

*André C. P. L. F. de Carvalho*  
**Inteligência Artificial**

Uma Abordagem de  
Aprendizado de Máquina



**LTC**  
EDITORA



[www.grupogen.com.br](http://www.grupogen.com.br)

<http://gen-io.grupogen.com.br>



Saúde



ROCA



Jurídico



Exatas

LTC  
EDITORA

Humanas



O **GEN | Grupo Editorial Nacional** reúne as editoras Guanabara Koogan, Santos, Roca, AC Farmacêutica, LTC, Forense, Método, E.P.U. e Forense Universitária



O **GEN-IO | GEN – Informação Online** é o repositório de material suplementar dos livros dessas editoras

[www.grupogen.com.br](http://www.grupogen.com.br)

<http://gen-io.grupogen.com.br>



## Capítulo 1 INTRODUÇÃO

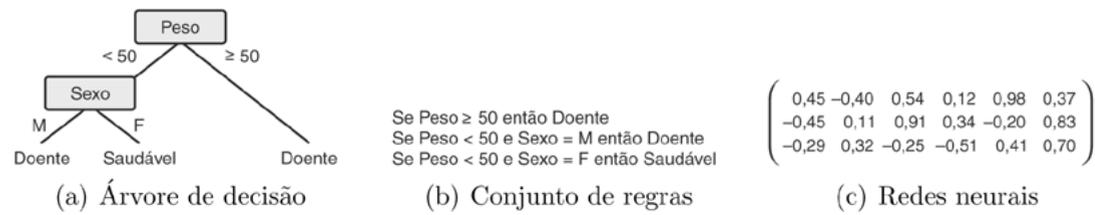


Figura 1.1 Diferentes vieses de representação.



Figura 1.2 Hierarquia de aprendizado.

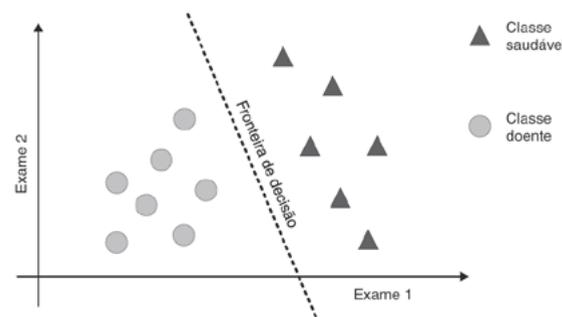


Figura 1.3 Espaço de objetos.



## Capítulo 2 ANÁLISE DE DADOS

Tabela 2.1 Conjunto de dados hospital com seus atributos

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	72	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Tabela 2.2 Tipo dos atributos do conjunto hospital

Atributo	Classificação
Id.	Qualitativo
Nome	Qualitativo
Idade	Quantitativo discreto
Sexo	Qualitativo
Peso	Quantitativo contínuo
Manchas	Qualitativo
Temp.	Quantitativo contínuo
#Int.	Quantitativo discreto
Est.	Qualitativo
Diagnóstico	Qualitativo

Tabela 2.3 Escala dos atributos do conjunto hospital

Atributo	Classificação
Id.	Nominal
Nome	Nominal
Idade	Racional
Sexo	Nominal
Peso	Racional
Manchas	Nominal
Temp.	Intervalar
#Int.	Racional
Est.	Nominal
Diagnóstico	Nominal

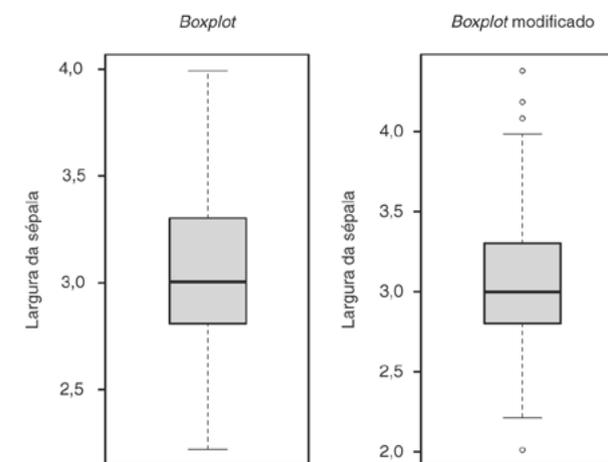


Figura 2.1 Boxplots para o atributo largura da sépala do conjunto de dados iris.

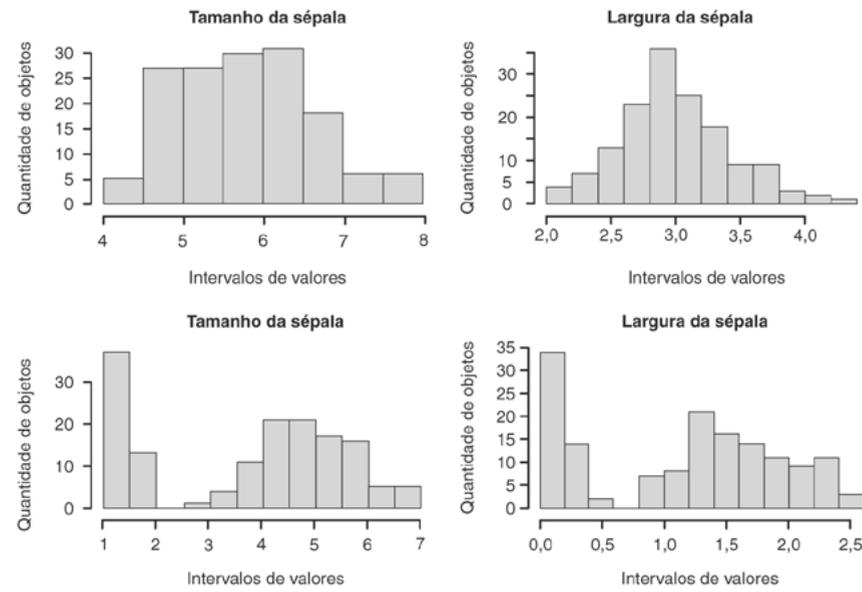


Figura 2.2 Histograma para a distribuição de valores dos atributos de entrada do conjunto iris.

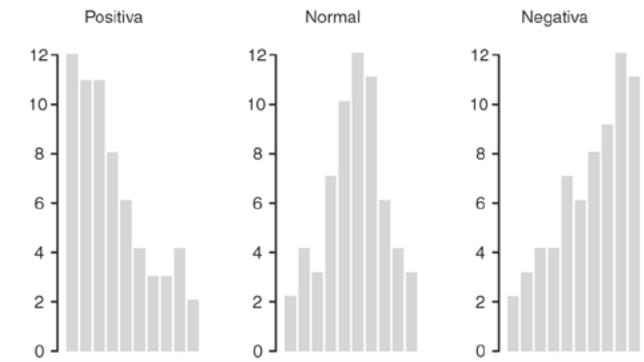


Figura 2.3 Distribuição dos valores de obliquidade.

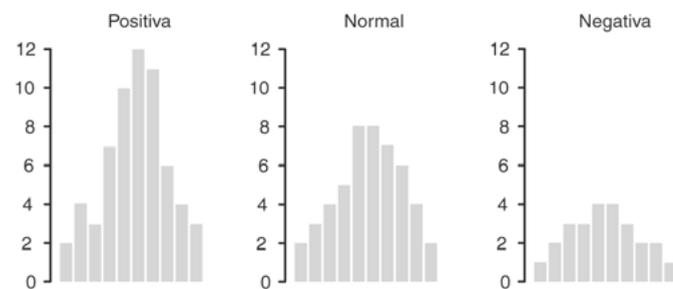


Figura 2.4 Distribuição dos valores de curtose.

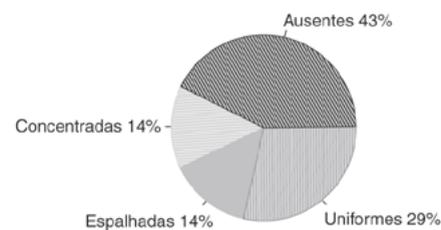


Figura 2.5 Gráfico de pizza para a distribuição de valores do atributo Manchas.

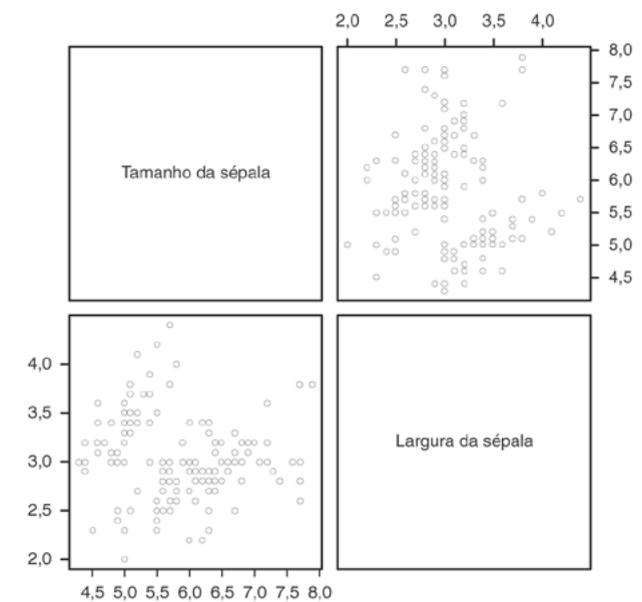


Figura 2.6 Matriz de scatter plot para dois atributos do conjunto de dados iris.

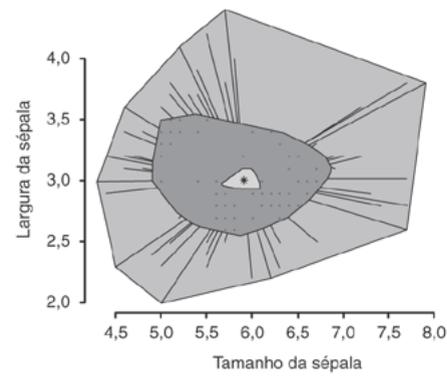


Figura 2.7 Diagramas de bagplot para dois atributos do conjunto de dados iris.

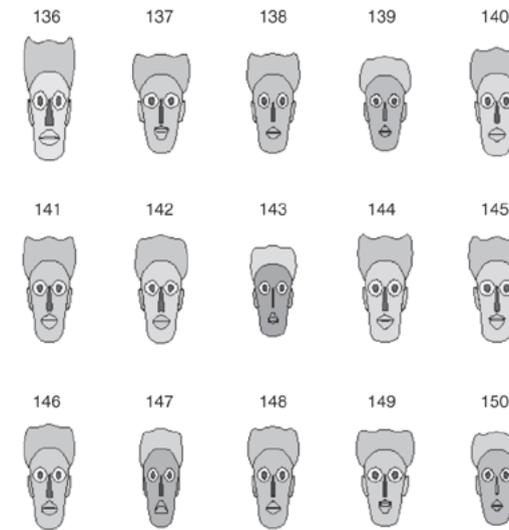


Figura 2.8 Diagramas de Chernoff para 15 objetos do conjunto de dados iris.

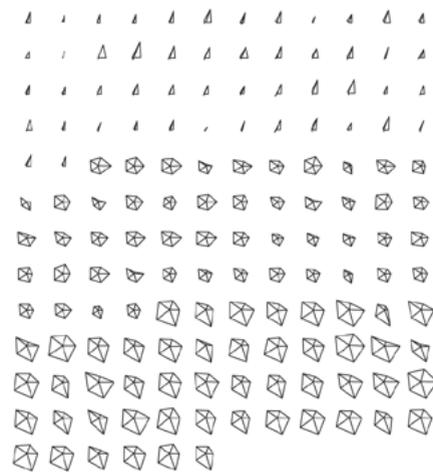


Figura 2.9 Star plots para os 15 objetos do conjunto de dados iris.

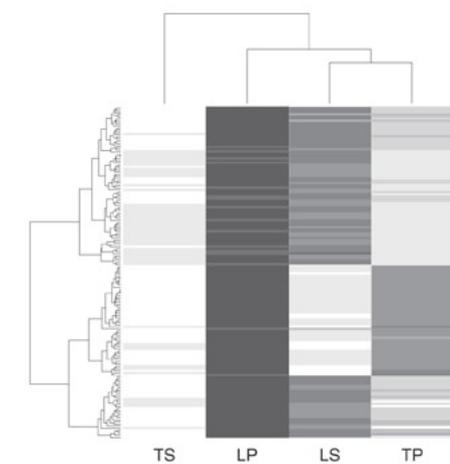


Figura 2.10 Heatmap para atributos de entrada do conjunto de dados iris.



