

Processamento Digital de Imagens

Classificação de Imagens

-Não supervisionada

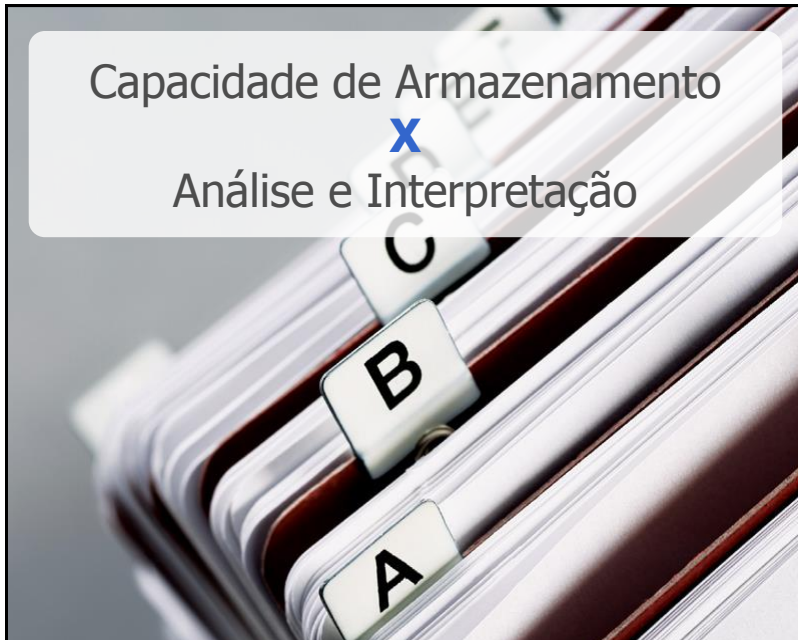
-Supervisionada

1

Capacidade de Armazenamento

X

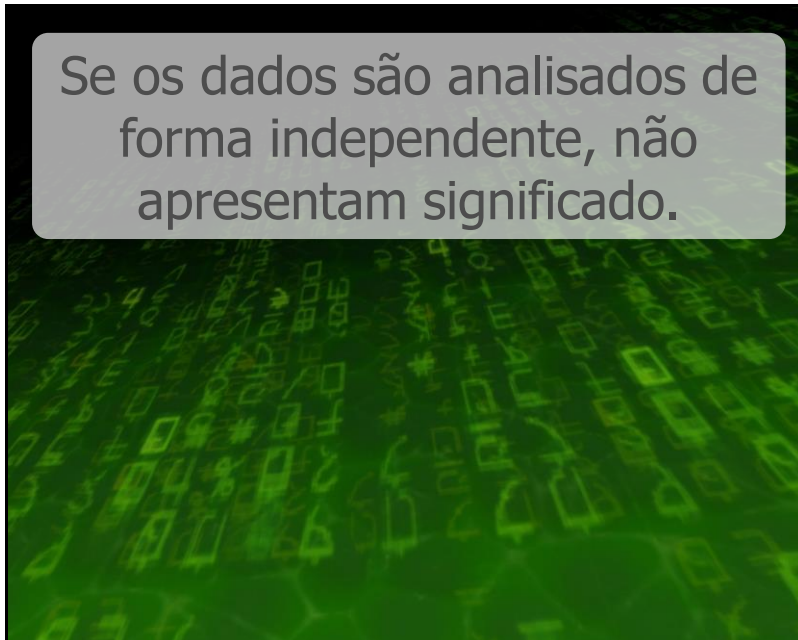
Análise e Interpretação



Satélites de SR
geram imagens
diariamente.



Se os dados são analisados de
forma independente, não
apresentam significado.



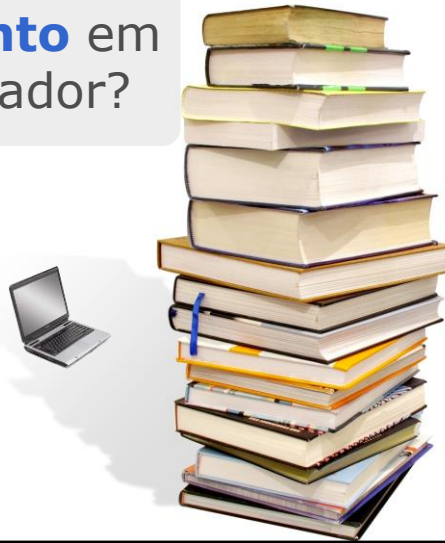


A interpretação de
imagens deve ser
automatizada.

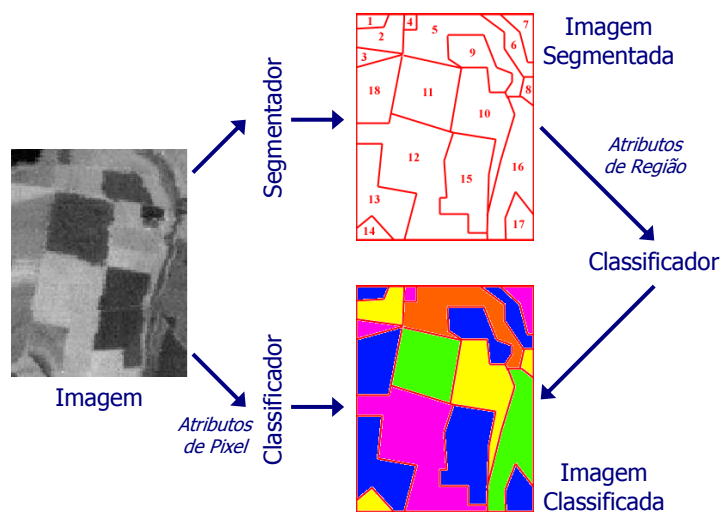
Como encontrar um alvo
em uma imagem de SR?



