

Reconhecimento de Padrões

Adrielle

O que é?

“O ato de tomar dados brutos e tomar uma ação baseada na ‘categoria’ do padrão” (Duda, Hart e Stork).

Aplicações no dia-a-dia

Reconhecimento de faces, de fala, de escrita.

Identificar suas chaves só de senti-las no bolso (bolsa).

Saber que uma fruta está madura.

Diferenciar entre animais da mesma espécie.

Identificação de impressão digital.

Como uma máquina faz reconhecimento de padrões?



NATUREZA



Conhecimento
do Processo!!!

Como se dá?

Medição/aquisição (sensor)



Pré-processamento



Extração das características (padrões)



Classificação

Alguns sistemas de classificação podem usar o sentido contrário!!!

Quais?

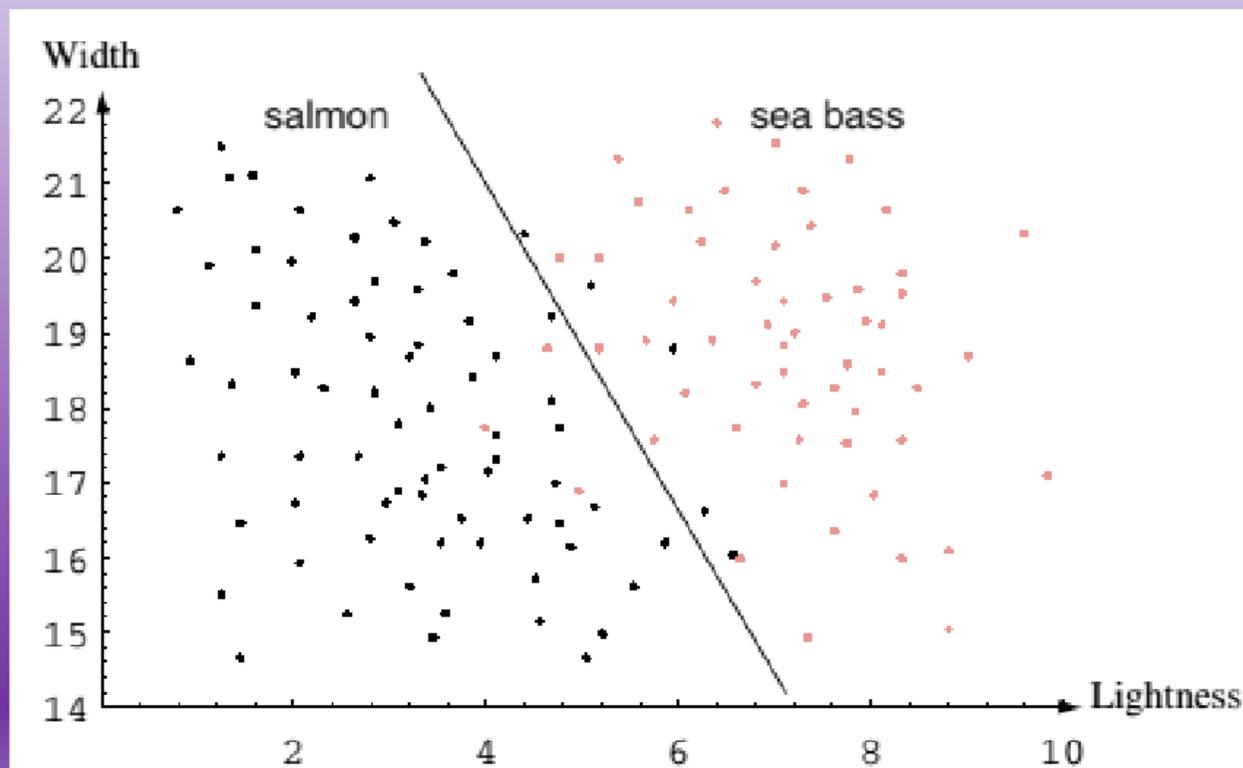
Exemplo

?