## Capítulo 3 PRÉ-PROCESSAMENTO DE DADOS

Tabela 3.1 Conjunto de dados sem atributos considerados irrelevantes

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Concentradas	38,0	2	Doente
18	F	67	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	M	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
19	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Tabela 3.2 Conjunto de dados com atributos com valores ausentes

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
—	M	79	—	38,0	—	Doente
18	F	67	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	—	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
—	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Tabela 3.3 Conjunto de dados com substituição dos valores ausentes

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
<b>27</b>	M	79	<b>Inexistentes</b>	38,0	4	Doente
18	F	67	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	<b>F</b>	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
<b>27</b>	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Tabela 3.4 Conjunto de dados com objetos inconsistentes

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Concentradas	38,0	2	Doente
18	F	67	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	M	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
<b>22</b>	<b>F</b>	<b>72</b>	<b>Inexistentes</b>	<b>38,0</b>	<b>3</b>	<b>Doente</b>
19	F	87	Espalhadas	39,0	6	Doente
<b>22</b>	<b>F</b>	<b>72</b>	<b>Inexistentes</b>	<b>38,0</b>	<b>3</b>	<b>Saudável</b>

Tabela 3.5 Conjunto de dados com objetos redundantes

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Concentradas	38,0	2	Doente
<b>18</b>	<b>F</b>	<b>67</b>	<b>Inexistentes</b>	<b>39,5</b>	<b>4</b>	<b>Doente</b>
49	M	92	Espalhadas	38,0	2	Saudável
<b>18</b>	<b>F</b>	<b>67</b>	<b>Inexistentes</b>	<b>39,5</b>	<b>4</b>	<b>Doente</b>
18	M	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
19	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Tabela 3.6 Conjunto de dados com atributos redundantes

Idade	Sexo	Peso	Manchas	Temp.	# Int.	# Vis.	Diagnóstico
28	M	79	Concentradas	38,0	2	<b>2</b>	Doente
18	F	67	Inexistentes	39,5	4	<b>4</b>	Doente
49	M	92	Espalhadas	38,0	2	<b>2</b>	Saudável
18	M	43	Inexistentes	38,5	8	<b>8</b>	Doente
21	F	52	Uniformes	37,6	1	<b>1</b>	Saudável
22	F	72	Inexistentes	38,0	3	<b>3</b>	Doente
19	F	87	Espalhadas	39,0	6	<b>6</b>	Doente
34	M	67	Uniformes	38,4	2	<b>2</b>	Saudável



Tabela 3.7 Conjunto de dados com ruído

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Concentradas	38,0	2	Doente
18	F	300	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	M	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
19	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Tabela 3.8 Codificação 1 – de – c

Atributo nominal	Código 1 – de – c
Azul	100000
Amarelo	010000
Verde	001000
Preto	000100
Marrom	000010
Branco	000001

Tabela 3.9 Pseudoatributos e seus possíveis valores

Pseudoatributo	#Valores
Continente	7 (b)
PIB	1 (i)
População	1 (i)
TMA	1 (i)
Área	1 (i)

Tabela 3.10 Conversão de valor ordinal para inteiro

Valor ordinal	Valor inteiro
Primeiro	0
Segundo	1
Terceiro	2
Quarto	3
Quinto	4
Sexto	5

Tabela 3.11 Conversão de valor ordinal para binário

Valor ordinal	Código cinza	Código termômetro
Primeiro	000	00000
Segundo	001	00001
Terceiro	011	00011
Quarto	010	00111
Quinto	110	01111
Sexto	100	11111

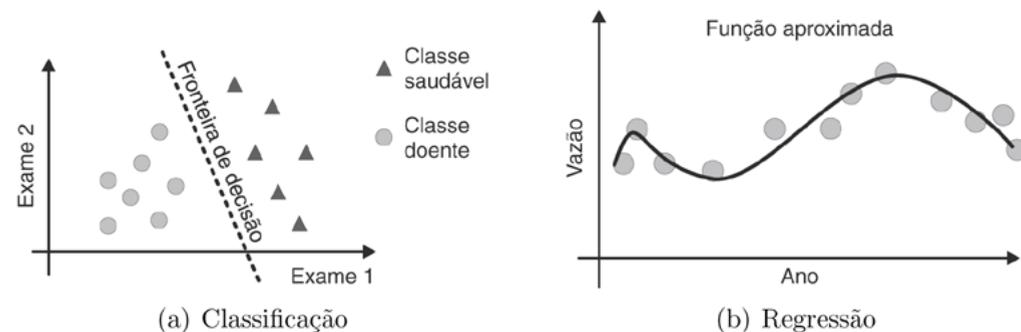


Figura 3.1 Gráfico ilustrativo das tarefas.

Tabela 3.12 Exemplo de conjunto de dados para problema de classificação

Tamanho (P)	Largura (P)	Tamanho (S)	Largura (S)	Espécie
5,1	3,5	1,4	0,2	Setosa
4,9	3,0	1,4	0,2	Setosa
7,0	3,2	4,7	1,4	Versicolor
6,4	3,2	4,5	1,5	Versicolor
6,3	3,3	6,0	2,5	Virgínica
5,8	2,7	5,1	1,9	Virgínica



Tabela 3.13 Exemplo de conjunto de dados para problema de regressão

Fertilidade	Agricultura	Educação	Renda	Mortalidade
80,2	17,0	12	9,9	22,2
83,1	45,1	9	84,8	22,2
92,5	39,7	5	93,4	20,2
85,8	36,5	7	33,7	20,3
76,9	43,5	15	5,2	20,6

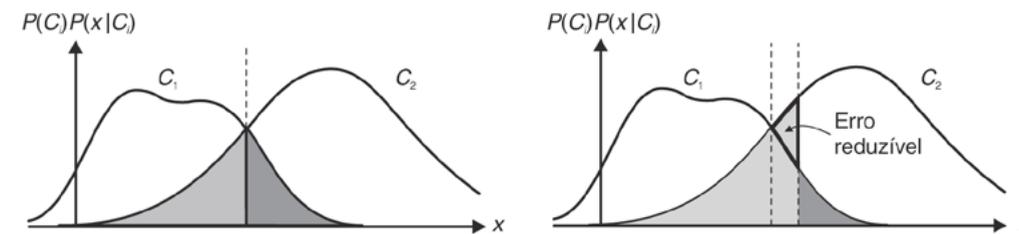


Figura 3.2 A figura da esquerda representa a superfície de decisão ótima. A área cinzenta corresponde à probabilidade de erro ao decidir por  $C_1$  quando a classe correta é  $C_2$ ; a área cinza-escura é o erro oposto. Se movemos a superfície de decisão, como se vê na figura da direita, o erro sempre aumenta.



## Capítulo 4 MÉTODOS BASEADOS EM DISTÂNCIAS

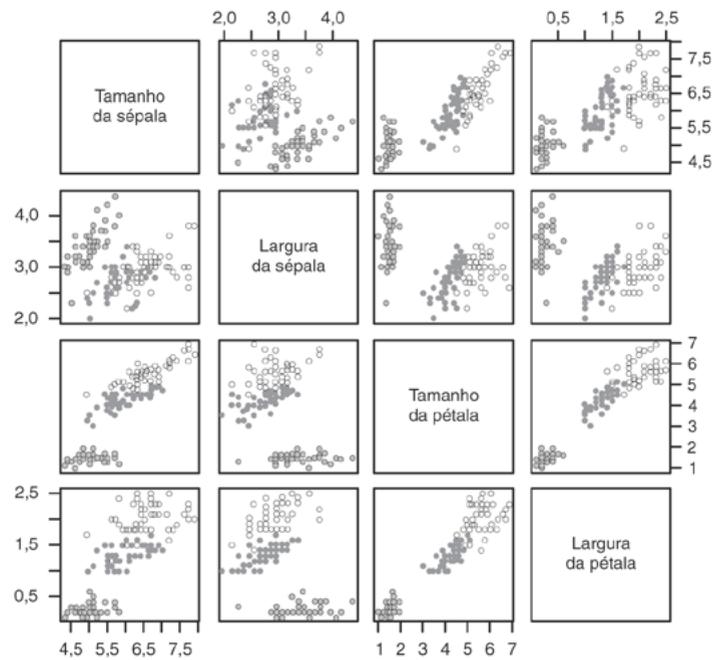


Figura 4.1 Projeção sobre o plano definido por dois atributos dos objetos do conjunto de dados iris.

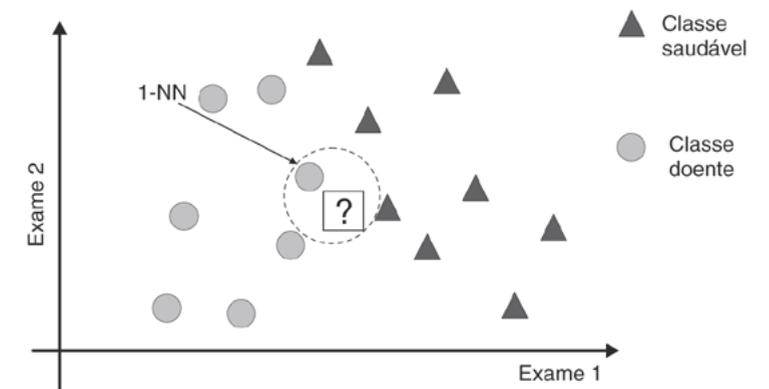


Figura 4.2 Exemplo ilustrativo do algoritmo 1-NN.

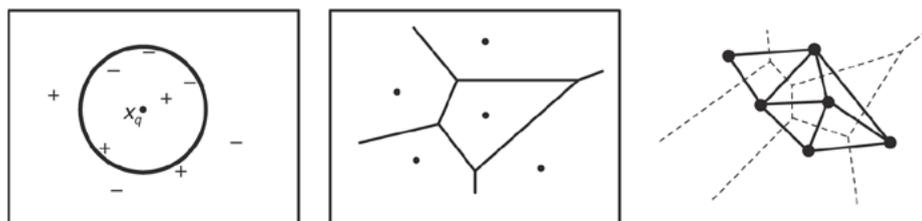


Figura 4.3 Superfície de decisão: diagrama de Voronoi.

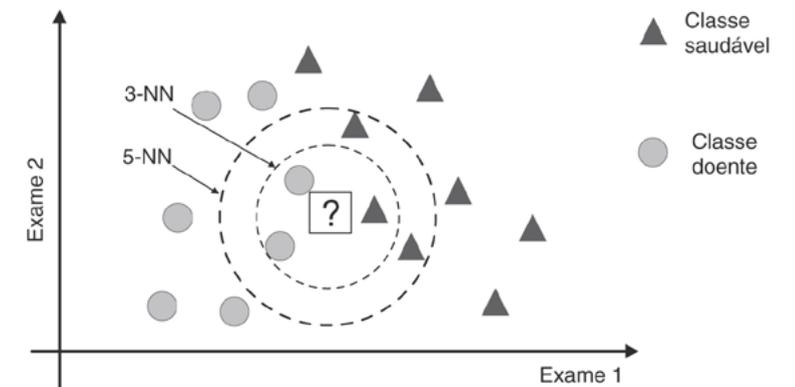


Figura 4.4 Impacto do valor de  $k$  no algoritmo  $k$ -NN.

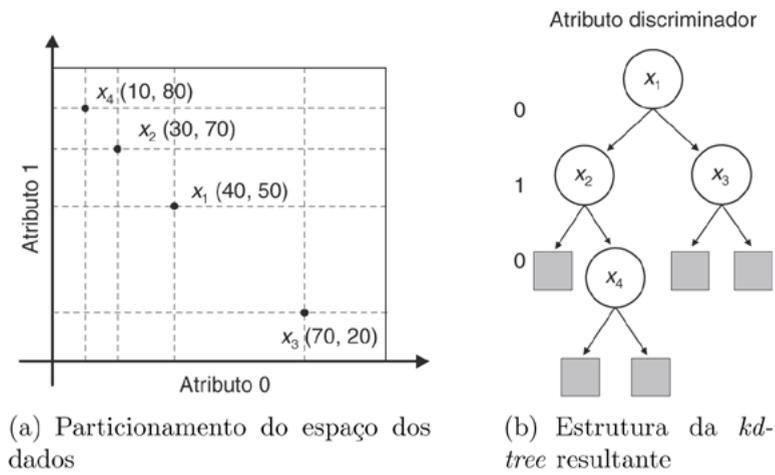


Figura 4.5 Os pontos mostrados em 4.5(a) identificam objetos armazenados como nós internos na *kd-tree* e são utilizados para particionar o espaço de busca.