Como representar o **conhecimento** em um computador?

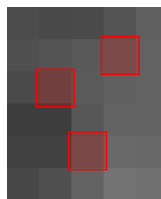


Processamento de Imagens

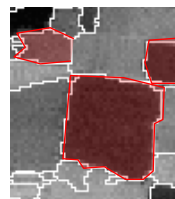




Objetos de Análise



Pixel



Região

Elementos

- Vetor de Atributos
 $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$
- Conjunto de k Classes
 $\{w_1, w_2, \dots, w_k\}$

Objetivos

- Particionar o espaço de Atributos
 - Associar uma classe a cada objeto
- <http://bit.ly/geodma-features>

Métodos de Classificação

- Não Supervisionados
 - Estrutura das classes determinada a partir de um conjunto de amostras não rotuladas
- Supervisionados
 - Conjunto de exemplos rotulados permite modelar classes e fronteiras de decisão

11

Métodos Não Supervisionados

- Definição de Everitt, 1981

“Regiões contínuas no espaço de atributos contendo uma densidade de pontos relativamente alta, separada de outras regiões de alta densidade, por regiões de densidade relativamente baixa de pontos”

12

Métodos Não Supervisionados

- Conjunto de padrões (amostras) não rotulados

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$$

- Objetivo básico:
 - Particionar as amostras em grupos (*clusters*) que permitam descobrir as similaridades/diferenças entre os padrões
 - O número de grupos é (em geral) desconhecido

13

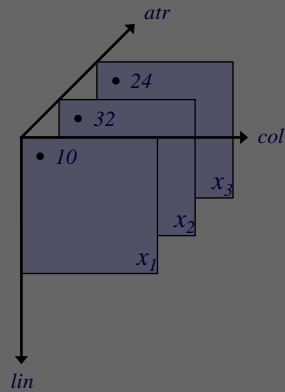
Agrupamento (*clustering*)
separa automaticamente
objetos em subconjuntos
similares entre si, e diferentes
dos demais.

Espaços

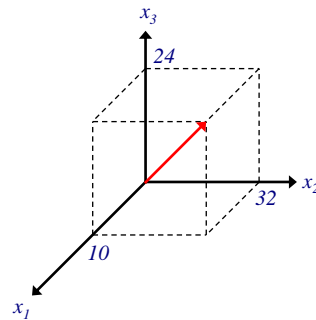
Espaço da Imagem

X

Espaço de Atributos

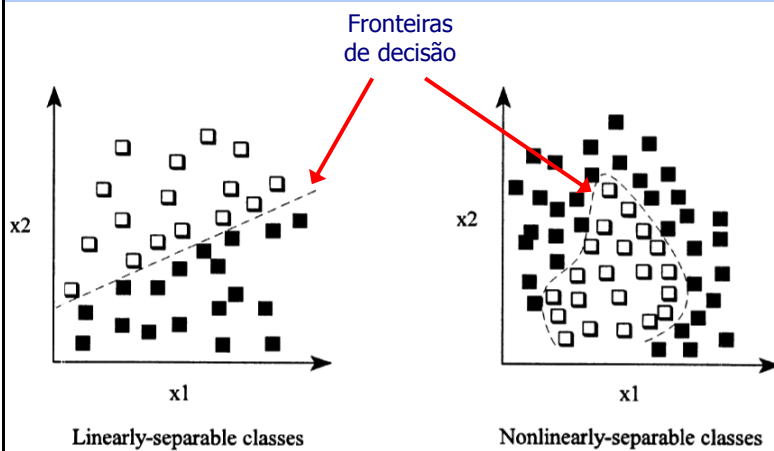


$\mathbf{x} = [x_1 \ x_2 \ x_3]^T \rightarrow$ vetor de atributos
 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \rightarrow$ amostras



15

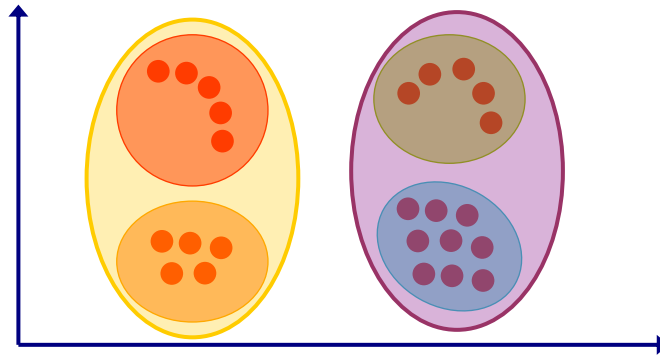
Separabilidade entre Classes



16

Critérios de Agregamento

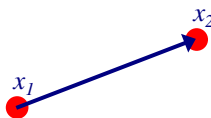
- Dependem do algoritmo e da medida de similaridade
- Quantas classes?



