGE, S. A. An Improved Method for Partitioning Radar Data into Convective and Stratiform Components. *27th Conference on Radar Meteorology*, 1995.
- FERREIRA, P. L. *Estatística Multivariada Aplicada*. Faculdade de Economia da Universidade de Coimbra, 2000.
- FILHO, A. J. P.; CRAWFORD, K. C. Mesoscale Precipitation Fields, Part I: Statistical Analysis and Hydrologic Response. *Journal of Applied Meteorology*, v. 38, p. 82–101, 1999.
- GOURLEY, J. J. et al. Evaluation of Incremental Improvements to Quantitative Precipitation Estimates in Complex Terrain. *Journal of Hydrometeorology*, v. 10, p. 1507–1520, 2009.
- HAIR, J. F. et al. *Análise Multivariada dos Dados*. 6^a. ed. [S.l.]: Bookman, 2009.
- HERBRICH, R. *Learning Kernel Classifiers: Theory and Algorithms*. MIT Press, 2002.
- ILLINOIS. University of Illinois WW2010 Project. Disponível em: [http://ww2010.atmos.uiuc.edu/\(Gh\)/guides/rs/rad/basics/sgnl.rxml](http://ww2010.atmos.uiuc.edu/(Gh)/guides/rs/rad/basics/sgnl.rxml). Acesso em: 22/11/2010. 2010.
- IZMAILOV, A.; SOLODOV, M. *Otimização - Volume 1. Condições de Otimalidade, Elementos de Análise Convexa e Dualidade*. Rio de Janeiro: IMPA - Instituto Nacional de Matemática Pura e Aplicada, 2005.
- LIMA, C. A. M. *Comitê de Máquinas: Uma Abordagem Unificada Empregando Máquinas de Vetores-Suporte*. Tese (Doutorado) — Universidade Federal de Campinas, Campinas - SP, 2004.
- LORENA, A. C.; CARVALHO, A. C. P. L. F. de. Uma Introdução às Support Vector Machines. *Revista de Informática Teórica e Aplicada*, v. 14, p. 43, 2007.
- MARSHALL, J. S.; PALMER, W. M. K. The Distribution of Raindrops with Size. *Journal of Meteorology*, 5, p. 165–166, 1948.
- MASSAMBANI, O. *Fundamentos do Radar*. São Paulo - SP, 2009. Disponível em <http://www.dca.iag.usp.br/www/material/omassamb/radar> .
- METEOPT. Fórum de Meteorologia. Disponível em: <http://www.meteopt.com/forum/instrumentos-meteorologicos/radar-meteorologico-da-madeira-4482-2.html>. Acesso em: 14/04/2011. 2011.

- MEYER, D. *Support Vector Machines - The Interface to libsvm in package e1071*. Technische Universität Wien, Austria, 2011.
- MOL, J. M. D. *Estimativa de Precipitação por Meio de Sensoriamento Remoto*. Dissertação (Mestrado) — Universidade de Brasília, 2005.
- QUEENSLAND, T. U. of. School of Information Technology and Electrical Engineerin. Disponível em: <http://itee.uq.edu.au/comp4702/material.html>. Acesso em: 28/06/2011. 2011.
- R-PROJECT. R version 2.9.2 (2009-08-24). Disponível em: <http://www.r-project.org>. 2009.
- RINEHART, R. E. *Radar For Meteorologists*. 4th. ed. [S.l.]: Rinehart Publishing, 2004.
- RODRIGUES, F. S. *Métodos de Agrupamento na Análise de Dados de Expressão Gênica*. Dissertação (Mestrado) — Universidade Federal de São Carlos, São Carlos - SP, 2009.
- SANTOS, E. M. dos. *Teoria e Aplicação de Support Vector Machines à Aprendizagem e Reconhecimento de Objetos Baseados na Aparência*. Dissertação (Mestrado) — Universidade Federal da Paraíba, 2002.
- SAUVAGEOT, H. Rainfall Measurement by Radar: A Review. *Atmospheric Research*, v. 35, p. 27–57, 1994.
- SEKHON, R. S.; SRIVASTAVA, R. C. Snow Size Spectra and Radar Reflectivity. *Journal of Atmospheric Science*, 27, p. 299–307, 1970.
- SEKHON, R. S.; SRIVASTAVA, R. C. Doppler Radar Observations of Drop-Size Distribution in a Thunderstorm. *Journal of Atmospheric Science*, 28, p. 983–994, 1971.
- SEMOLINI, R. *Support Vector Machines, Inferência Transdutiva e o Problema de Classificação*. Brasil, Dezembro 2002.
- SILVA, T. de S. *Aplicação do Método Amostragem Seletiva Global na Construção da Árvore de Cenários de Afluências Utilizada no Planejamento da Operação Energética de Médio Prazo*. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, Rio de Janeiro - RJ, 2010.
- SIMPSON, J.; ADLER, R. F.; NORTH, G. R. A Proposed Tropical Rainfall Measuring Mission (TRMM) Satellite. *Bulletin of the American Meteorological Society*, v. 69, p. 278–295, 1988.
- SMOLA, A. J.; SCHÖLKOPF, B. *Learning with Kernels*. The MIT Press, Cambridge-MA, 2002.
- STEINER, M.; HOUZE, R. A.; YUTER, S. E. Climatological Characterization of Three-Dimensional Storm Structure from Operational Radar and Rain Gauge Data. *Journal of Applied Meteorology*, v. 34, p. 1978–2007, 1995.
- VAPNIK, V. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.
- VAPNIK, V. *Statistical Learning Theory*. New York: John Wiley and Sons, 1998.
- VAPNIK, V.; CHERVONENKIS, A. *On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities*. [S.l.]: Theoretical Probability and Its Applications, 1971. 264-280 p.

ZHANG, J. et al. Three-Dimensional Gridding and Mosaic of Reflectivities from Multiple WRS-88D Radars. *The 30th International Conference on Radar Meteorology*, v. 22, p. 719–721, July 2001.

ZHANG, J.; HOWARD, K.; GOURLEY, J. J. Constructing Three-Dimensional Multiple-Radar Reflectivity Mosaics: Examples of Convective Storms and Stratiform Rain Echoes. *Journal of Atmospheric and Oceanic Technology*, v. 22, p. 30–42, 2005.

ZHANG, J.; LANGSON, C.; HOWARD, K. Brightband Identification Based on Vertical Profiles of Reflectivity from the WSR-88D. *Journal of Atmospheric and Oceanic Technology*, v. 25, p. 1859–1872, 2008.