RECONHECIMENTO DE PADRÕES(MIM)

Faculdade de Ciências da Universidade do Porto
Hugo Pedro Proença (DI-UBI), 10/04/2010

Página da Disciplina

http://www.di.ubi.pt/~hugomcp/PR

- Slides
- Fichas de Exercícios
- Conjuntos de Dados

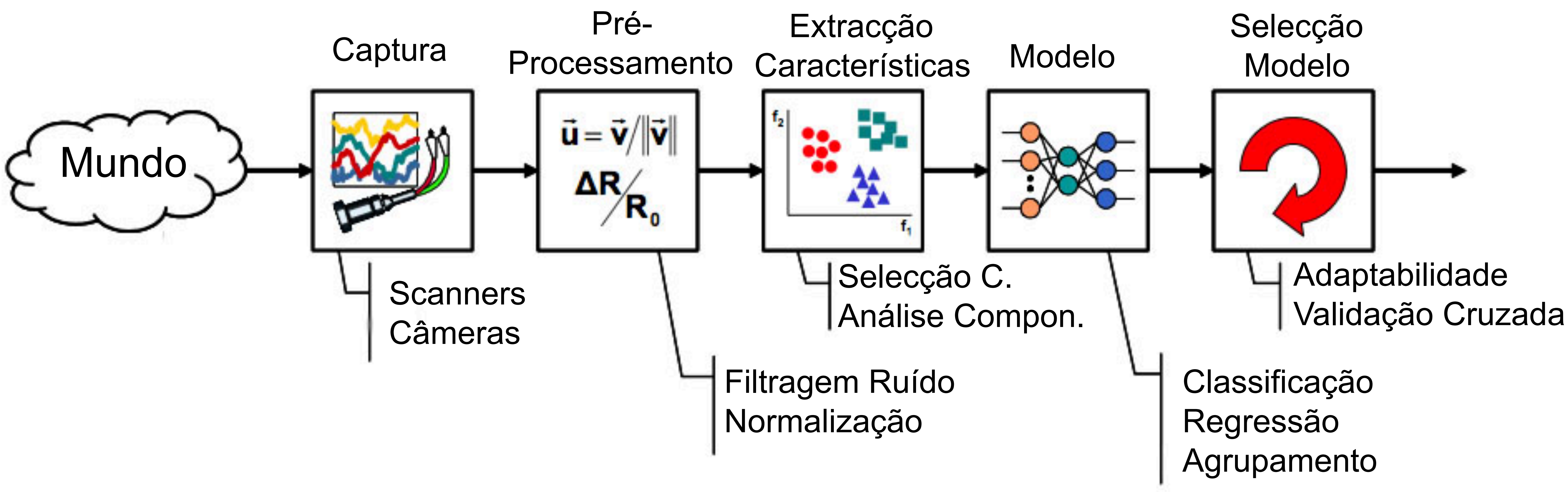
Objectivos

- Este módulo visa a apresentação das noções básicas de um sistema de Reconhecimento de Padrões.
  - Perspectiva de Alto-Nível
  - Definições
  - Estrutura
  - Aplicações
- O ambiente MATLAB (Mathematica) será usado na exemplificação das noções mais importantes.