Exemplo

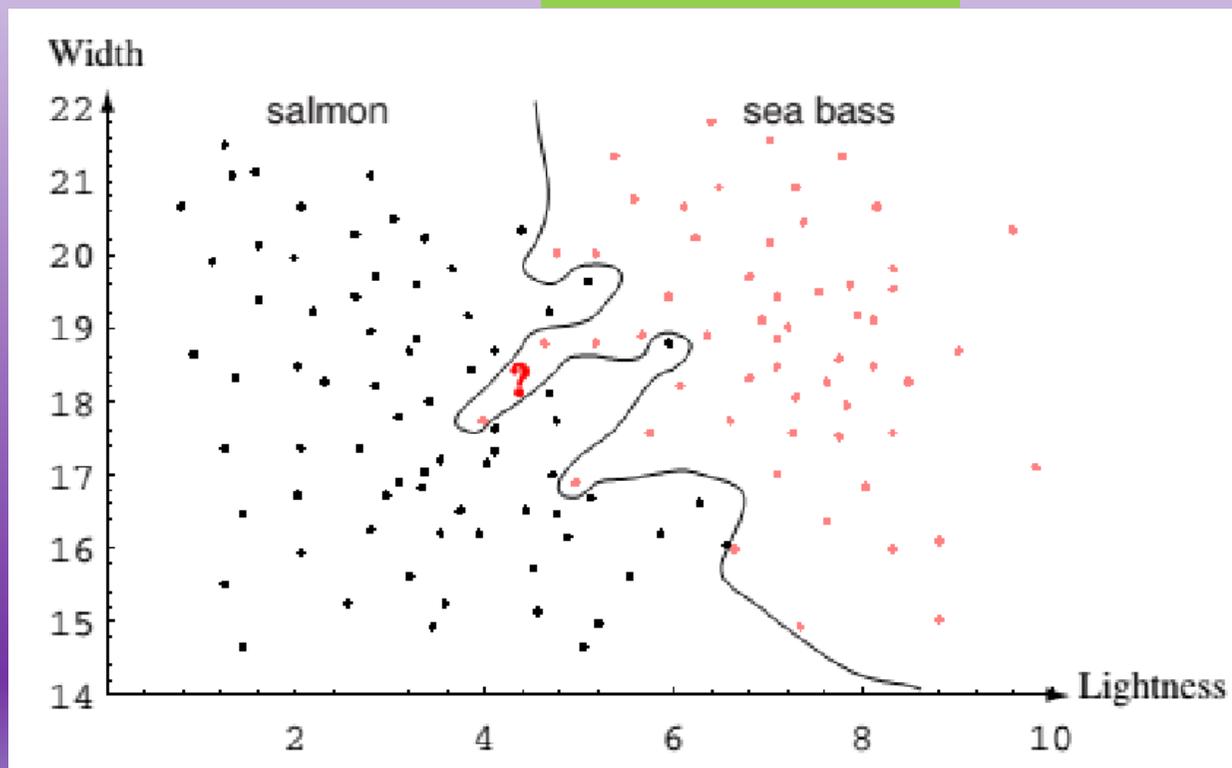


$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$



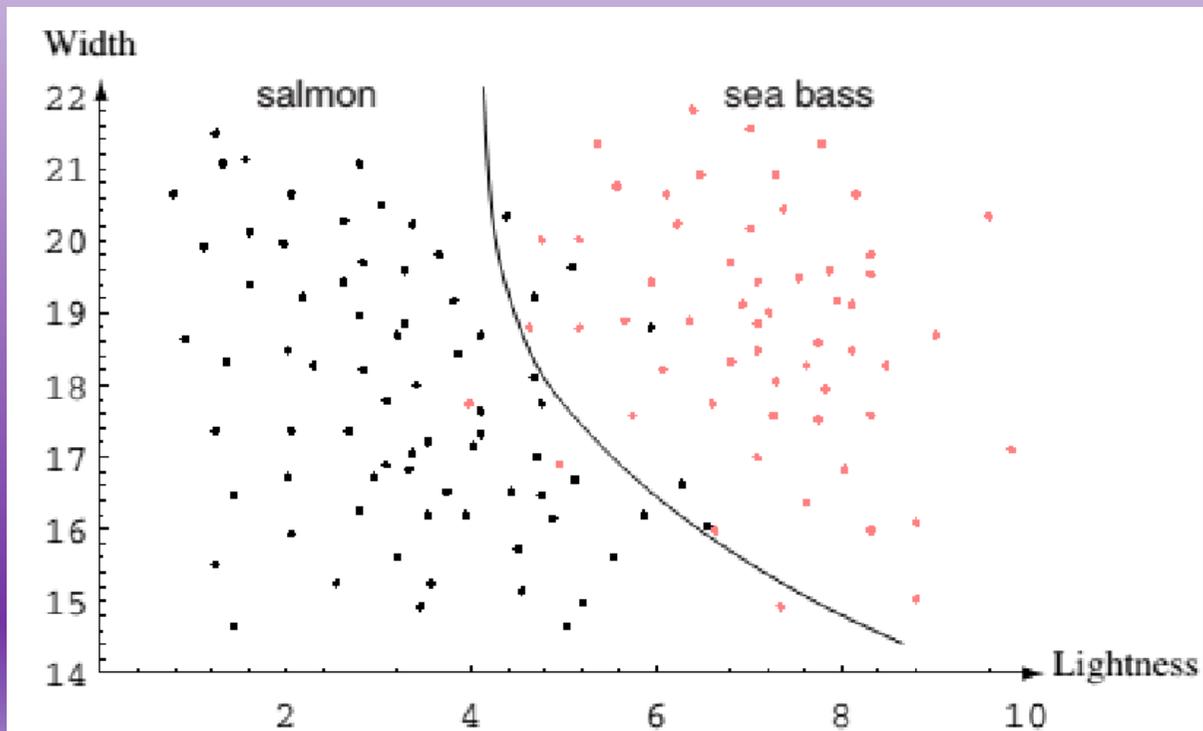
Exemplo

Overfitting!!!