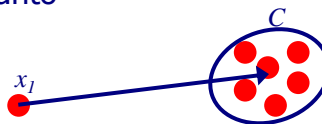
17

Medidas de Similaridade

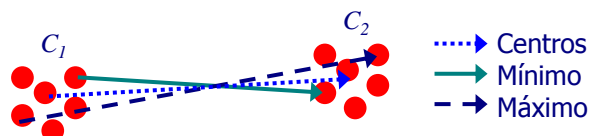
- Ponto a ponto



- Ponto a Conjunto



- Conjunto a Conjunto



18

Algoritmos Não Supervisionados

- Sequenciais
- Hierárquicos
- Baseados em Função de Custo
 - k-médias
 - ISODATA
 - SOM
 - Fuzzy sets
 - ...
- Híbridos

19

Método sequencial

- Produz um conjunto de grupos
- Rápido e direto
- Padrões são apresentados poucas vezes
 - Influência no resultado
- Número de grupos desconhecido *a priori*

20

Método sequencial

- Algoritmo básico:

Definir um grupo $C_1 = \{\mathbf{x}_1\}$, $k = 1$

Para cada padrão \mathbf{x}_i , $i = 1, \dots, N$

Procurar o grupo mais próximo C_j

Se $d(\mathbf{x}_i, C_j) > D$ e $k < k_{max}$

Criar novo grupo $C_{k+1} = \{\mathbf{x}_i\}$

Atualizar $k \leftarrow k+1$

Senão

Inserir \mathbf{x}_i em $C_j \rightarrow C_j = C_j \cup \{\mathbf{x}_i\}$

Entrada

Conjunto de padrões

$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$

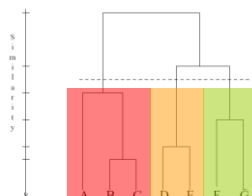
Máximo de grupos $\rightarrow k_{max}$

Raio máximo do grupo $\rightarrow D$

21

Método hierárquico

- Produz uma hierarquia de partições em vez de uma única partição
- Solução é encontrada em N passos
 - N é número de padrões
- Em cada passo (t) uma nova partição é obtida a partir da partição obtida no passo anterior ($t-1$)
- Representação por Dendograma



22

Método hierárquico

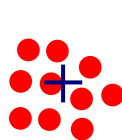
- Algoritmo básico:

1. Calcular a matriz de proximidade de padrões
Inicialmente cada padrão é um grupo
2. Encontrar o par mais similar (C_i, C_j)
3. Juntar os grupos $C_{new} = C_i \cup C_j$
Remover C_i e C_j
Incluir C_{new}
Atualizar matriz de proximidade
4. Se todos os padrões estão em um único grupo
Parar
Senão
Voltar ao passo 2

23

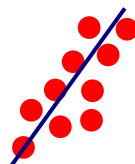
Funções de Custo

- Algoritmos com objetivo de encontrar uma partição que otimize uma função de custo
- Elementos
 - Conjunto de padrões $\rightarrow \mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$
 - Vetor de parâmetros $\rightarrow \theta$
 - Função de Custo $\rightarrow J(\mathbf{X}; \theta)$

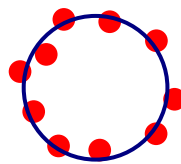


Compacto

$$\theta = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_k]^T$$



Linear



Esférico

$$\theta = [\mathbf{c}_1, r_1, \dots, \mathbf{c}_k, r_k]^T$$

24

Algoritmo k-médias

- Recupera grupos compactos
- Sensível a ruído e "outliers"
- Limitado a atributos numéricos
- Existem variações do algoritmo (ISODATA):
 - Divisão de grupos
 - Junção de grupos
 - Descarte de grupos



25

Algoritmo k-médias

- Algoritmo básico:

1. Definir solução inicial ($t=0$)
 $\mathbf{m}(0) = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_k]^T$
2. Classificar cada padrão no grupo mais próximo
obter as partições C_1, C_2, \dots, C_k
Calcular novas médias em $(t+1)$
Se $\|\mathbf{m}(t) - \mathbf{m}(t-1)\| < \zeta$ (convergência)
Parar \rightarrow solução $\{U, \mathbf{m}\}$
Senão
Voltar ao passo 2

Entrada

Conjunto de padrões

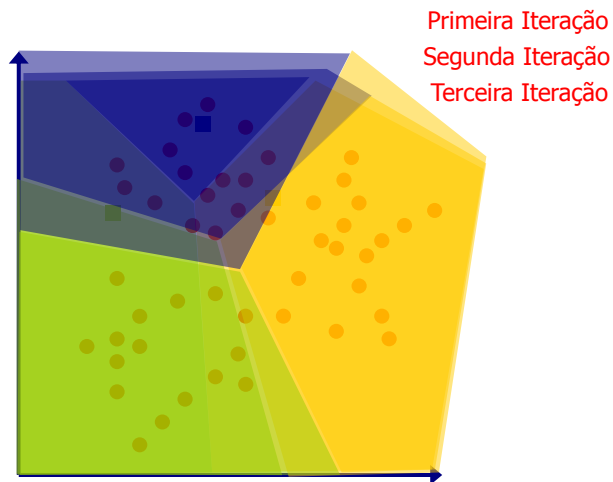
$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$

Número de grupos $\rightarrow k$

Limiar ζ de convergência

26

K-Médias com 3 grupos



27

Algoritmo ISODATA

- Iterative Self-Organizing Data Analysis Techniques
- Parâmetros
 - N_{min} tamanho mínimo do grupo (para descartar grupos)
 - k número desejado de grupos
 - σ_s^2 variância máxima de um atributo (divisão de grupo)
 - D_m distância máxima entre grupos (junção de grupos)
 - N_m número máximo de grupos a ser criado

28

Algoritmo ISODATA

- Algoritmo básico:

1. Realizar o k -médias sobre o conjunto
 2. Dividir os grupos com padrões dissimilares
 3. Juntar grupos similares
- Repetir 2 e 3 até obter solução estável