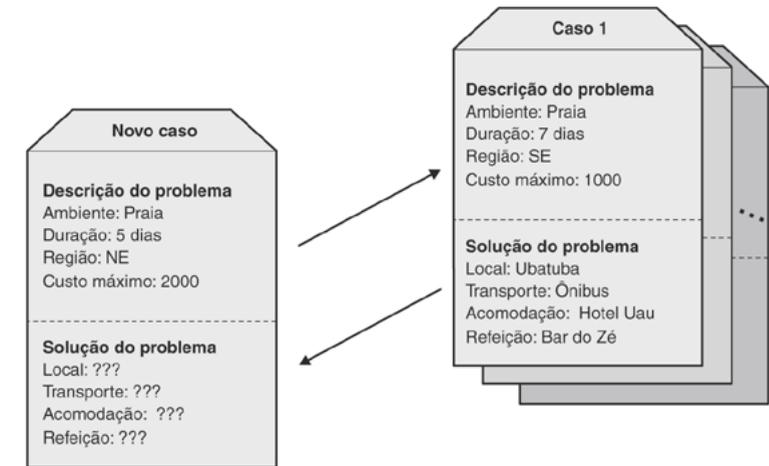


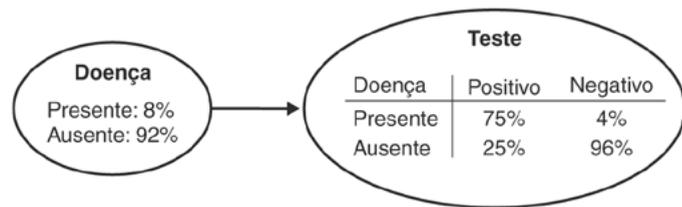
Figura 4.6 Exemplo de um novo caso e de uma BC.



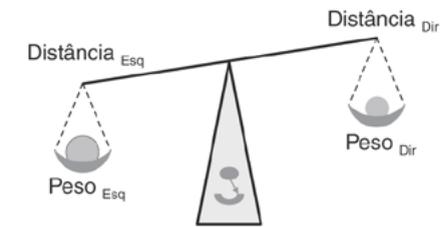
Figura 4.7 Ciclo de raciocínio baseado em casos (Aamodt e Plaza, 1994).



## Capítulo 5 MÉTODOS PROBABILÍSTICOS



**Figura 5.1** Modelo probabilístico gráfico para representar a informação no problema médico. A figura mostra o modelo qualitativo e o modelo quantitativo.



**Figura 5.2** O problema do equilíbrio da balança.

**Tabela 5.1** Contagem de valores e probabilidade a priori para as classes

	Equilibrada	Esquerda	Direita
Contagem	49	288	288
$P(\text{Classe})$	0,078	0,461	0,461

**Tabela 5.2** Tabelas de distribuição dos valores dos atributos por classe

	Distribuição normal		Discretização				
<b><math>Peso_{Esq}</math></b>	Média	Desvio padrão	V1	V2	V3	V4	V5
Equilibrado	2,938	1,42	10	11	9	10	9
Esquerda	3,611	1,23	17	43	63	77	88
Direita	2,399	1,33	98	71	53	38	28
<b><math>Distância_{Esq}</math></b>	Média	Desvio padrão	V1	V2	V3	V4	V5
Equilibrado	2,938	1,42	10	11	9	10	9
Esquerda	3,611	1,22	17	43	63	77	88
Direita	2,399	1,33	98	71	53	38	28
<b><math>Peso_{Dir}</math></b>	Média	Desvio padrão	V1	V2	V3	V4	V5
Equilibrado	2,938	1,42	10	11	9	10	9
Esquerda	2,399	1,33	71	53	38	28	17
Direita	3,611	1,22	17	43	63	77	88
<b><math>Distância_{Dir}</math></b>	Média	Desvio padrão	V1	V2	V3	V4	V5
Equilibrado	2,938	1,42	10	11	9	10	9
Esquerda	2,399	1,33	98	71	53	38	28
Direita	3,611	1,22	17	43	63	77	88

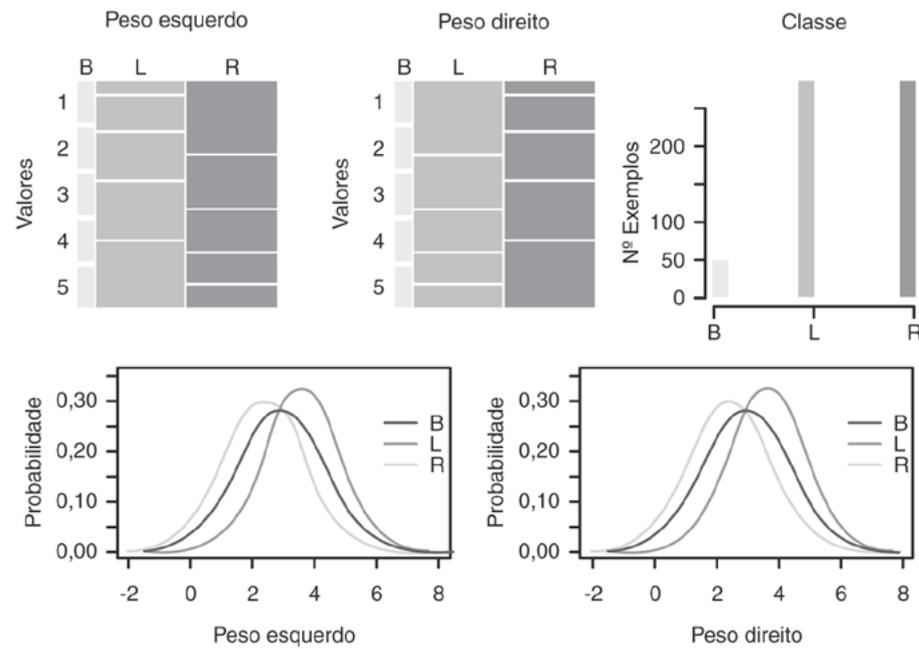
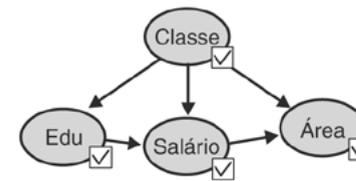


Figura 5.3 Naive Bayes para o problema da balança.



Classe	Não		Sim		
Edu	Não	Sim	Não	Sim	
Salário	Baixo	0,6	0,4	0,33	0,66
Alto	0,4	0,4	0,66	0,33	

Figura 5.4 A Figura mostra o modelo qualitativo – um grafo cujos nós representam variáveis – e o modelo quantitativo para a variável Salário – tabelas com a distribuição de probabilidades dos valores da variável Salário dado o valor das variáveis que a influenciam.

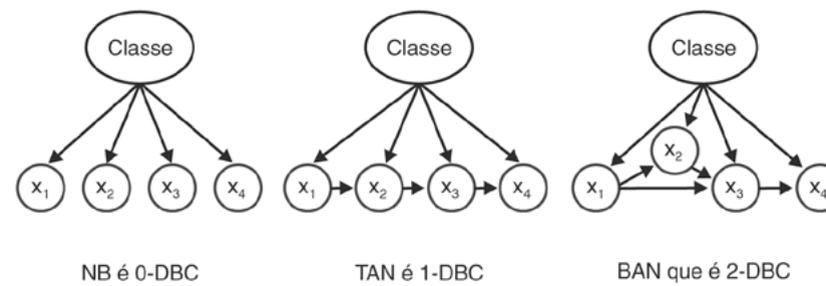


Figura 5.5 Exemplos de classificadores bayesianos com k-dependências.



## Capítulo 6 MÉTODOS BASEADOS EM PROCURA

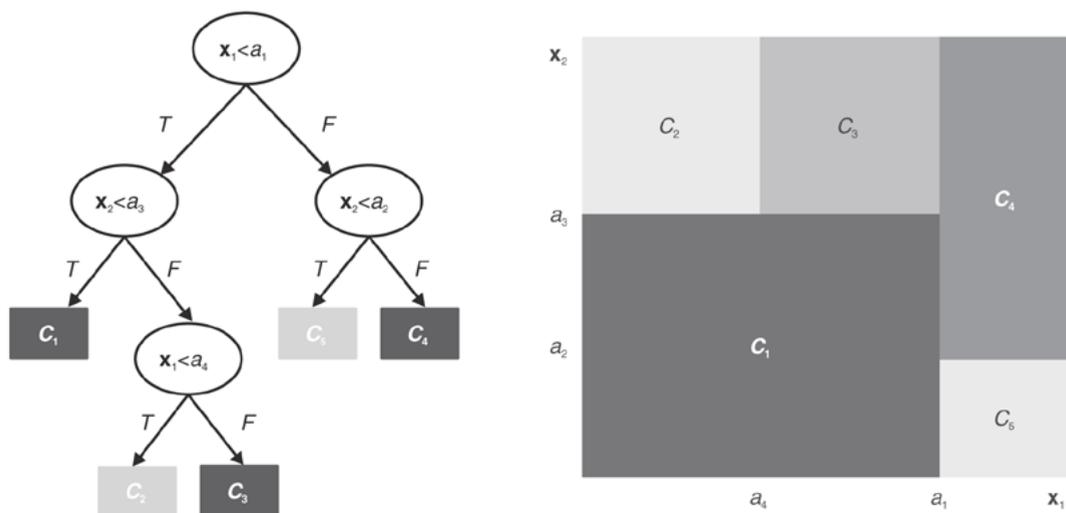


Figura 6.1 Uma árvore de decisão e as regiões de decisão no espaço de objetos.

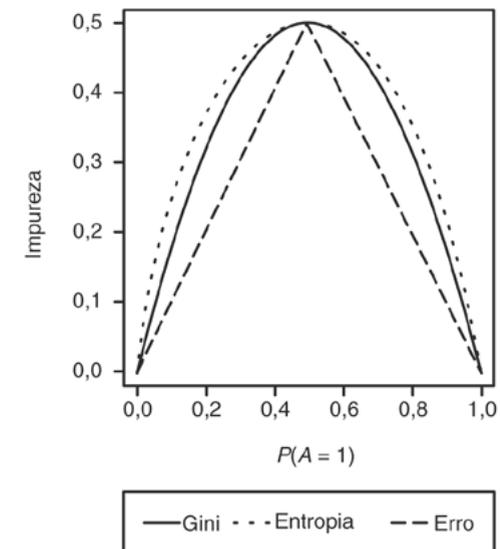


Figura 6.2 Gráfico da entropia, índice Gini e da taxa de erro de uma variável booleana aleatória A.