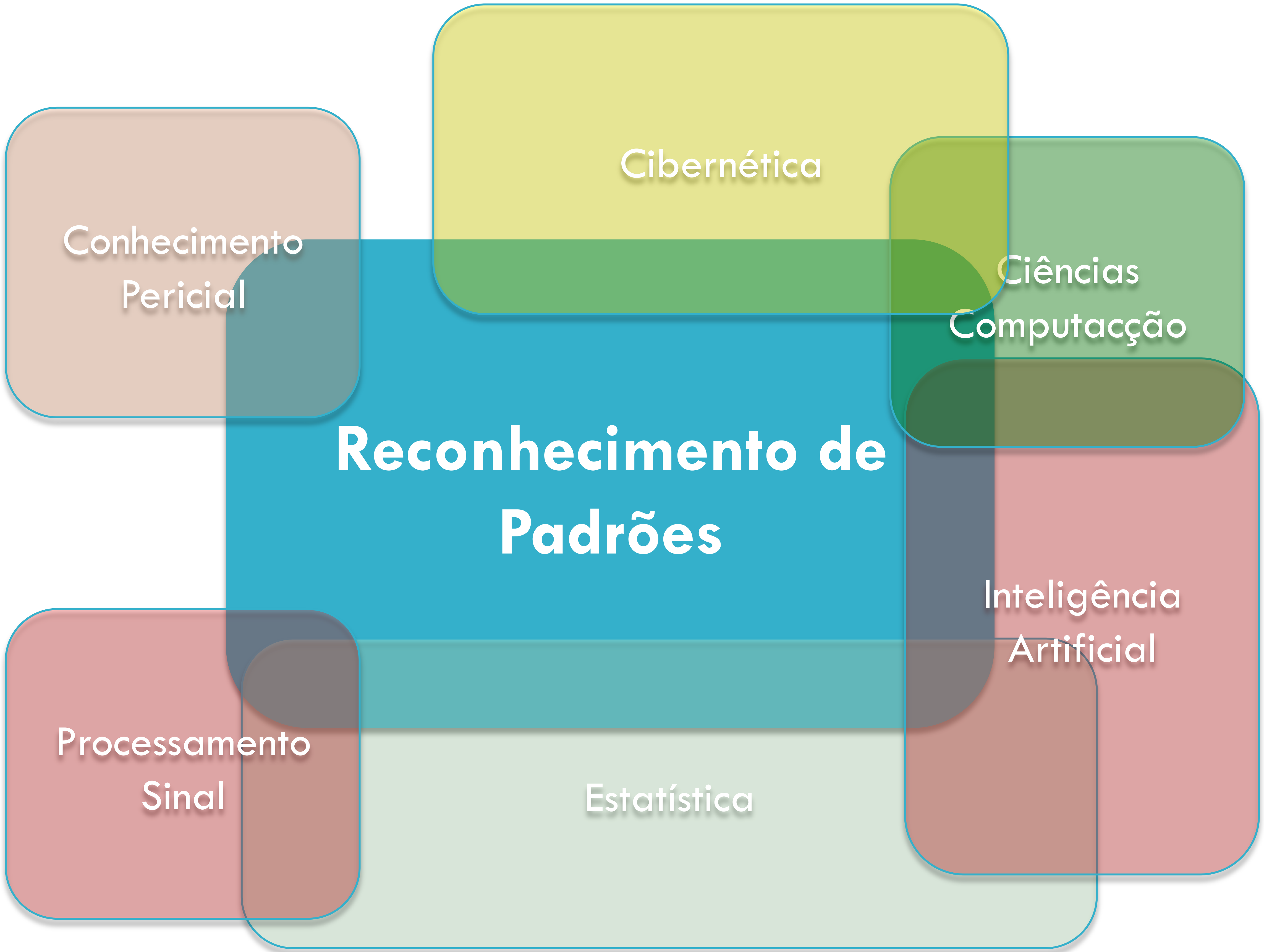
Reconhecimento de Padrões

- *"The assignment of a physical object or event to one of several pre-specified categories"* – Duda and Hart
- *"Given some examples of complex signals and the correct decisions for them, make decisions automatically for a stream of future examples"* – Ripley
- *"The science that concerns the description or classification (recognition) of measurements"* – Schalkoff
- *"The process of giving names ω to observations x"*, – Schürmann
- *"Pattern Recognition is concerned with answering the question "What is this?""* – Morse
- *"The act of taking in raw data and taking an action based on the category of the pattern"* – Duda et al.

Reconhecimento de Padrões



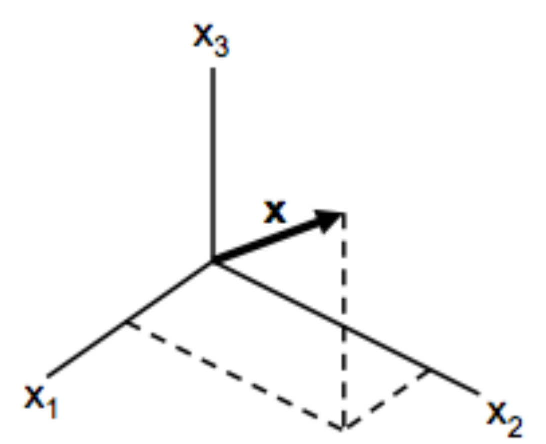
Reconhecimento de Padrões



Características (Features)

- A noção de **característica (feature)** é **<u>a fundamental</u>** do reconhecimento de padrões.
  - Medição (qualitativa ou quantitativa) do objecto de interesse.
- A combinação de "d" características é representada por 1 vector d-dimensional (vector de características).
- O espaço d-dimensional definido pelos vectores de caracterísiticas é chamado o espaço de características.
- Os objectos são representados como pontos no espaço de características.