29

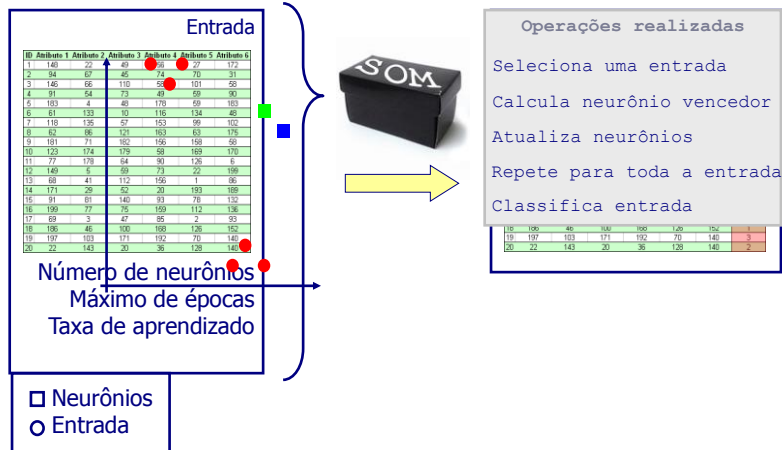
Mapas Auto-Organizáveis

- SOM – *Self Organizing Maps*
- Baseado em redes neurais
- Aprendizado competitivo
 - Neurônios mais próximos ao padrão são vencedores
- Atualização dos pesos proporcional
 - Convergência suavizada
- Soluções armazenadas nos neurônios
- <http://bit.ly/video-self-organizing-maps>



30

Self Organizing Maps



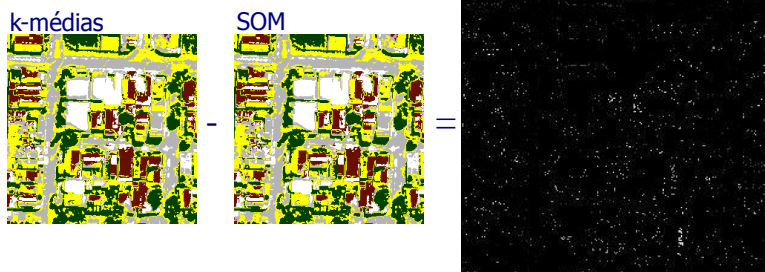
31

Comparação de resultados



32

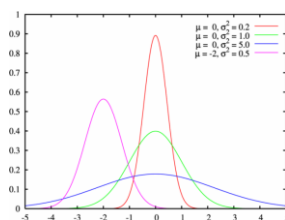
Diferenças



33

Modelos de Mistura

- O valor do pixel em qualquer banda é a combinação linear da resposta de cada componente dentro do pixel
- GMM: *Gaussian Mixture Model*
- Assume-se que as formas das funções densidade de probabilidade são conhecidas, apenas é necessário aprender o valor de um vetor de parâmetros



34

Modelos de Mistura

- As amostras pertencem a um conjunto de c classes, de tamanho conhecido
- Os valores dos c vetores de $\theta_1, \dots, \theta_c$ são desconhecidos

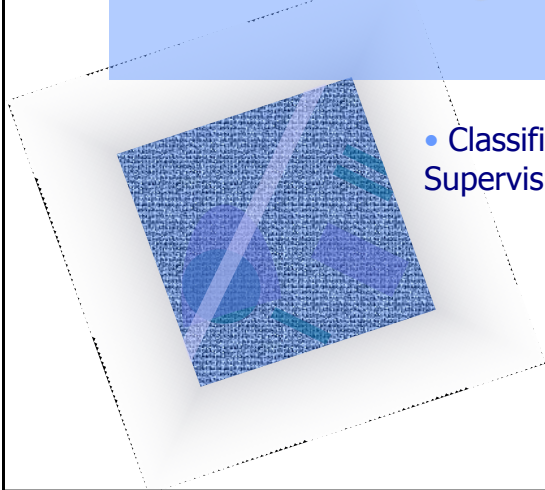
35

Modelos de Mistura

- A função de densidade de probabilidade é denominada função de mistura
- Objetivo
 - Usar os dados para estimar os valores dos parâmetros desconhecidos θ_c
- Com a estimativa de θ , é possível decompor a mistura em seus componentes

36

Processamento Digital de Imagens



- Classificação Supervisionada

Classificação Supervisionada

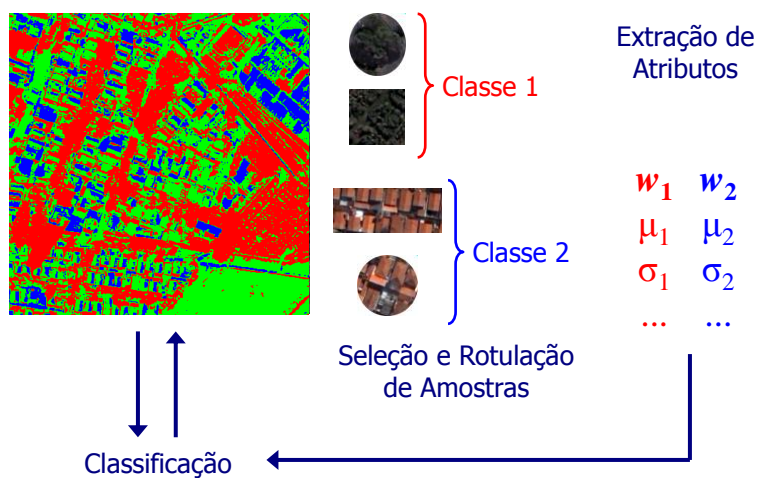
- Conjunto de exemplos (padrões, amostras) rotulados permitem modelar as classes e as fronteiras de decisão do classificador
- Exige treinamento
- Métodos
 - Paralelepípedo
 - Árvores
 - Distância mínima
 - Estatísticos

Classificação Supervisionada

- Definição das classes de interesse
- Treinamento do classificador
 - Coleta de amostras
- Classificação
- Estimativa de erros
 - Coleta de amostras para teste