Tabela 6.1 Exemplo de conjunto de dados

Tempo	Temperatura	Umidade	Vento	Joga
Chuvoso	71	91	Sim	Não
Ensolarado	69	70	Não	Sim
Ensolarado	80	90	Sim	Não
Nublado	83	86	Não	Sim
Chuvoso	70	96	Não	Sim
Chuvoso	65	70	Sim	Não
Nublado	64	65	Sim	Sim
Nublado	72	90	Sim	Sim
Ensolarado	75	70	Sim	Sim
Chuvoso	68	80	Não	Sim
Nublado	81	75	Não	Sim
Ensolarado	85	85	Não	Não
Ensolarado	72	95	Não	Não
Chuvoso	75	80	Não	Sim

Tabela 6.2 Distribuição dos valores da classe pelos valores do atributo Tempo

	Ensolarado	Nublado	Chuvoso
Sim	2	4	3
Não	3	0	2

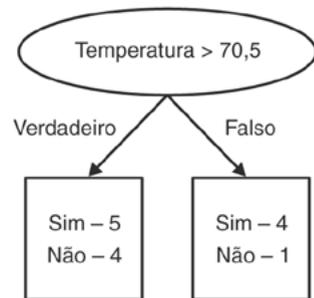


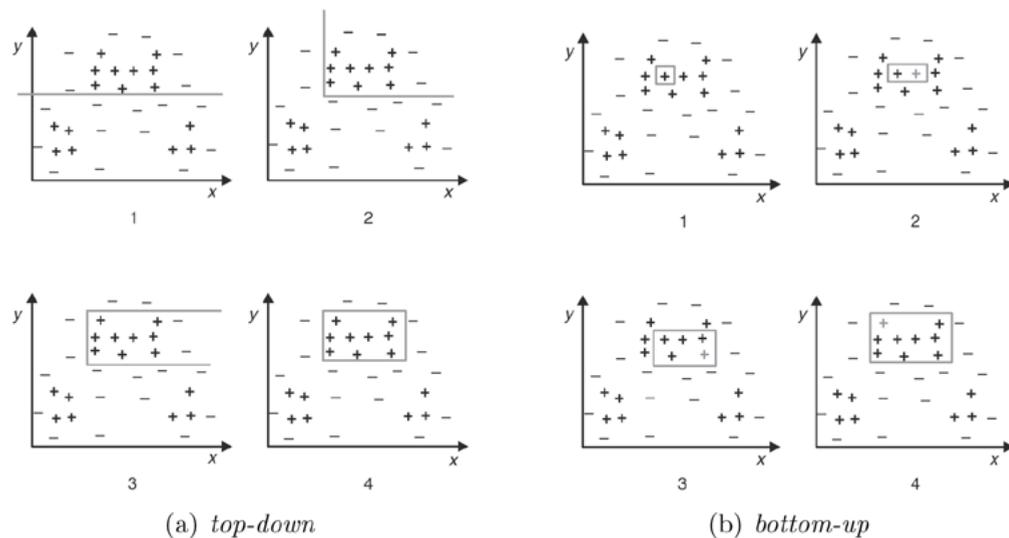
Figura 6.3 Partições usando o teste  $Temperatura > 70,5$ .



(a) Generalização menos geral

(b) Generalização de especificação máxima

Figura 6.4 Dois exemplos da superfície de decisão desenhadas por um conjunto de regras.



(a) top-down

(b) bottom-up

Figura 6.5 Exemplo ilustrativo do funcionamento do algoritmo para induzir uma regra.

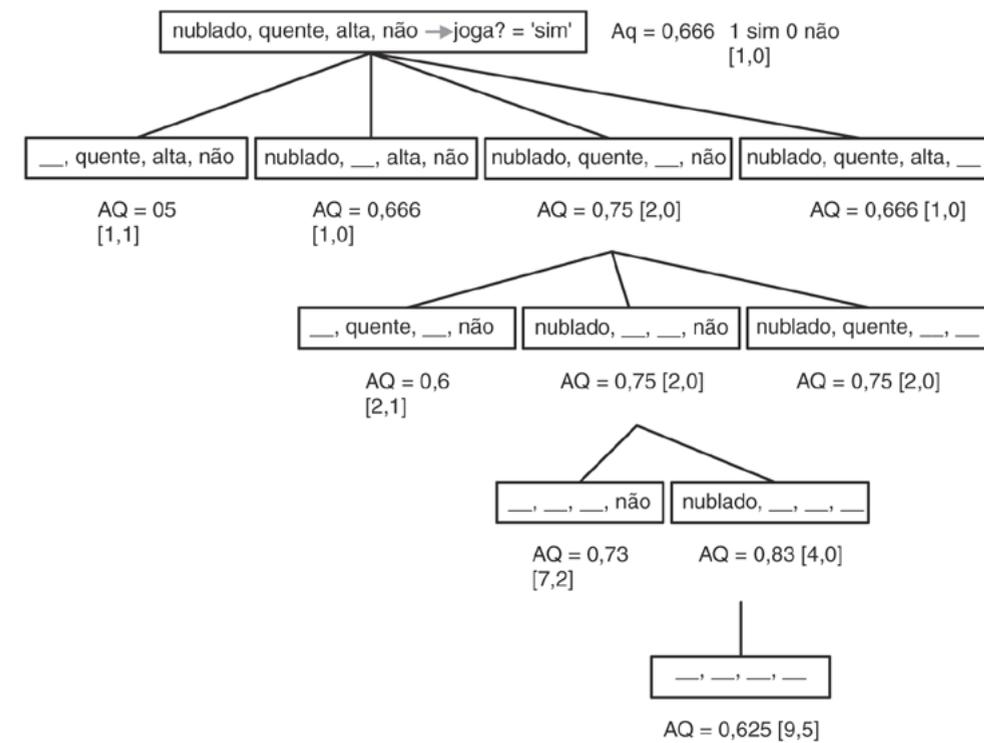


Figura 6.6 Exemplo do processo bottom-up de construção de uma regra.



## Capítulo 7 MÉTODOS BASEADOS EM OTIMIZAÇÃO

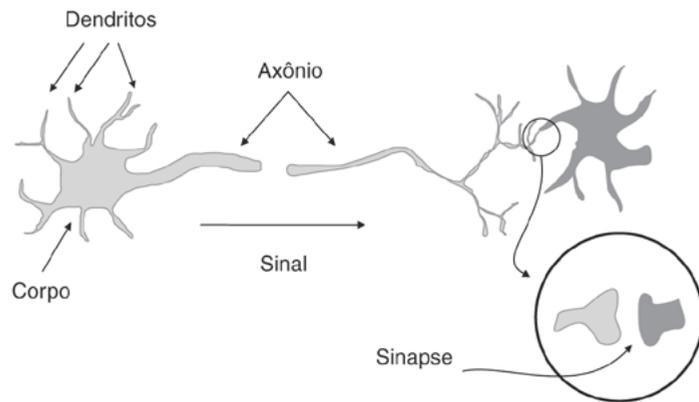


Figura 7.1 Neurônio biológico simplificado.

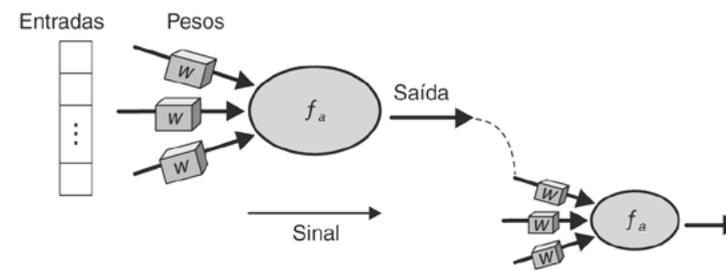


Figura 7.2 Neurônio artificial.

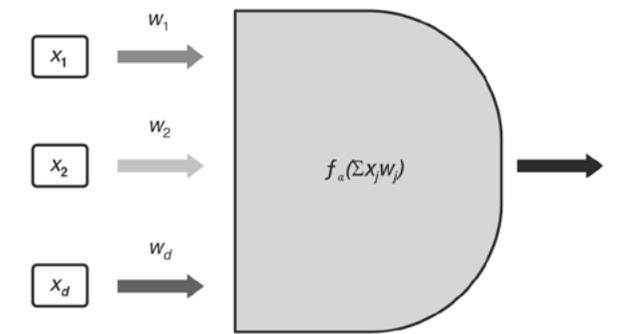


Figura 7.3 Entrada total em um neurônio artificial.

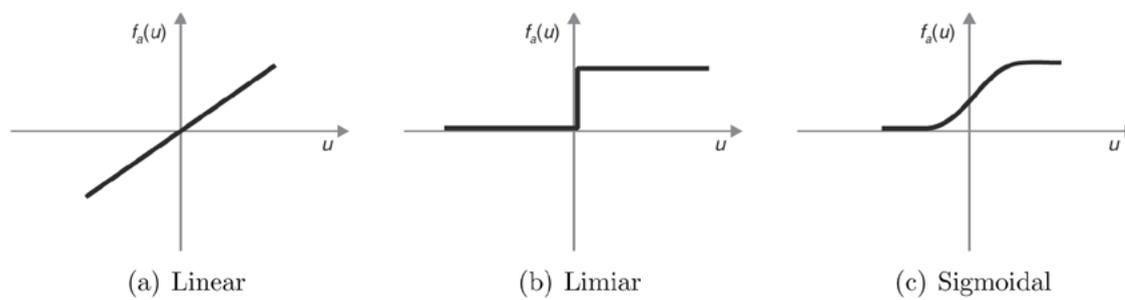


Figura 7.4 Exemplos de funções de ativação.

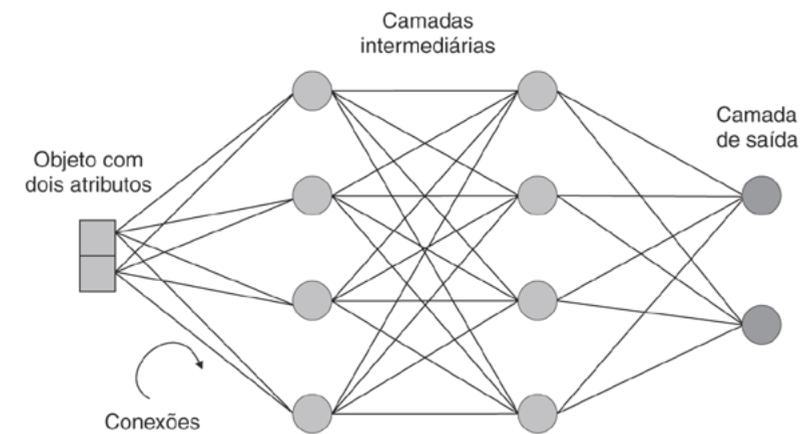


Figura 7.5 Exemplo de RNA multicamadas típica.

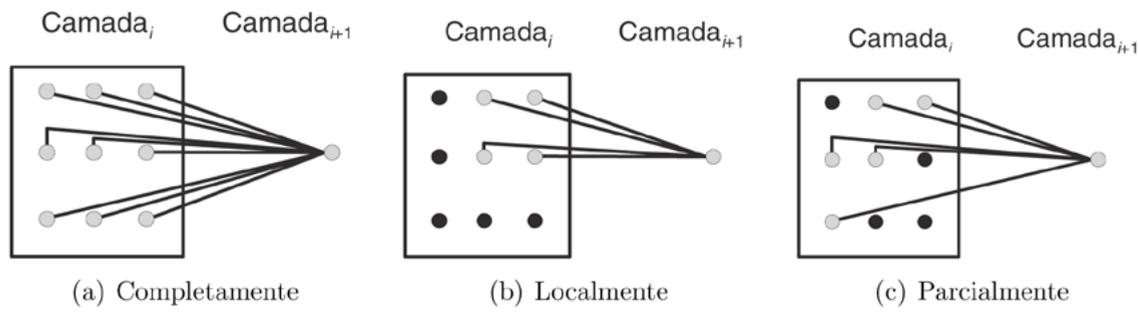


Figura 7.6 Diferentes padrões de conexão em uma RNA multicamadas.

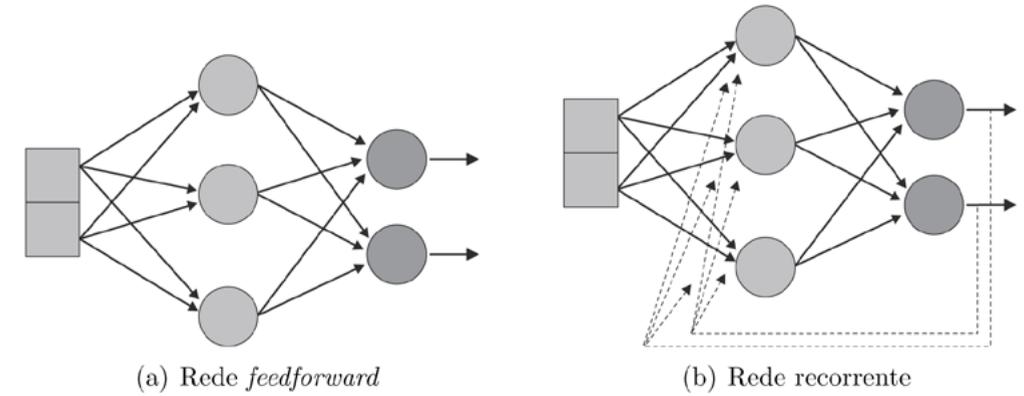


Figura 7.7 Redes neurais feedforward e recorrente.