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



Características (Features)

- Em abstracto, um problema de reconhecimento de padrões é representado por uma matriz de dados, com "n" instâncias (observações) e "d" características:

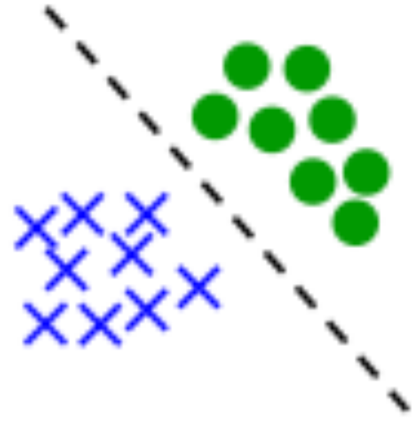
Instâncias

Obs	Caract. 1	Caract. 2	…	Caract. d	CLASSE
1	X11	X12		X1d	A
2	X21	X22		x2d	B
…	…	…	…	…	…
N	Xn1	Xn2		Xnd	C

Características (Features)

- O que são boas características?
  - A **qualidade** de um vector de características é dada pela sua capacidade para **discriminar** os exemplos de diferentes classes.
    - Exemplos da mesma classe deverão ter valores semelhantes para cada característica.
    - Exemplos de classes diferentes deverão ter valores evidentemente diferentes.

Boas Características



Más Características

Características (Features)

- Suponhamos que pretendíamos discriminar entre jogadores de basket (X) e jogadores de xadrez (O).
- Característica "Altura":

OOO O O X OOO O XX X X O XX XX X X X X

1,6 1,7 1,8 1,9 2,0 2,1 2,2 2,3

- Característica "Côr dos Olhos":

OOXXOXOOX XXOXOXOO OXXXOX

Claros Médios Escuros

Características (Features)

- As características a extrair para um dado problema são bastantes dependentes do **conhecimento pericial** e determinam em grande medida o sucesso do sistema de reconhecimento de padrões.
  - Quantitativas
    - Contínuas (Exlº: peso, altura, temperatura…)
    - Discretas (Exlº: nº de vértices, total de arestas,…)
  - Qualitativas
    - Exlº nível de nublosidade (muito/pouco nublado, limpo)

Características (Features)

- Para as características qualitativas, é comum atribuir uma etiqueta numérica, sendo posteriormente normalizadas como tal.

Observ.	Nublosidade	Temperatura
1	Pouco nublado	25.2
2	Limpo	27.3
3	Muito Nublado	17.2

Observ.	Nublosidade	Temperatura
1	2	25.2
2	1	27.3
3	3	17.2

Características (Features)

- Normalmente, diferentes características têm unidades de grandeza muito diferentes e distribuem-se por intervalos heterogéneos.
- O processo de **normalização** consiste em uniformizar esses intervalos, por forma a que o processo de classificação não sofra desvios resultantes de diferentes unidades de grandeza.
  - **Min-Max:** Passagem para intervalo fixo ([0,1]). Ao maior valor observado atribui-se "1", ao menor "0". Os restantes recebem valores de forma linear.
  - **Média-Variância:** Igual média e variância. Calcula-se a média e variância dos valores observados. Os novos valores são dados pelo valor antigo subtraído da média e dividido pela variância.