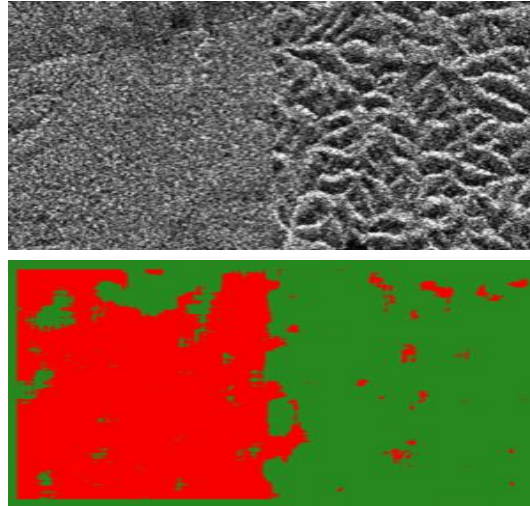
39

Processo de Classificação Supervisionada



40

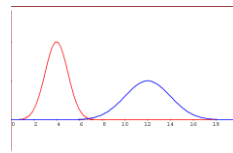
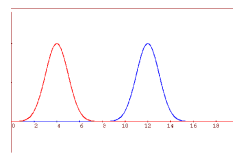
Escolha dos atributos



41

Qualidade da Classificação

- Natureza estatística das classes
- Grau de afastamento das hipóteses
- Qualidade das amostras
 - Representatividade
 - Tamanho
 - Independência
- Amostragem
 - Aleatória
 - Sistemática



42

Classificador Paralelepípedo

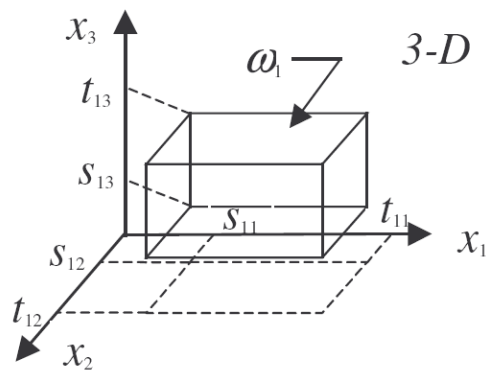
- Define fronteiras de decisão para cada classe
- Entrada
 - Amostras de treinamento
 - $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$
 - $\{w_1, w_2, \dots, w_k\}$
- Saída
 - Limites para cada atributo

43

Classificador Paralelepípedo

- Fronteiras de decisão

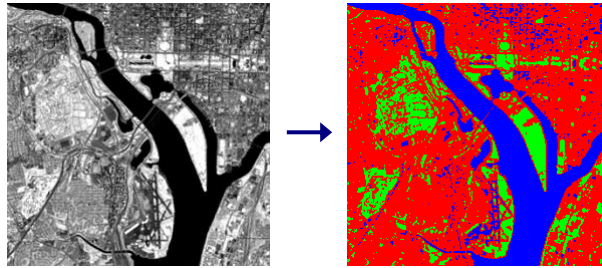
$$\mathbf{x} \in \omega_k \Leftrightarrow x_i \in [s_{ki}, t_{ki}], i = 1, \dots, n$$



44

Classificador Paralelepípedo

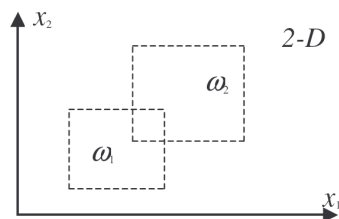
- Exemplo 1-D
 - $[0, 60) \rightarrow$ classe 1
 - $[60, 192) \rightarrow$ classe 2
 - $[192, 255] \rightarrow$ classe 3



45

Classificador Paralelepípedo

- Problema
 - Confusão entre classes
- Possíveis soluções
 - Não classificar
 - Múltiplas classes
 - Sorteio
 - Distância ao centro do paralelepípedo



46

Teoria de Decisão

- Uso de funções de decisão: $g(\mathbf{x}) \rightarrow$ grau de pertencimento
 - $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$
 - $\{w_1, w_2, \dots, w_k\}$
- Meta \rightarrow Encontrar k funções de decisão

$$g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_k(\mathbf{x})$$

- Regra de Classificação

$$\mathbf{x} \in \omega_i \Leftrightarrow g_i(\mathbf{x}) > g_j(\mathbf{x}), \forall j \neq i$$

- Fronteiras de decisão entre classes

$$g_{ij}(\mathbf{x}) = g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0$$

47

Classificador por Distância Euclidiana

- Representante de cada classe w_j
 - Vetor de médias

$$\mathbf{m}_j = \frac{1}{N_j} \sum_{\mathbf{x} \in \omega_j} \mathbf{x}$$

- Distância Euclidiana

$$d_j(\mathbf{x}) = \sqrt{\sum_{i=1}^n (x_i - m_i)^2}$$

- Padrão é classificado segundo a regra

$$\mathbf{x} \in \omega_i \Leftrightarrow d_i(\mathbf{x}) < d_j(\mathbf{x}), \forall j \neq i$$

48

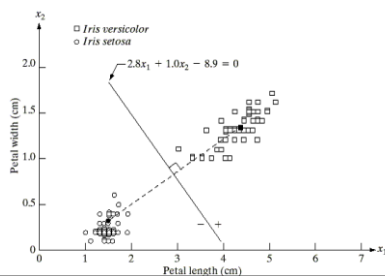
Classificador por Distância Euclidiana

- Regra de decisão

$$g_i(\mathbf{x}) = \mathbf{x}^t \mathbf{m}_i - \frac{1}{2} \mathbf{m}_i^t \mathbf{m}_i$$

- Fronteira de Decisão

$$g_{ij}(\mathbf{x}) = \mathbf{x}^t (\mathbf{m}_i - \mathbf{m}_j) - \frac{1}{2} (\mathbf{m}_i - \mathbf{m}_j)^t (\mathbf{m}_i + \mathbf{m}_j) = 0$$