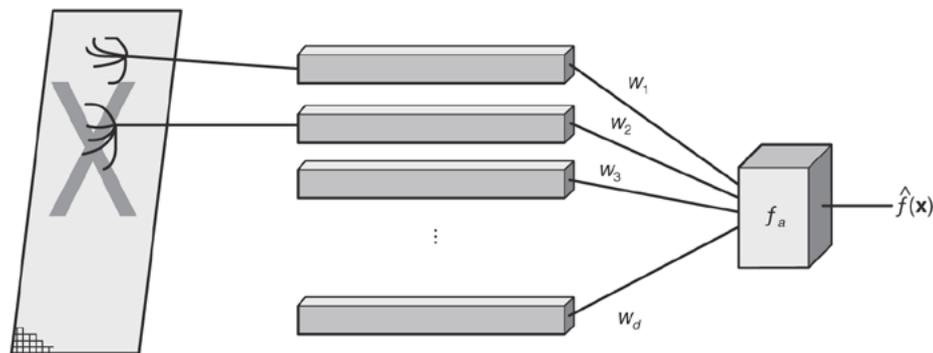


Figura 7.8 Rede perceptron.

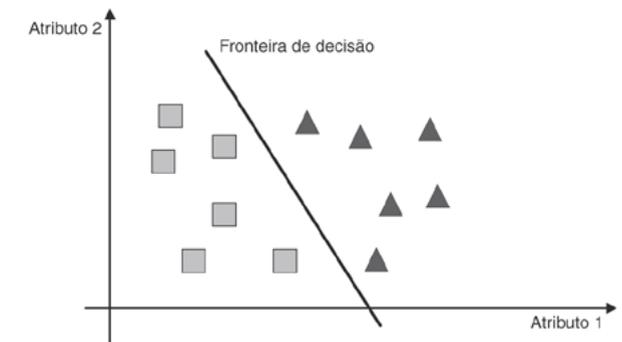


Figura 7.9 Objetos linearmente separáveis.

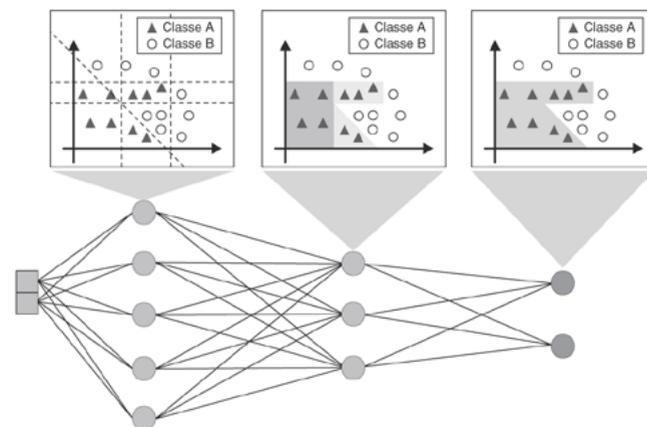


Figura 7.10 Papel desempenhado pelos neurônios das diferentes camadas da rede MLP.

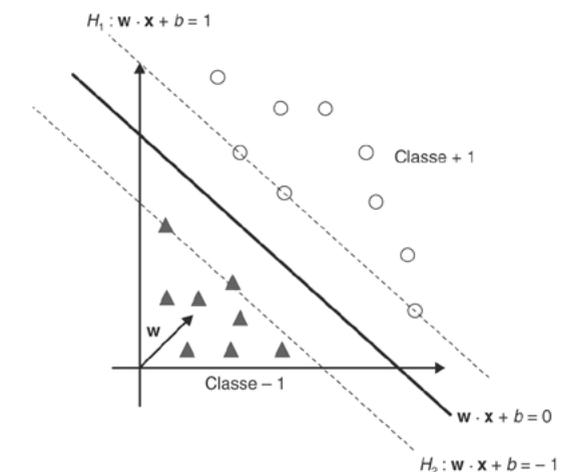


Figura 7.11 Ilustração de hiperplanos canônicos e separador.

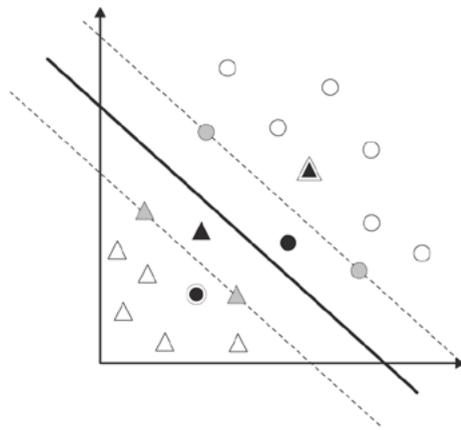


Figura 7.12 Tipos de SVs: livres (cor cinza) e limitados (cor preta).

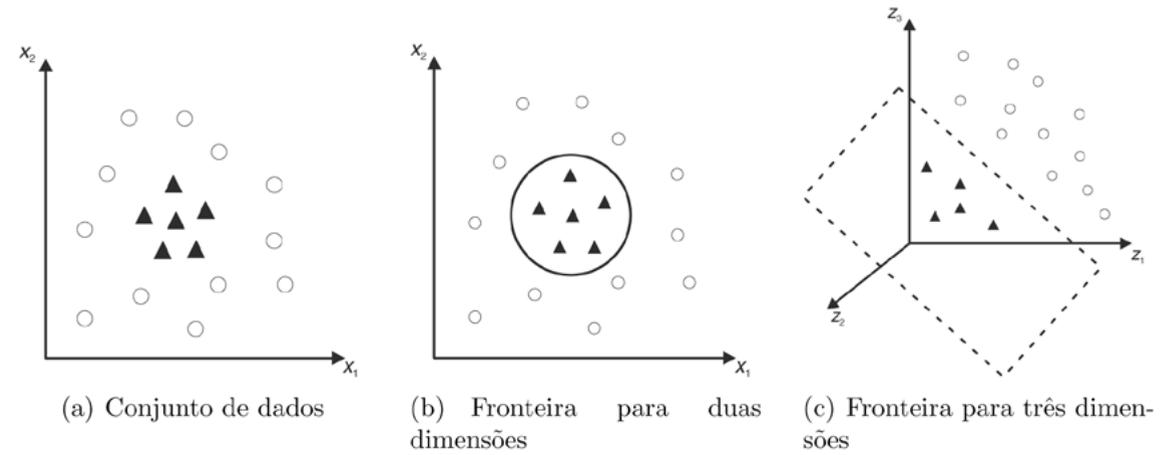


Figura 7.13 Exemplo de transformação realizada em conjunto de dados não linear para espaço de características.

Tabela 7.1 Funções kernel mais comuns

Tipo de kernel	Função $K(\mathbf{x}_i, \mathbf{x}_j)$	Parâmetros
Polinomial	$(\delta(\mathbf{x}_i \cdot \mathbf{x}_j) + \kappa)^d$	$\delta, \kappa$ e $d$
RBF	$\exp(-\sigma \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$	$\sigma$
Sigmoidal	$\tanh(\delta(\mathbf{x}_i \cdot \mathbf{x}_j) + \kappa)$	$\delta$ e $\kappa$

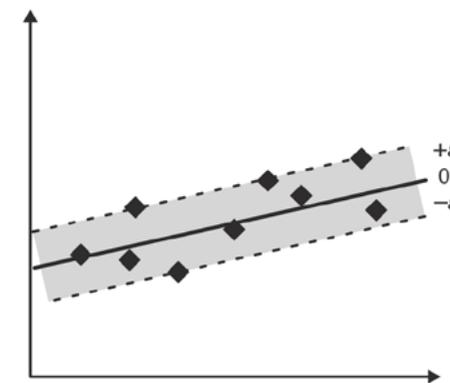


Figura 7.14 Ilustração simplificada de procedimento realizado por SVR.



## Capítulo 8 MODELOS MÚLTIPLOS PREDITIVOS

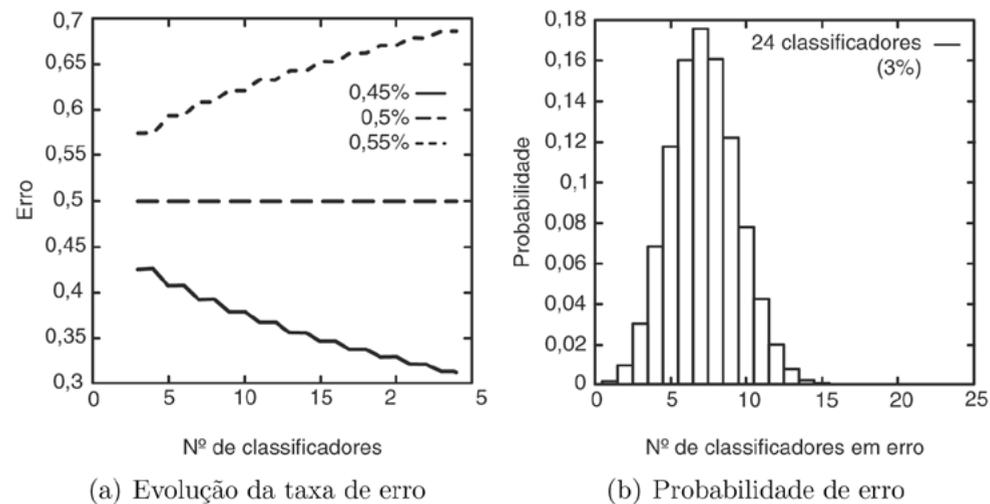


Figura 8.1 (a) Evolução da taxa de erro variando o número de classificadores em um modelo múltiplo. (b) Probabilidade de que exatamente  $i$  dos 24 classificadores cometerão um erro.

Tabela 8.1 Exemplo ilustrativo do MAI: tabela mostra o conjunto de dados original e o conjunto de dados de  $Nível_1$

V1	V2	V3	V4	V5	Classe	V1	V2	V3	V4	V5	Erro
t	a	c	t	a	membro	t	a	c	t	a	+
t	g	c	t	a	membro	t	g	c	t	a	-
g	t	a	c	t	não membro	g	t	a	c	t	+
a	a	t	t	g	membro	a	a	t	t	g	+
t	c	g	a	t	não membro	t	c	g	a	t	-
a	g	g	g	g	membro	a	g	g	g	g	+

Conjunto de dados original                      Conjunto de dados  $Nível_1$

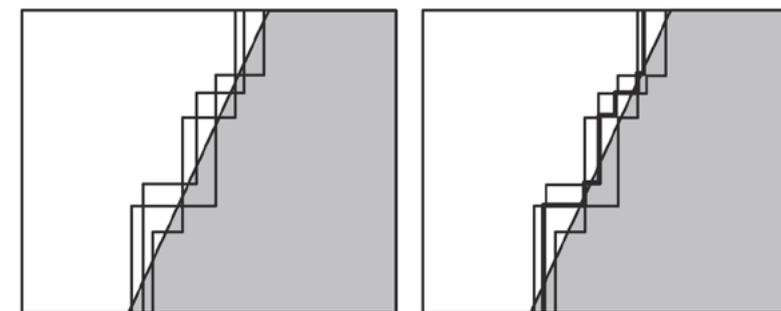


Figura 8.3 A figura da esquerda representa as superfícies de decisão de três classificadores num problema de duas classes. A figura da direita apresenta, em negrito, a superfície de decisão obtida por votação uniforme dos três classificadores.

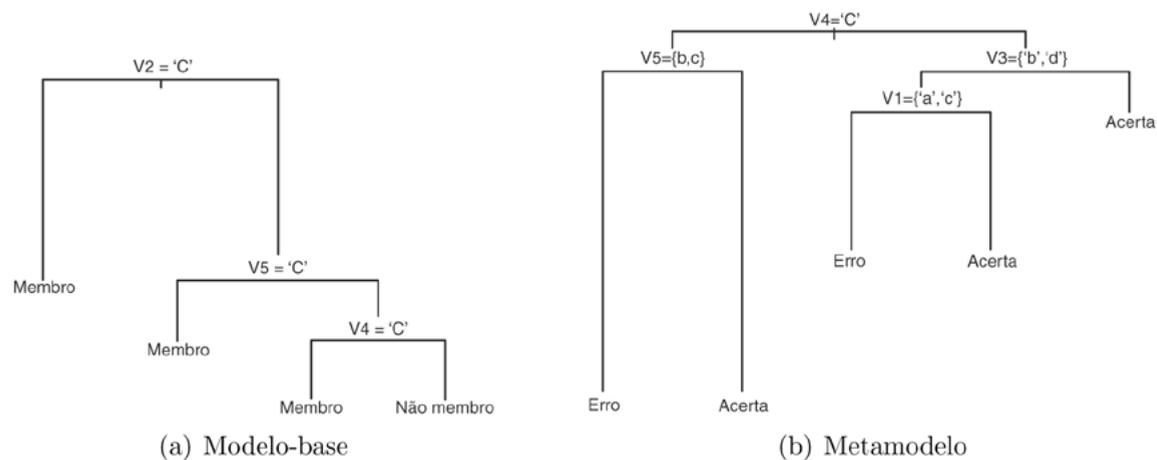


Figura 8.2 Exemplo de uma árvore de decisão originada a partir dos dados originais 8.2(a) e um metaclassificador 8.2(b), também na forma de uma árvore de decisão, gerado a partir dos dados de  $Nível_1$ .