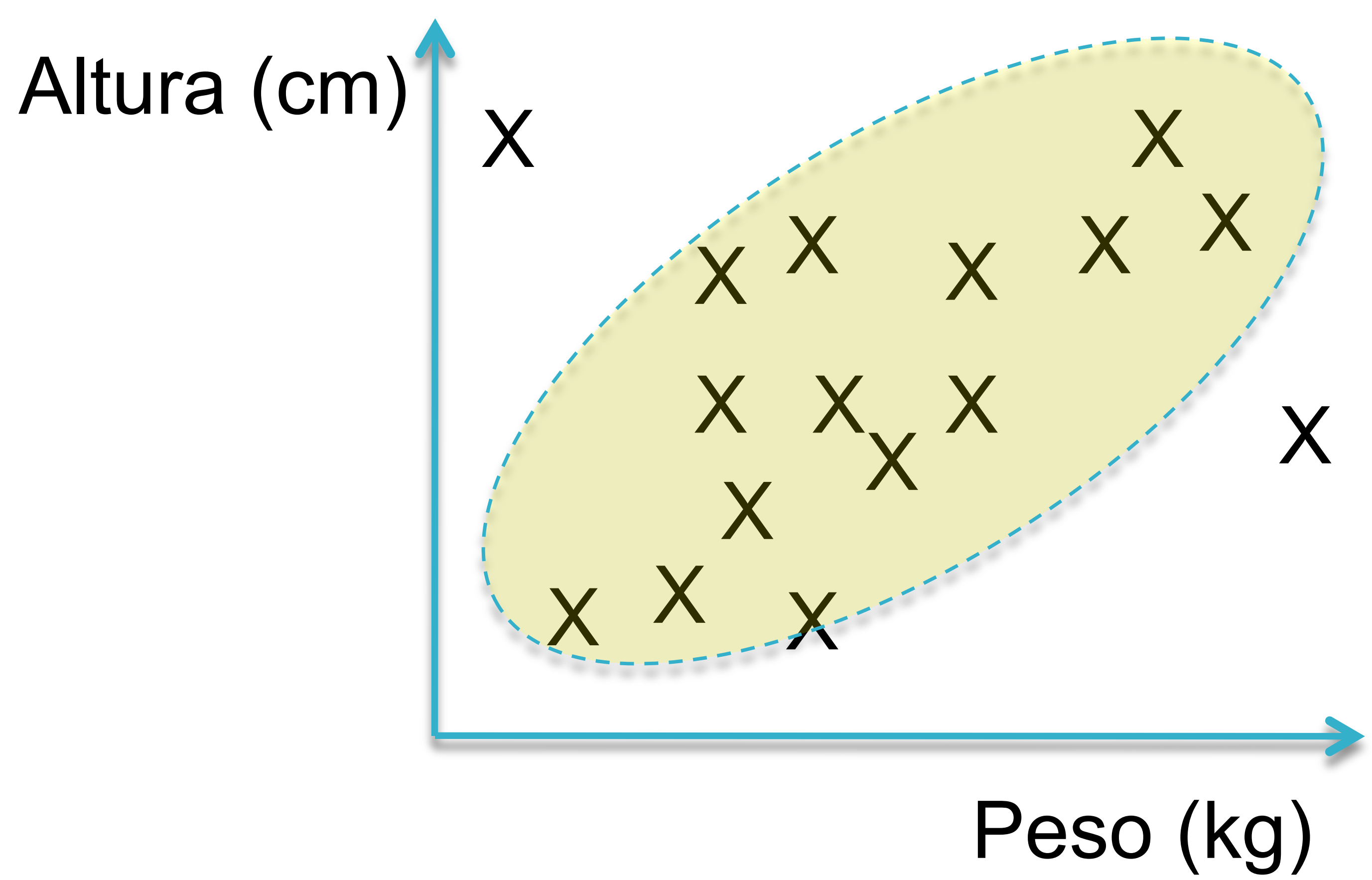
Normalização: Exercício

- Considere a seguinte matriz de dados, referente a uma estatística sobre a produtividade dos empregados de uma empresa de entrevistas. Normalize-a das duas formas diferentes descritas no acetato anterior.

Obs.	Salário	Idade	Peso	Côr Olhos
1	1000	27	75	Castanhos
2	2100	26	86	Verdes
3	750	42	68	Castanhos
4	1562	29	65	Azuis
5	2100	23	56	Castanhos

Características e Dependência

- De entre o conjunto de características julgadas relevantes para um problema, é comum que algumas delas "variem em conjunto".
  - Por exemplo, se fizermos a medição do peso e altura de um conjunto de pessoas, será provável que – em média – as pessoas maiores tenham mais peso.