49

Exercício – Duas classes

- Amostras de pixels da **Classe 1**
 - X1 = [1, 7, 7, 12]
 - X2 = [3, 5, 7, 4]
- Amostras de pixels da **Classe 2**
 - X3 = [4, 8, 9, 10]
 - X4 = [6, 6, 15, 20]
- N = ?
- Total de atributos = ?
- Vetor de medias = ?
- Classificar os seguintes pixels usando distância Euclideana
- E1 = [12, 4, 5, 5]
- E2 = [1, 2, 4, 9]

$$\mathbf{m}_j = \frac{1}{N_j} \sum_{\mathbf{x} \in \omega_j} \mathbf{x}$$

$$d_j(\mathbf{x}) = \sqrt{\sum_{i=1}^n (x_i - m_i)^2}$$

$$\mathbf{x} \in \omega_i \Leftrightarrow d_i(\mathbf{x}) < d_j(\mathbf{x}), \forall j \neq i$$

50

Respostas

- $M1 = [2, 6, 7, 8]$
- $M2 = [5, 7, 12, 15]$
- $d(E1, M1) = 10.81$
- $d(E1, M2) = 14.38$
- $d(E2, M1) = 5.19$
- $d(E2, M2) = 11.87$

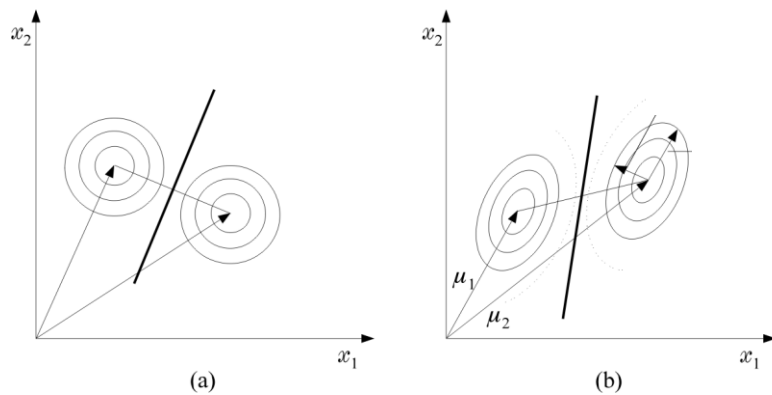
51

Distância Mahalanobiana

- Representantes de cada classe w_j
 - Vetor de médias
 - Matriz de covariâncias
- Distância de Mahalanobis
$$d_j^2(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_j\|_{\Sigma_j}^2 = (\mathbf{x} - \mathbf{m}_j)^T \Sigma_j^{-1} (\mathbf{x} - \mathbf{m}_j)$$
$$d_j(\mathbf{x}) = \sqrt{d_j^2(\mathbf{x})}$$
- Padrão é classificado segundo a regra
$$\mathbf{x} \in \omega_i \Leftrightarrow d_i(\mathbf{x}) < d_j(\mathbf{x}), \forall j \neq i$$

52

Euclides x Mahalanobis



53

Distância de Bhattacharyya

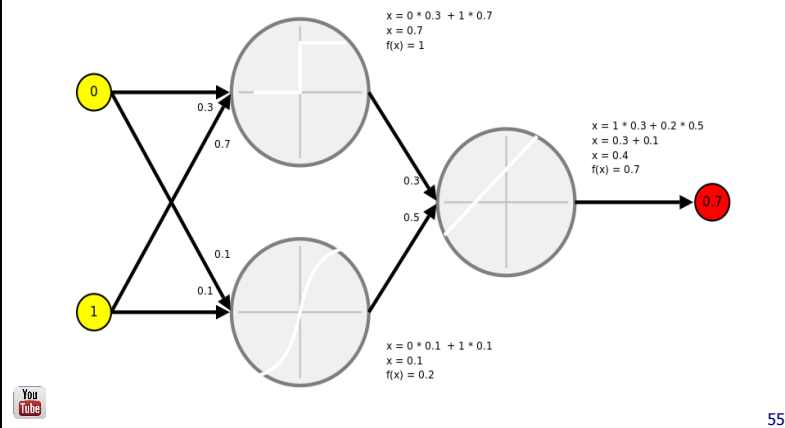
- Representantes de cada classe w_j
 - Vetor de médias
 - Matriz de covariâncias
- Distância de Bhattacharyya

$$B = \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \frac{\left| \frac{\Sigma_i + \Sigma_j}{2} \right|}{\sqrt{|\Sigma_i| |\Sigma_j|}}$$

54

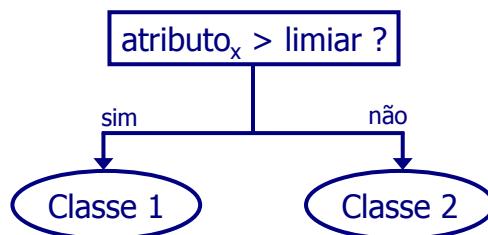
Redes neurais artificiais

- Geram separadores não lineares
- <http://bit.ly/neural-networks>



Árvores de Decisão

- Múltiplos estágios de decisão
- Espaço de atributos particionado em forma de árvore
- Gera fronteiras paralelas aos eixos
 - Regras de decisão