Exemplo

Limite de classificação ótimo!



Deseja-se um classificador com capacidade de **generalização**.

Escolha do classificador

- Quantidade de características disponíveis (dimensões).
- Número de amostras.
- Velocidade necessária na classificação.
- Memória e processador disponível.
- Sensibilidade ao ruído do classificador.
- Entre outros.

Problemas

- Segmentação : como segmentar antes de categorizar? Como categorizar antes de segmentar?
- Overfitting
- Como usar um classificador quando uma ou outra característica de uma amostra está faltando?
- Invariâncias: o classificador não pode ser influenciado pela posição do peixe na esteira. Um reconhecedor de fala não pode ser influenciado pelo tempo de pausa entre uma palavra e outra.
- Dimensionalidade.
- Entre outros.

Tipos de aprendizagem em classificadores

Refere-se a uma forma de algoritmo para a redução do erro em um conjunto de dados de treino.

Aprendizado Supervisionado: Tem-se dados de treino. Um custo é atribuído a cada padrão e busca-se reduzir a soma dos custos para esse padrão no treinamento.

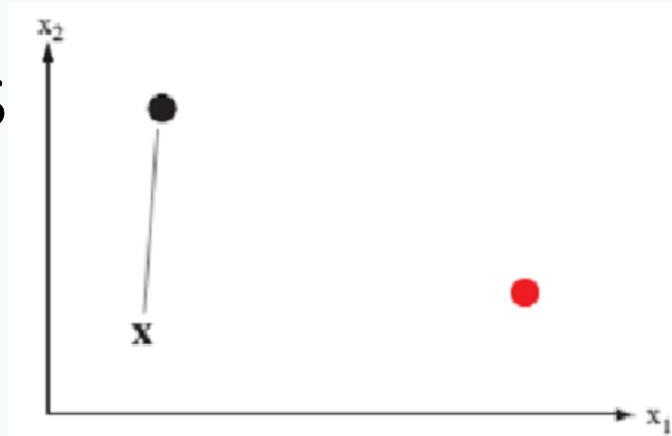
Aprendizado não Supervisionado: Não se tem dados de treino. O classificador forma *clusters* com os padrões de entrada naturalmente criando seu próprio modelo de classificação.

Aprendizado Reforçado: Dados de treino são disponibilizados. O classificador primeiro realiza a classificação. Os dados são usados para comparar se a classificação foi certa ou errada (nesse caso gerando um custo). O treinamento é feito até o custo chegar a um valor baixo o suficiente ou zerar.

Técnicas

- Histogramas
- Análise em Discriminante Linear (LDA) - Discriminante de Fisher
- Máquina de Vetor de Suporte (SVM)
- Análise em Componentes Principais (PCA)
- Análise em Componentes Independentes (ICA)
- Janelas de Parzen
- kNN
- RNN - Perceptron
- K-Means
- Etc.

Exemplo MATLAB – K-Médias



Aprendizado Não-supervisionado

1. Define-se o número desejado de classes. Ex.: 2
2. Escolhe-se a posição inicial dos centroides. Podem ser duas amostras sem que sejam substituídas no processo.
3. Calcula-se a matriz de distâncias Euclidianas (ou outra métrica) de cada amostra a cada centroide.

Métrica Minkowski

$$L_k(\mathbf{a}, \mathbf{b}) = \left(\sum_{i=1}^d |a_i - b_i|^k \right)^{\frac{1}{k}}$$



Distância Euclidiana

$$D(\mathbf{a}, \mathbf{b}) = \left(\sum_{k=1}^d (a_k - b_k)^2 \right)^{\frac{1}{2}}$$

Exemplo MATLAB – K-Médias

4. Pela matriz, analisa-se de qual centroide, cada amostra está mais próxima e atribui-se à amostra o rótulo da classe daquele centroide.
5. Tira-se a média das distâncias de todas as amostras de cada classe até o centroide daquela classe. Para cada dimensão.
6. A média encontrada em cada dimensão se torna a nova coordenada daquele centroide que é então reposicionado.
7. Recalcula-se a matriz de distâncias.
8. Analisa-se se a amostra continua perto do novo centroide de sua classe ou está mais próxima do centroide da outra classe e reclassifica-se a amostra.

Exemplo MATLAB – K-Médias

9. Em caso de troca de classe pela amostras, retorna-se ao passo 3. Quando todas as amostras não trocarem mais de classe a classificação é concluída.

MATLAB

Referências

- Duda, R. O.; Hart, P. E. e Stork, D. G.; **Pattern Classification**, 2000. WileyInterscience, 2 ed., 680p..