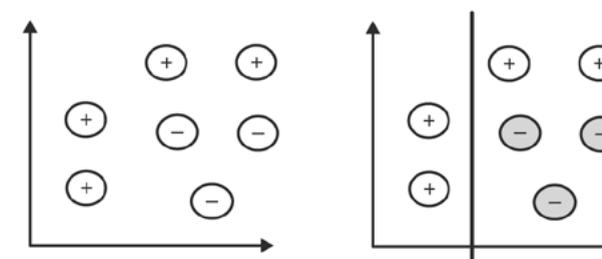
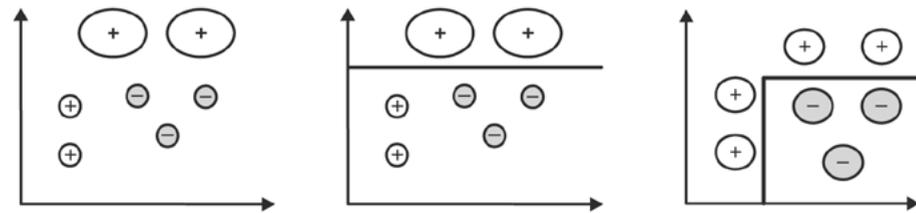
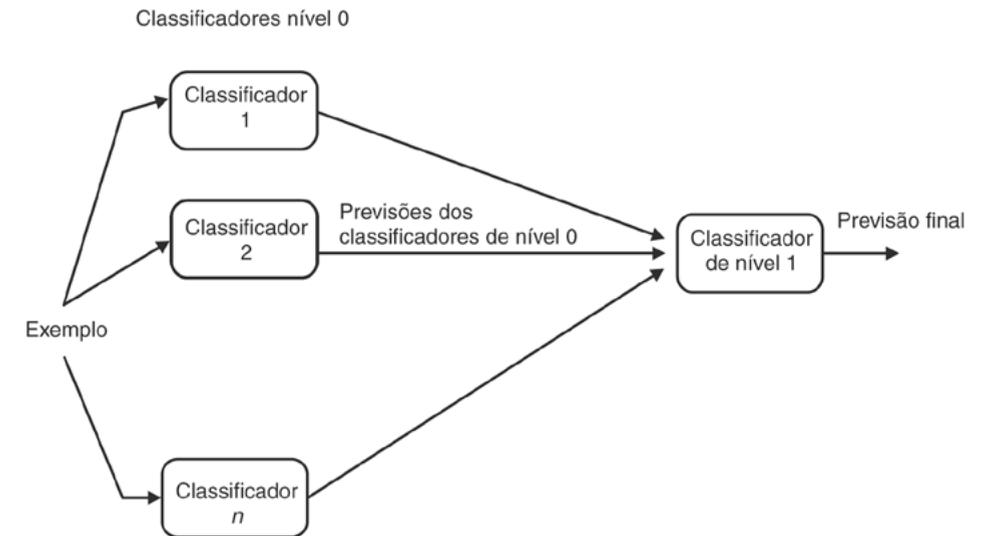


Figura 8.4 Boosting: primeira iteração. A figura da esquerda mostra a distribuição uniforme original dos exemplos. A figura da direita mostra a superfície de decisão para essa distribuição.



**Figura 8.5** Boosting: segunda iteração. O peso dos exemplos classificados incorretamente na primeira iteração aumentou. A figura do centro mostra a superfície para a nova distribuição dos exemplos. O modelo final, obtido conjugando os dois modelos anteriores, é apresentado na figura da direita.



**Figura 8.6** Arquitetura generalização em pilha.

**Tabela 8.2** Exemplo ilustrativo de generalização em pilha: A tabela mostra o conjunto de dados original e o conjunto de dados de  $Nível_1$

**Tabela 8.2** Exemplo ilustrativo de generalização em pilha: A tabela mostra o conjunto de dados original e o conjunto de dados de  $Nível_1$

V1	V2	V3	V4	V5	Classe	$P_{1,1}$	$P_{1,2}$	$P_{2,1}$	$P_{2,2}$	$P_{3,1}$	$P_{3,2}$	Classe
t	a	c	t	a	Membro	0,51	0,49	0,13	0,87	0,12	0,88	Membro
t	g	c	t	a	Membro	0,19	0,81	0,07	0,93	0,81	0,19	Membro
g	t	a	c	t	Não Membro	0,68	0,32	0,55	0,45	0,69	0,31	Não Membro
a	a	t	t	g	Membro	0,74	0,26	0,66	0,34	0,94	0,06	Membro
t	c	g	a	t	Não Membro	0,62	0,38	0,01	0,99	0,78	0,22	Não Membro
a	g	g	g	g	Membro	0,65	0,35	0,90	0,10	0,55	0,45	Membro

Conjunto de dados original

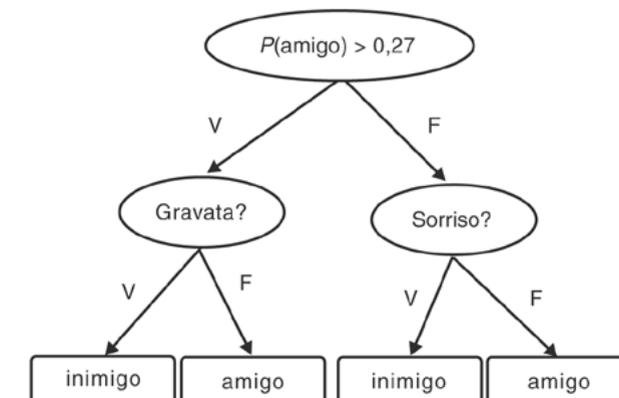
Conjunto de dados de  $Nível_1$

**Tabela 8.3** Dois exemplos do conjunto de dados de  $Nível_0$  no problema Monks-2

Cabeça	Corpo	Sorriso	Objeto	Cor	Gravata	Classe
redonda	redondo	sim	espada	vermelho	sim	inimigo
redonda	redondo	não	balão	azul	não	amigo

**Tabela 8.4** Dois exemplos do conjunto de dados de  $Nível_1$

Cabeça	Corpo	Sorriso	Objeto	Cor	Gravata	$P(\text{amigo})$	$P(\text{inimigo})$	Classe
redonda	redondo	sim	espada	vermelho	sim	0,135	0,864	inimigo
redonda	redondo	não	balão	azul	não	0,303	0,696	amigo



**Figura 8.7** A árvore gerada pelo C4.5  $\nabla$  naive Bayes.



## Capítulo 9 AVALIAÇÃO DE MODELOS PREDITIVOS

Classe verdadeira	Classe predita		
	1	2	3
1	11	1	3
2	1	4	0
3	2	1	6

Figura 9.1 Exemplo de matriz de confusão para um problema com três classes.

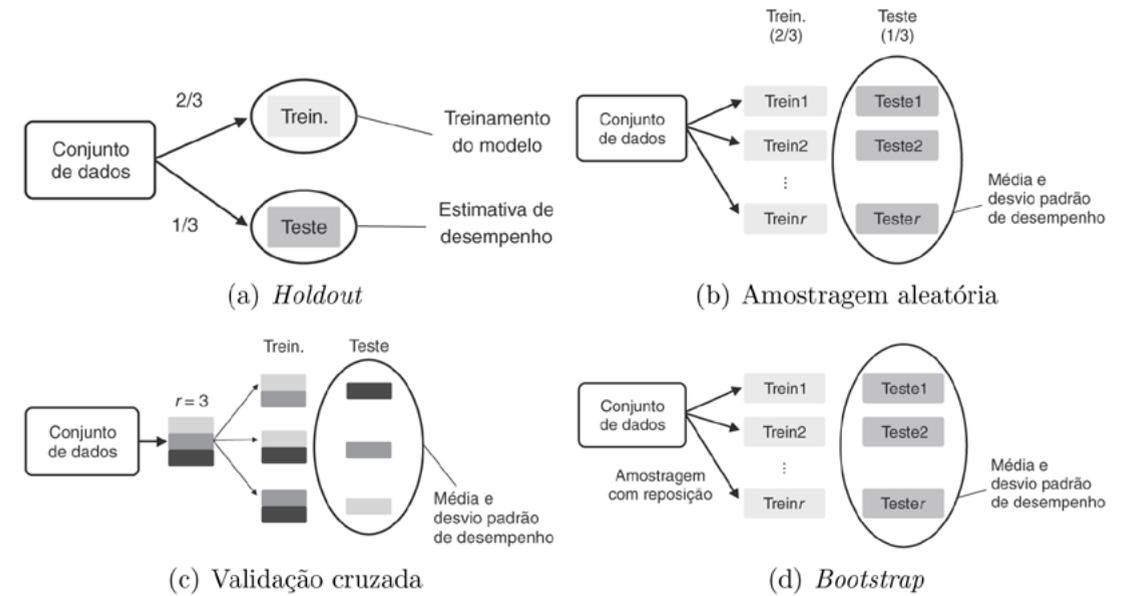


Figura 9.2 Métodos de amostragem.

Tabela 9.1 Matriz de confusão para um problema com duas classes

Classe verdadeira	Classe predita	
	+	-
+	VP	FN
-	FP	VN

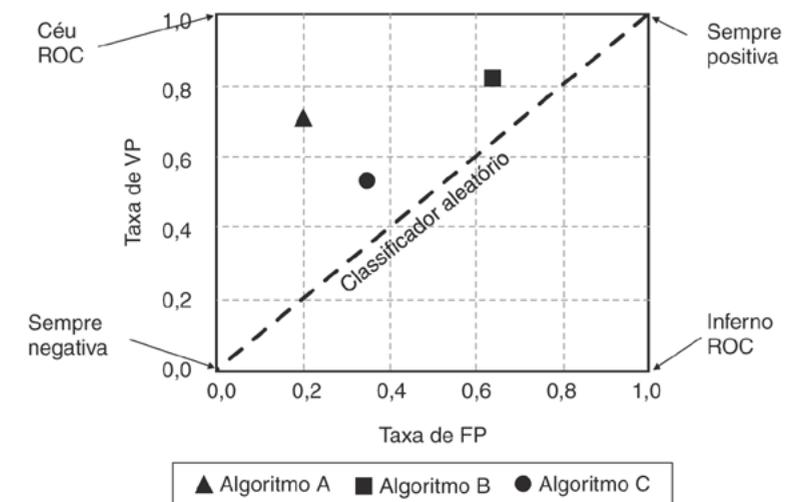


Figura 9.3 Espaço ROC com três classificadores.

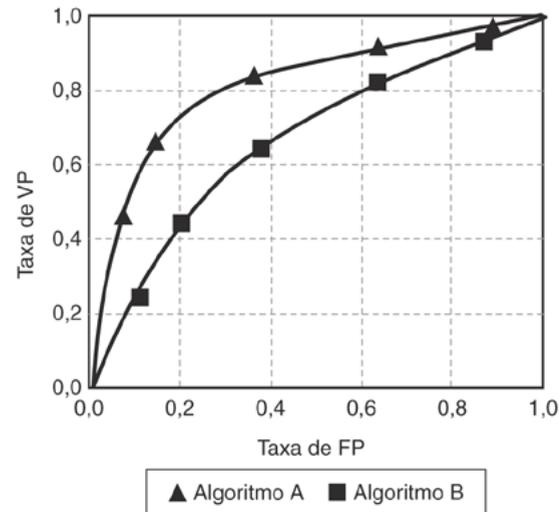


Figura 9.4 Exemplos de curva ROC.