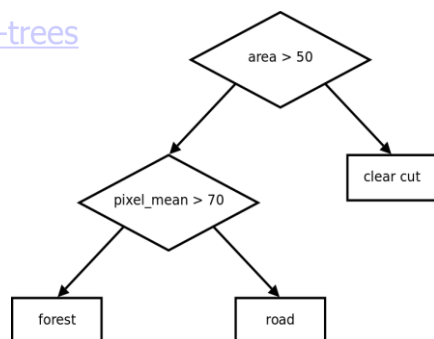
Árvores de Decisão

- Premissas
 - Se todos os casos forem da mesma classe, a árvore se torna uma folha
 - Para cada atributo, calcular sua potencial *separabilidade*
 - Calcular o *ganho de informação* ao testar um atributo
- Cálculo da entropia → Medida de desordem
- Operações de poda
 - Árvore mais generalista

57

Árvores de Decisão

- Independência de:
 - Número de atributos
 - Amplitude dos atributos
- Resultados de fácil entendimento
- <http://bit.ly/decision-trees>



Algoritmo C4.5

- Entropia

$$H(p_1, p_2, \dots, p_N) = -p_1 \log p_1 - p_2 \log p_2 \cdots - p_N \log p_N$$

- Valor da Informação

$$info([v_1, v_2, \dots, v_N]) = H\left(\frac{v_1}{D}, \frac{v_2}{D}, \dots, \frac{v_N}{D}\right)$$

$$D = \sum_{i=1}^N v_i$$

- Ganho \rightarrow Vantagem em usar um atributo ao invés de outro

info para todas as classes menos *info* por ramificação

Exemplo de classificação

- 2 atributos

pixel_mean, area

- 3 classes de treinamento

4 amostras para floresta

3 para corte raso

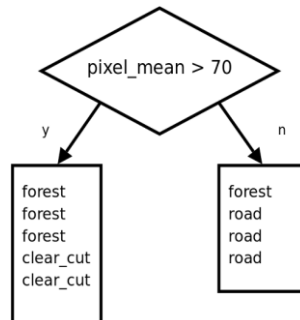
2 para estrada

Usando o atributo *pixel_mean*

$$\text{gain}(\text{pixel_mean}) = \text{info}([4, 2, 3]) - \text{info}([3, 2, 0], [1, 0, 3])$$

↓ ...

$$\text{gain}(\text{pixel_mean}) = 1.5305 - 0.89999 = 0.63051$$

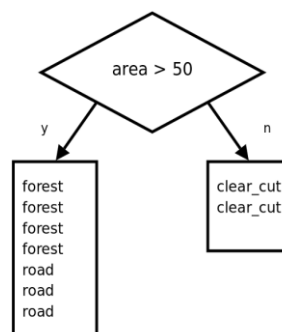


Usando o atributo *area*

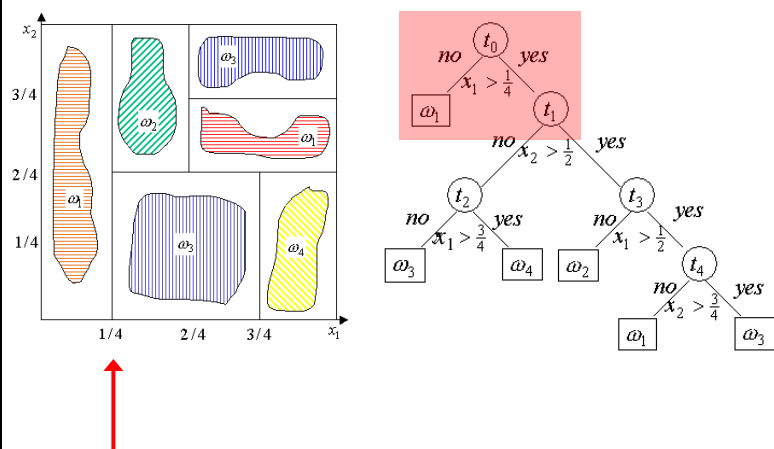
$$\text{gain}(\text{area}) = \text{info}([4, 2, 3]) - \text{info}([4, 0, 3], [0, 2, 0])$$

↓ ...