Características e Dependência

- A matriz de co-variância fornece uma medida da **variação relativa** de cada par de características.
- Uma matriz de dados com "d" características dá origem a uma matriz de co-variância de "d" linhas e "d" colunas.
  - Simétrica em relação à diagonal principal, Valor(x,y)=Valor(y,x)

v1	c12	c13	c14	c15
c21	v2	c23	c24	c25
c31	c32	v3	c34	c35
c41	c42	c43	v4	c45
c51	c52	c53	c54	v5

Características e Dependência

- Para calcular a matriz de co-variância começa-se por calcular o vector de médias (por característica):

$$E[X] = \left[E[X_1]E[X_2]\ldots E[X_N]\right]^T = \left[\mu_1 \mu_2 \ldots \mu_N\right] = \mu$$

- A matriz é dada por:

$$COV[X] = \Sigma = E[(X-\mu)(X-\mu)^T]$$

$$= \begin{bmatrix} E[(x_1-\mu_1)(x_1-\mu_1)] & \ldots & E[(x_1-\mu_1)(x_N-\mu_N)] \\ \ldots & \ldots & \\ E[(x_N-\mu_N)(x_1-\mu_1)] & \ldots & E[(x_N-\mu_N)(x_N-\mu_N)] \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \ldots & c_{1N} \\ \ldots & \ldots & \\ c_{1N} & \ldots & \sigma_N^2 \end{bmatrix}$$