Tabela 9.2 Exemplo de tabela de diferenças de resultados em teste de Wilcoxon

Conj. dados	C4.5	C4.5+m	Diferença	Dif_absoluta	Posição
Pulmão	0,583	0,583	0,000	0,000	1,5
Fungo	0,583	0,583	0,000	0,000	1,5
Atmosfera	0,882	0,888	+0,006	0,006	3,0
Mama	0,599	0,591	-0,008	0,008	4,0

Tabela 9.3 Valores de  $q_{0,05}$  para diferentes pós-testes

A	2	3	4	5	6	7	8	9	10
Nemenyi	1,960	2,343	2,569	2,728	2,850	2,949	3,031	3,102	3,164
Bonferroni-Dunn	1,960	2,241	2,394	2,498	2,576	2,648	2,690	2,724	2,773

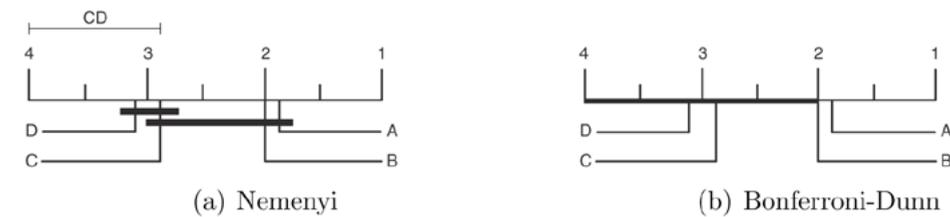


Figura 9.5 Representação gráfica de pós-testes.



## Capítulo 10 MINERAÇÃO DE PADRÕES FREQUENTES

TID	Itens	0 itens	1 item	2 itens	3 itens
1	{a,d,e}	∅: 10	{a}: 7	{a,c}: 4	{a,c,d}: 3
2	{b,c,d}		{b}: 3	{a,d}: 5	{a,c,e}: 3
3	{a,c,e}		{c}: 7	{a,e}: 6	{a,d,e}: 4
4	{a,c,d,e}		{d}: 6	{b,c}: 3	
5	{a,e}		{e}: 7	{c,d}: 4	
6	{a,c,d}			{c,e}: 4	
7	{b,c}			{d,e}: 4	
8	{a,c,d,e}				
9	{b,c,e}				
10	{a,d,e}				

Figura 10.1 Um banco de dados de transações, com 10 transações, e a enumeração de todos os conjuntos de itens frequentes usando o suporte mínimo de  $s_{\min} = 3$ .

TID	Item set
1	{a,d,e}
2	{b,c,d}
3	{a,c,e}
4	{a,c,d,e}
5	{a,e}
6	{a,c,d}
7	{b,c}
8	{a,c,d,e}
9	{b,c,e}
10	{a,d,e}

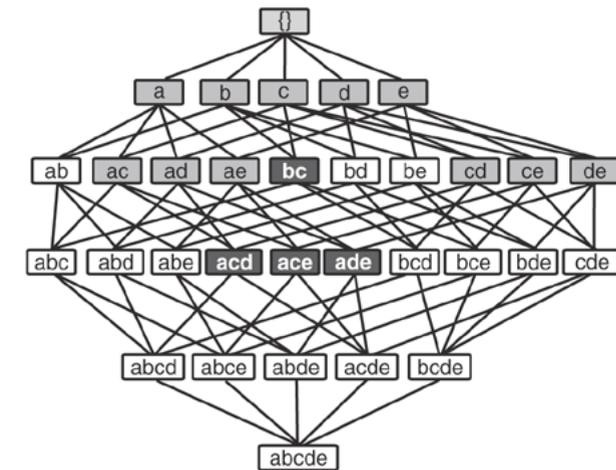


Figura 10.2 Um banco de dados de transações, com 10 transações, e o espaço de busca para encontrar todos os possíveis itemsets frequentes usando o suporte mínimo de  $s_{\min} = 3$ .

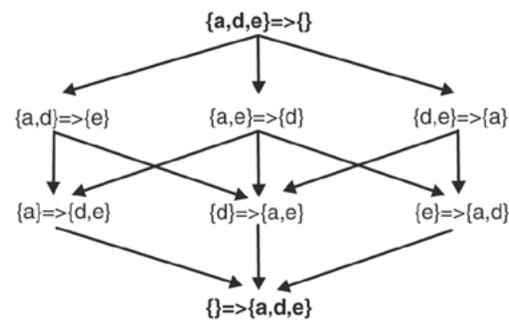


Figura 10.3 Espaço de busca de regras de associação para um itemset frequente.

Tabela 10.1 Regras extraídas

Regra	Confiança	Suporte do itemset
{b} → {c}	100,0%	30%
{d, e} → {a}	100,0%	40%
{e} → {a}	85,7%	60%
{a} → {e}	85,7%	60%
{d} → {a}	83,3%	50%
{a, d} → {e}	80,0%	40%

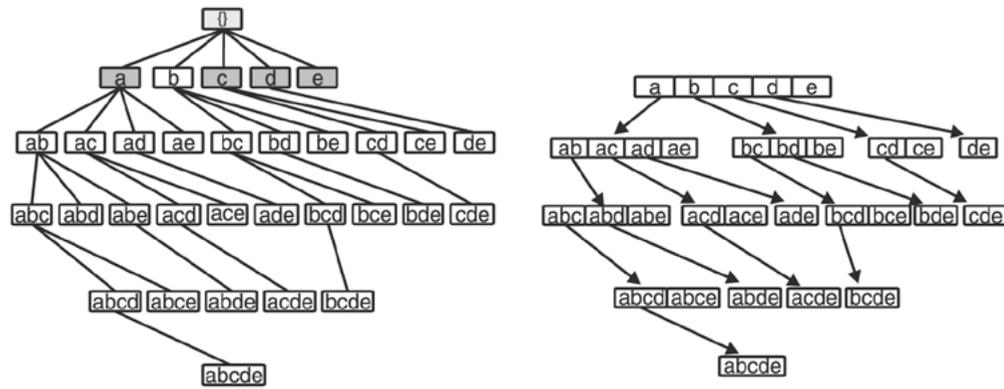


Figura 10.4 O espaço de pesquisa utilizando a profundidade e a árvore de prefixos correspondente para 5 itens.

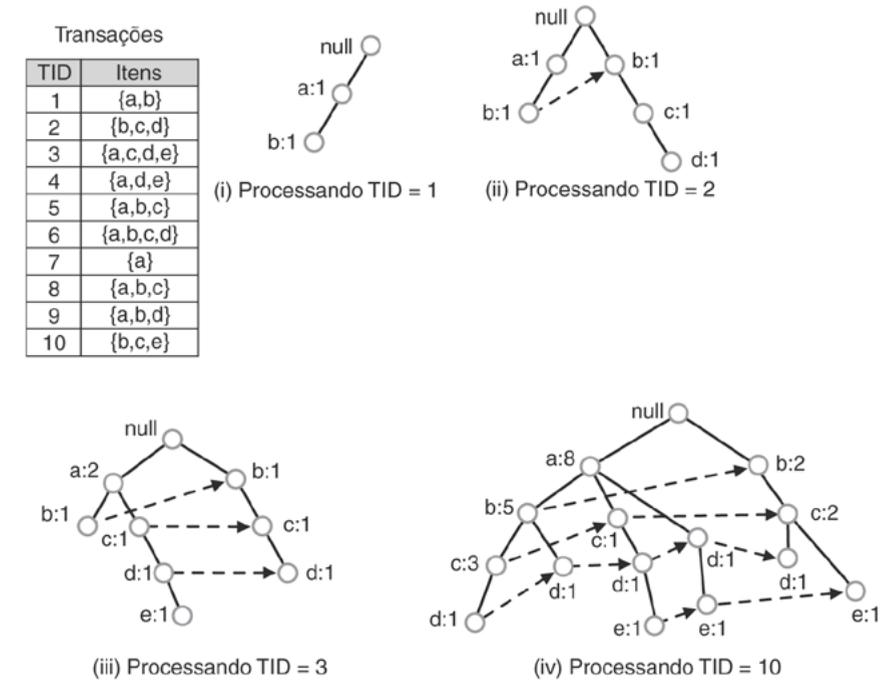


Figura 10.5 Construção de uma FP-tree.

Tabela 10.2 Preferências sobre o consumo de Chá e Café de 1000 consumidores

	Café	Não Café	Total
Chá	150	50	200
Não Chá	650	150	800
<b>Total</b>	800	200	1000



## Capítulo 11 ANÁLISE DE AGRUPAMENTOS

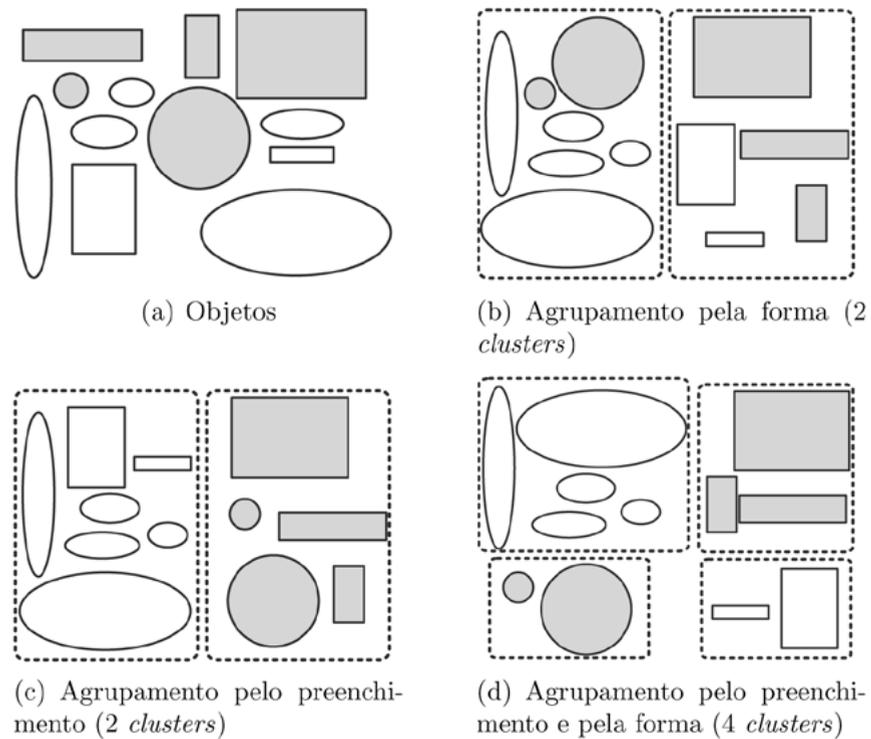


Figura 11.1 Objetos agrupados de diferentes maneiras.

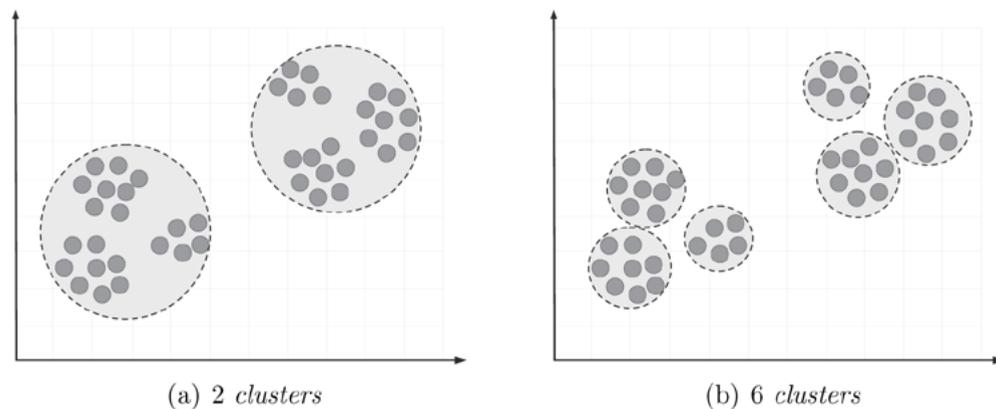


Figura 11.3 Dados com clusters em diferentes níveis de refinamento.