$$\text{gain}(\text{area}) = 1.5305 - 0.76629 = 0.76421$$



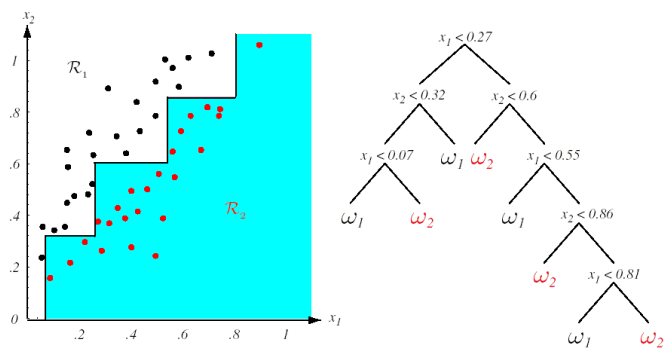
Exemplo



63

Árvores de Decisão

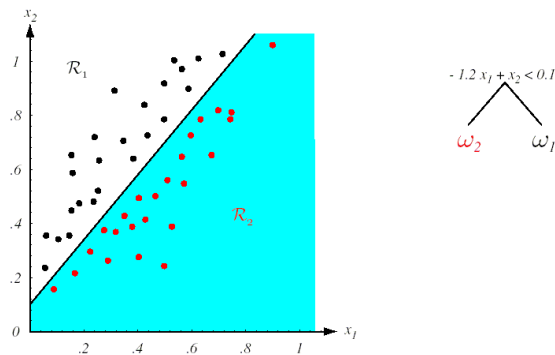
- Problema
 - Classes correlacionadas



64

Árvores de Decisão

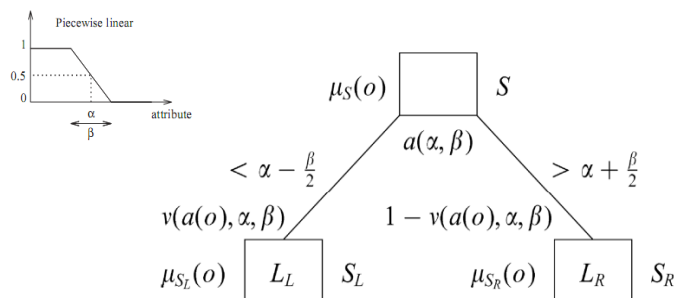
- Solução
 - Árvores oblíquas



65

Árvores de Decisão Fuzzy

- Extensão das árvores tradicionais, incluindo limiares não rígidos



Classificadores Estatísticos

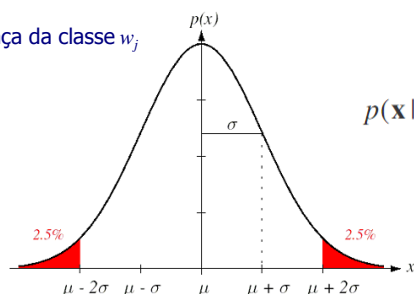
- MAXVER → Máxima verossimilhança

$$\mathbf{x} \in \omega_i \Leftrightarrow p(\mathbf{x} | \omega_i) > p(\mathbf{x} | \omega_j), \forall j \neq i$$

$$p(x | \omega_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{(x - m_j)^2}{2\sigma_j^2}\right]$$

$m_j \rightarrow$ média
 $\sigma_j \rightarrow$ variância

Verossimilhança da classe w_j



Atributos
Independentes

$$p(\mathbf{x} | \omega_i) = \prod_{j=1}^n p(x_j | \omega_i)$$

67

Generalização

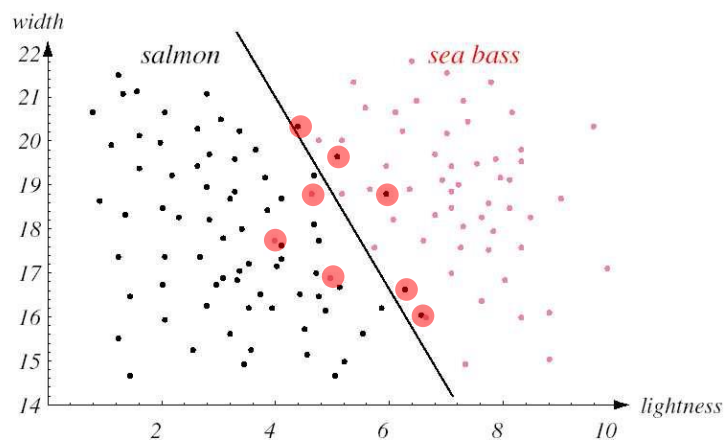
- Compromisso entre
 - Simplicidade do classificador
 - Desempenho sobre os dados de treinamento
- Objetivo
 - Obter maior exatidão sobre novos dados

Occam's razor:

"Prefer the simplest hypothesis that fits the data"

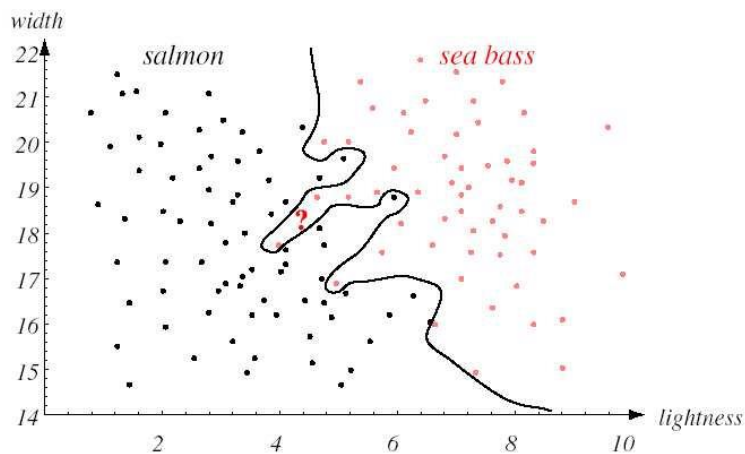
68

Generalização



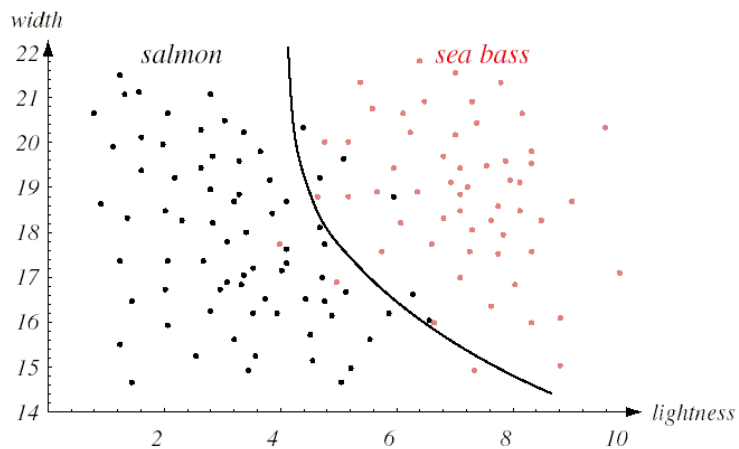
69

Generalização



70

Generalização



71

Referências

- Remote sensing digital image analysis: an introduction (*Richards, 05*)
- Pattern Recognition (*Theodoridis-Koutroumbas, 06*)
- Digital Image Processing (*Gonzalez-Woods, 06*)
- Pattern Recognition Principles (*Tou-Gonzalez, 74*)
- Statistical Pattern Recognition: a review (*Jain, 00*)
- Data Clustering: a review (*Jain, 99*)

72