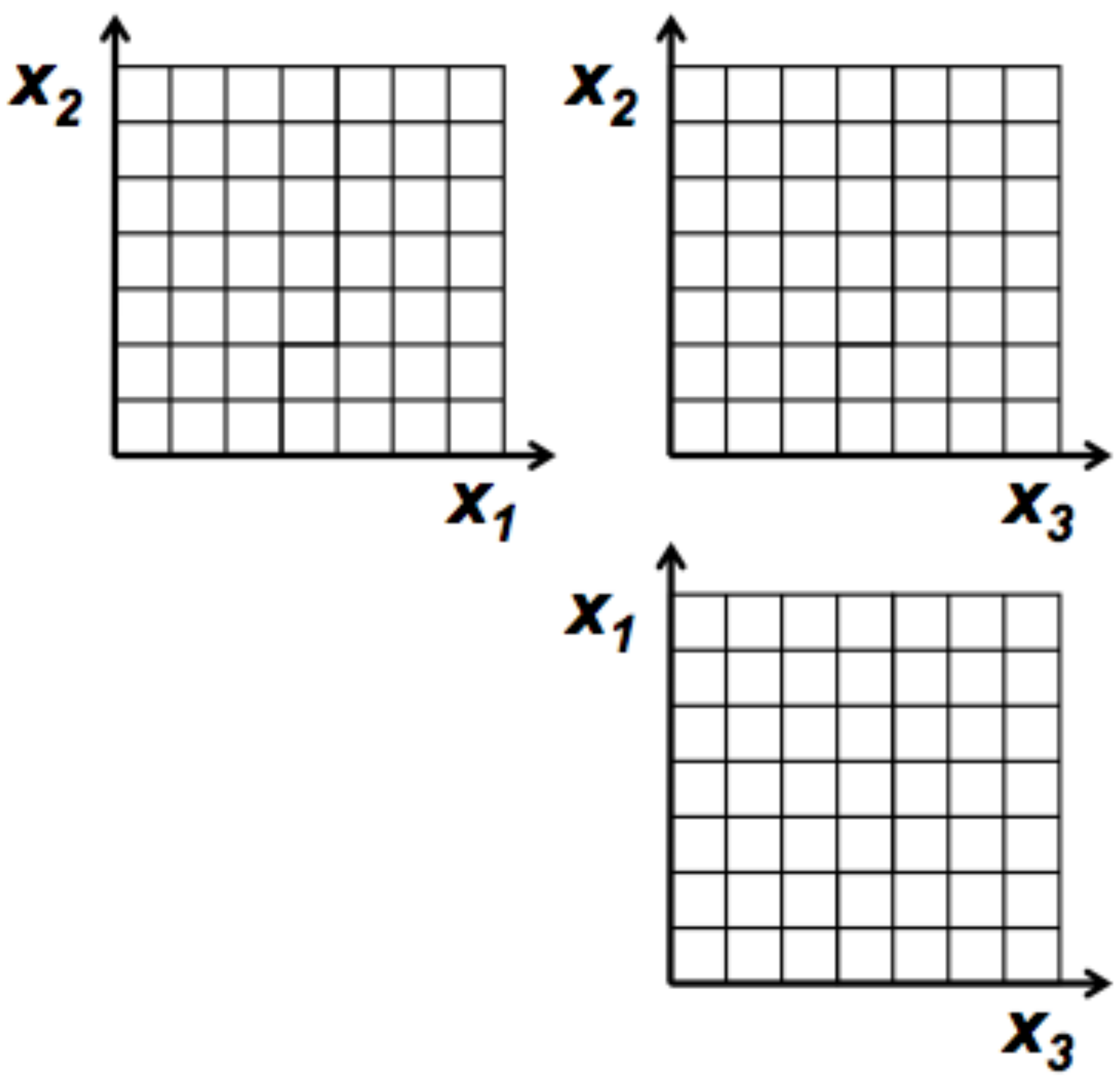
Características e Dependência

- A matriz de co-variâncias tem propriedades importantes, uma vez que permite inferir padrões nos dados:
  - Quando x(i) e x(j) aumentam/diminuem conjuntamente, c(ij) >0
  - Quando x(i) e x(j) variam inversamente, c(ij)<0
  - Se x(i) e x(j) são independentes, c(ij)=0
  - Os valores na diagonal principal expressam a variância da característica.

Características e Dependência

- Exercício. Calcule a matriz de Co-Variância para o seguinte conjunto de dados:

Obs.	X1	X2	X3
1	2	2	4
2	3	4	6
3	5	4	2
4	6	6	4



	x_1	x_2	x_3	$x_1-\mu_1$	$x_2-\mu_2$	$x_3-\mu_3$	$(x_1-\mu_1)^2$	$(x_2-\mu_2)^2$	$(x_3-\mu_3)^2$	$(x_1-\mu_1)(x_2-\mu_2)$	$(x_1-\mu_1)(x_3-\mu_3)$	$(x_2-\mu_2)(x_3-\mu_3)$
1												
2												
3												
4												

Classificador Bayesiano

- O Teorema de Bayes é fundamental na variante estatística de reconhecimento de Padrões.
  - Utilizando a noção de probabilidade condicional e o teorema de probabilidade total, o Reverendo Thomas Bayes (1702-1761) propôs como forma de classificação a maximização da probabilidade à-posteriori:

$$P[\omega_j \mid x] = \frac{P[x \mid \omega_j] \cdot P[\omega_j]}{\sum_{k=1}^{N} P[x \mid \omega_k] \cdot P[\omega_k]} = \frac{P[x \mid \omega_j] \cdot P[\omega_j]}{P[x]}$$

Classificador Bayesiano

- $P(A|B)$: Probabilidade condicional. É a probabilidade de ocorrer "A", sabendo que "B" ocorreu. É dada por:

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

- $P(W_i)$: Probabilidade à-priori.
- $P(W_i|x)$: Probabilidade à-posteriori. Probabilidade de ocorrer a classe "W_i", para uma dada observação "x" (ocorrida).
- $P(x|W_i)$: Função de (densidade de) probabilidade. Probabilidade de ocorrência de valores de "x" na classe "W_i".
- $P(x)$: Probabilidade de ocorrência de uma observação.

Classificador Bayesiano

- Assim, um exemplo é classificado na classe "k" se essa classe maximizar as probabilidades à-posteriori, isto é:

$$\mathbf{k = \arg\max_{j} P[W_i \mid x]}$$

- Na prática, calculam-se as probabilidades à-posteriori para cada classe e assume-se a classe com maior valor.

Classificador Bayesiano

- Em termos práticos, o modelo de classificação Bayesiano tem um problema sério: a aproximação das funções de probabilidade $P(x|W_i)$.
  - Mais evidente em problemas de elevada dimensionalidade.
  - Exigência de conjuntos de dados de dimensões gigantescas, ou mesmo impraticáveis.
  - Como obter um valor fiável para $P([x1,x2,...xd]|W_i)$?

Naive-Bayes

- O classificador Naïve-Bayes é baseado no teorema Bayesiano e é particularmente útil quando a dimensionalidade do espaço de características é elevada.
- Funciona como uma simplificação (aproximação) ao problema cujo desempenho é comummente comparável aos métodos de classificação mais sofisticados.
- A ideia fundamental é considerar os valores de cada característica (xi) independentes.
  - $P([x1,x2,\ldots xd]|W_i)=P(x1|W_i) \times P(x2|W_i) \times \ldots P(xd|W_i)$

Classificador Bayesiano

- Exercício: A partir do seguinte conjunto de dados pretende-se construir 1 sistema de reconhecimento de padrões para auxílio no diagnóstico de gripe:

Training Example	N (running nose)	C (coughing)	R (reddened skin)	F (fever)	Classification
d_1	+	+	+	−	positive (ill)
d_2	+	+	−	−	positive (ill)
d_3	−	−	+	+	positive (ill)
d_4	+	−	−	−	negative (healthy)
d_5	−	−	−	−	negative (healthy)
d_6	−	+	+	−	negative (healthy)

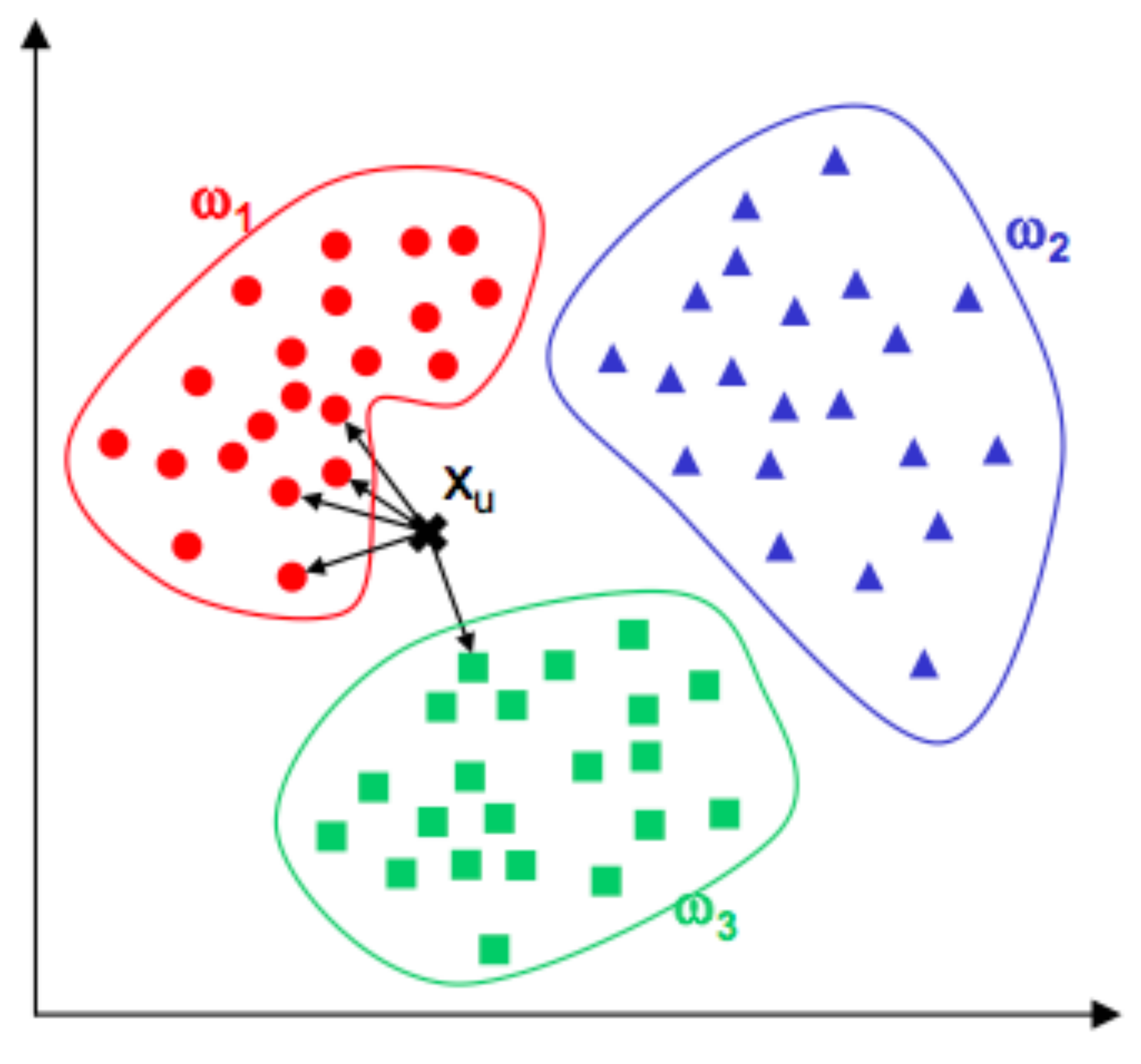
- Qual a classificação dada pelo modelo "Naïve-Bayes" para a observação "+ - - +"?