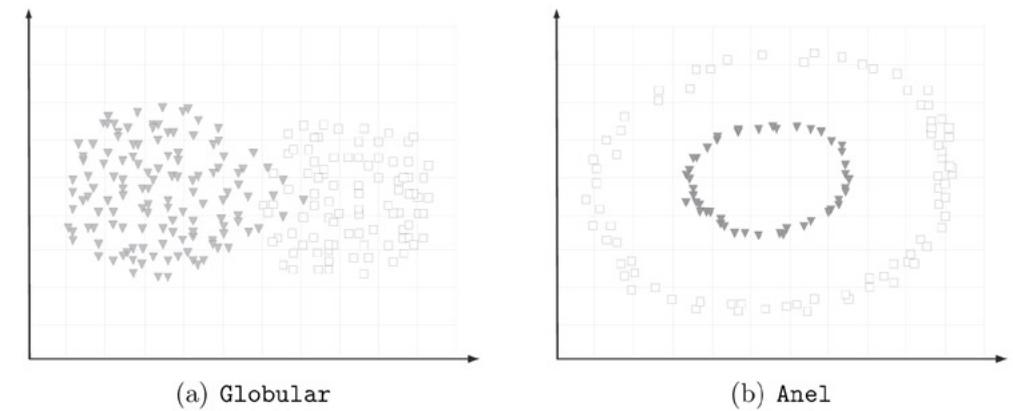


Figura 11.2 Dados com clusters em conformidade com diferentes critérios.

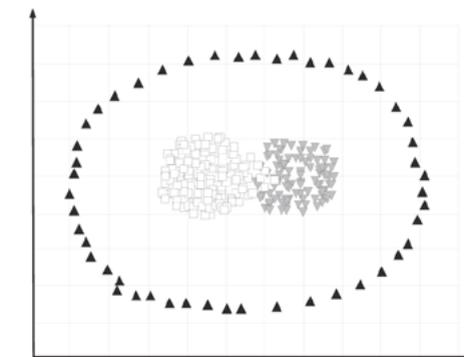


Figura 11.4 Conjunto de dados com uma estrutura de clusters heterogênea.

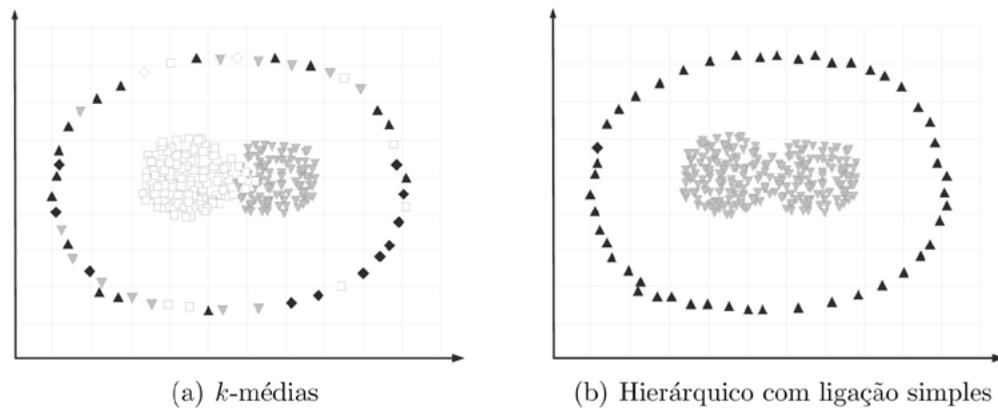


Figura 11.5 Resultados dos algoritmos em dados com estrutura heterogênea.

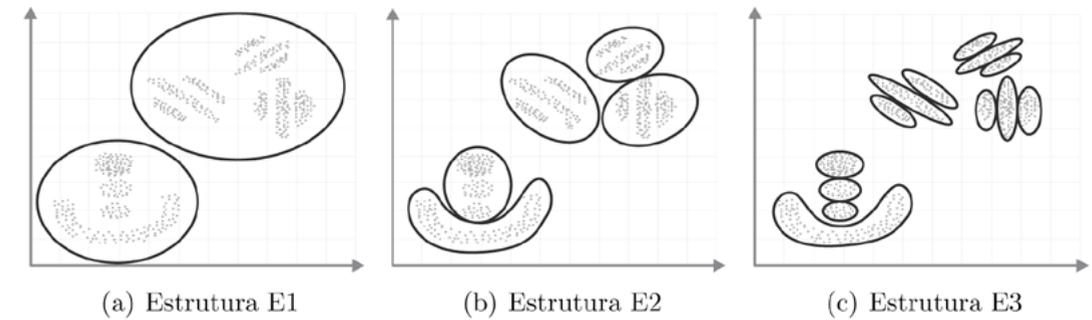


Figura 11.6 Conjunto de dados com várias estruturas heterogêneas.

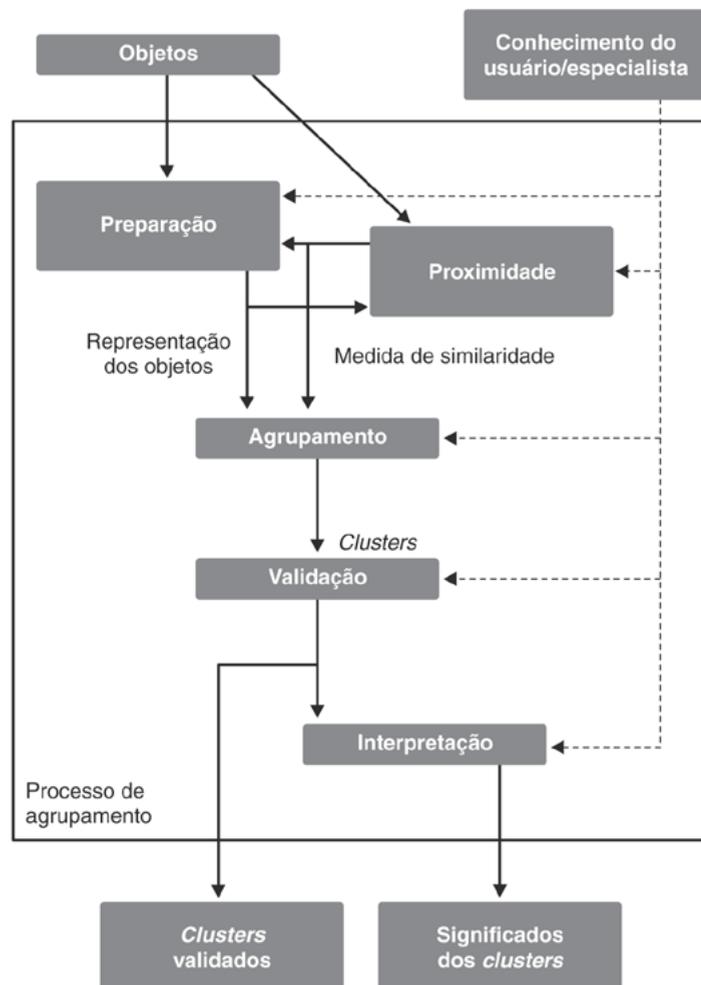


Figura 11.7 Etapas do processo de agrupamento.

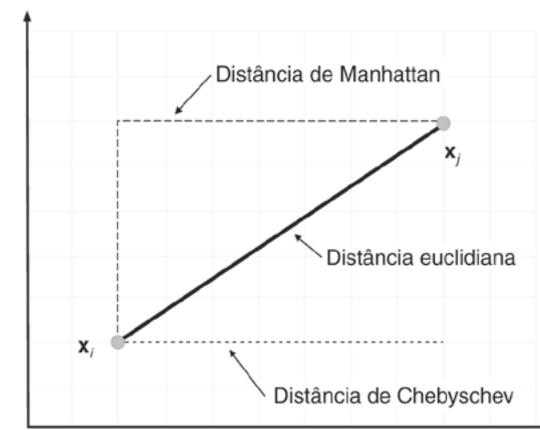


Figura 11.8 Interpretação das métricas de Minkowski para  $d = 2$ .

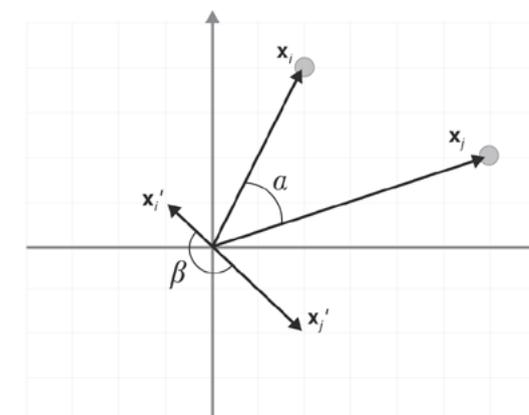


Figura 11.9 Interpretação geométrica da separação angular e da correlação de Pearson para  $d = 2$ .



## Capítulo 12 ALGORITMOS DE AGRUPAMENTOS

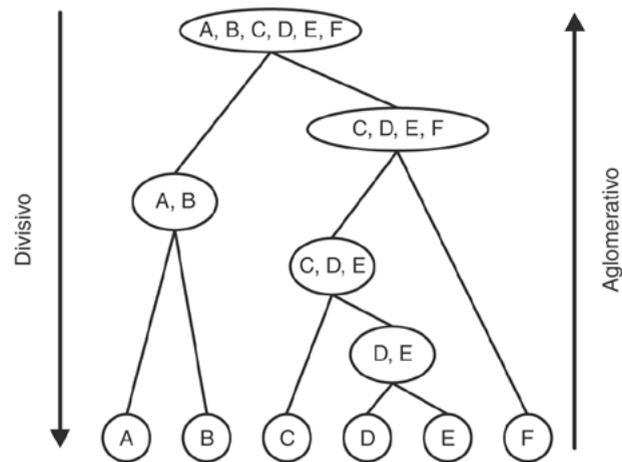


Figura 12.1 Funcionamento dos algoritmos hierárquicos aglomerativos e divisivos.

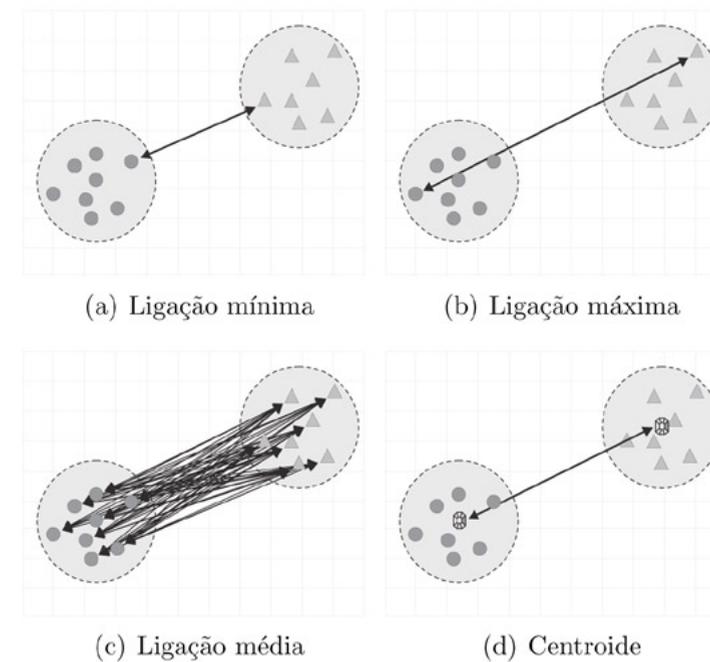


Figura 12.2 Distâncias entre clusters.

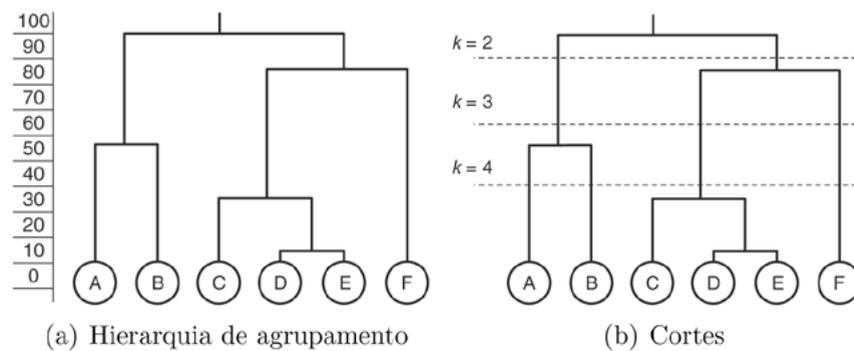


Figura 12.3 Exemplo de dendrograma.