KNN

- O algoritmo de classificação dos K-vizinhos mais próximos é especialmente simples e intuitivo.
- Classifica 1 exemplo simplesmente de acordo com a sua semelhança em relação ao conjunto de treino.
- Algoritmo:
  - Seja x=[x1,…xd] o exemplo a classificar.
  - Encontrar os K-vizinhos mais próximos de "x", de acordo com uma medida de distância (Euclideana, Mahalanobis, …)
  - Atribuir a "x" a classe mais frequente entre os k-vizinhos.

KNN

- Neste exemplo, sería atribuído ao exemplo “x” a classe (W_1), uma vez que está em maioria entre os seus vizinhos mais próximos.

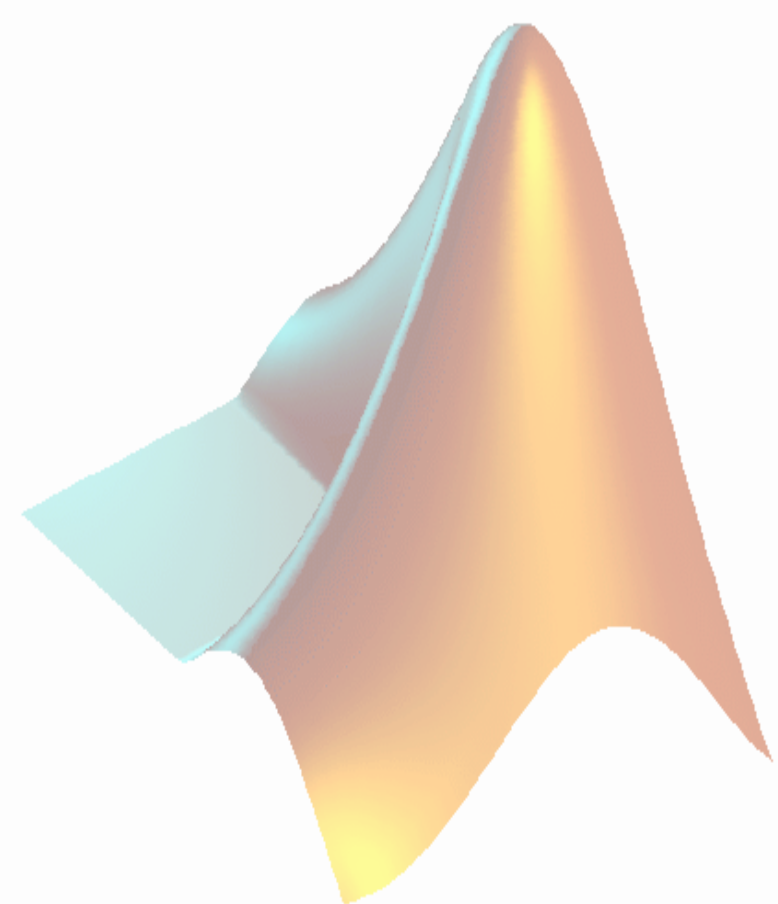


Aplicação Prática

- Suponhamos que um sistema de automático de captura de imagem regista caracteres manuscritos.
  - Para simplificar, caracteres numéricos [0,9]
- Pretendemos desenvolver um sistema de reconhecimento de padrões que seja capaz de "entender" essa informação.
  - Não homogénea, irregular.

Aplicação Prática

- O ambiente *MATLAB* facilita de sobremaneira a interacção "programador (utilizador) ←→ máquina".
  - Linguagem de programação de alto (altíssimo) nível.
  - Conjunto de funções (*toolboxes*) disponíveis.
  - Linguagem interpretada. Código-fonte das funções disponível.
  - Vasta comunidade de utilizadores
    - Google, ...
- *Mathematica*, *R* funcionam como alternativas menos dispendiosas.



Aplicação Prática

- As diferentes imagem de caracteres (30x20) foram previamente gravadas numa matriz binária com dimensões 90.000 (300x30x10) x 20.
  - Linha 1:30 (1° caract), 31:100 (2° caract)...
- Passo 1: Leitura da Matriz
  - >> Dados=load('OCRDataSet.txt');
- Passo 2: Função para aceder individualmente a cada caracter (30x20)

Aplicação Prática

% x=y(a:b,c:d) %Acesso a matrizes

Desde a coluna "c" até "d"

Desde a linha "a" até "b"

Retorno

Parâmetros

Function ret=getImagem(Matriz, indice)

ret= Matriz((indice-1)*30+1:indice*30,1:end);