

**Aprendizagem Automática e Data Mining
2012/2013**

2º teste
– com consulta –
Duração: 1h 30

4 de Junho de 2013

ATENÇÃO: 1) O teste é constituído por **4 questões**. 2) Escreva o seu Nome e Número **em todas as folhas de resposta**, as quais são facultadas pelo docente. 3) Só pode abandonar a sala **45 minutos** após o início do teste. 4) Tem de entregar sempre as folhas de resposta.

Bom trabalho!

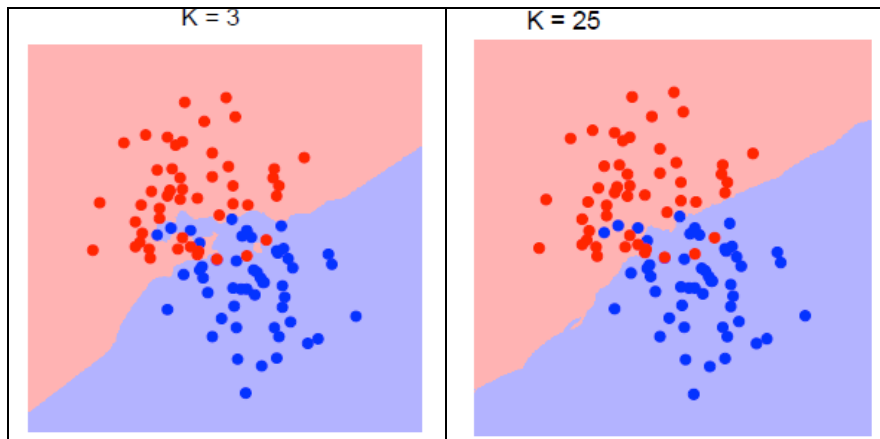
Questão 1 [2.5 + 2.5 valores]

Considere o seguinte conjunto de pontos do espaço bi-dimencional, $X = \{\mathbf{x}_1 = (-0.1, -0.1), \mathbf{x}_2 = (0.7, 0.7), \mathbf{x}_3 = (1.0, 1.0), \mathbf{x}_4 = (1.6, 1.6), \mathbf{x}_5 = (2.0, 2.0), \mathbf{x}_6 = (2.5, 2.5), \mathbf{x}_7 = (3.2, 3.2), \mathbf{x}_8 = (3.5, 3.5), \mathbf{x}_9 = (4.1, 4.1), \mathbf{x}_{10} = (4.9, 4.9)\}$. Estes dados encontram-se classificados por uma classe binária com valores '+' e '-', definindo o conjunto $D = \{(\mathbf{x}_1, '-'), (\mathbf{x}_2, '+'), (\mathbf{x}_3, '+'), (\mathbf{x}_4, '-'), (\mathbf{x}_5, '+'), (\mathbf{x}_6, '+'), (\mathbf{x}_7, '-'), (\mathbf{x}_8, '-'), (\mathbf{x}_9, '+'), (\mathbf{x}_{10}, '+')\}$.

1.1. Com o conjunto de dados D treinou-se um classificador K -NN com distância Euclideana não pesada e parâmetro $K=3$.

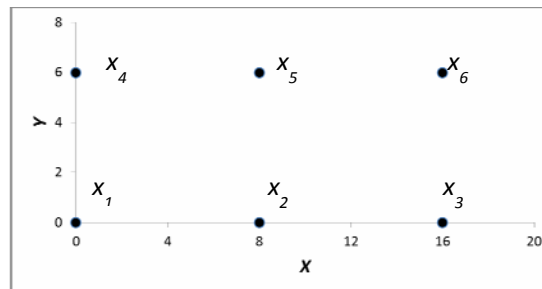
Calcule o valor da classe mais votada para o ponto $\mathbf{x}_2$, considerando esse ponto como exemplo de teste. Apresente todos cálculos auxiliares.

1.2. Considere as seguintes figuras onde se apresentam um conjunto de pontos pertencentes a duas classes distintas. Os pontos da classe C_1 estão assinalados a cinzento e os pontos da classe C_2 estão assinalados a preto. Cada figura mostra os resultados de classificação obtidos com o classificador K -NN para valores de $K=3$ e $K=25$, respectivamente. Explique, referindo-se à(s) propriedade(s) do classificador K -NN, porque é que para o valor de K mais elevado se obtém uma fronteira de decisão "suavizada".



Questão 2 [5 valores]

Considere o seguinte conjunto de pontos $X = \{\mathbf{x}_1 = (0,0), \mathbf{x}_2 = (8,0), \mathbf{x}_3 = (16,0), \mathbf{x}_4 = (0,6), \mathbf{x}_5 = (8,6), \mathbf{x}_6 = (16,6)\}$, abaixo representados num gráfico XY



O K -means devolve uma K -partição do conjunto X , a qual corresponde a K subconjuntos não vazios. Por exemplo, para $K=3$, podem obter-se os clusters $C_1 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5\}$, $C_2 = \{\mathbf{x}_3, \mathbf{x}_4\}$ e $C_3 = \{\mathbf{x}_6\}$, aos quais correspondem os seguintes valores de pertença b_k^i das entidades aos clusters: $b_1^1 = 1, b_1^2 = 1, b_1^5 = 1$; $b_2^3 = 1, b_2^4 = 1$; $b_3^6 = 1$.

Naturalmente, a uma K -partição está associado um conjunto de K centroides $\{m_k\}_{k=1}^K$

Correu-se o algoritmo K -means para $K=3$, com os seguintes protótipos iniciais: $\mathbf{m}_1 = \mathbf{x}_2$, $\mathbf{m}_2 = \mathbf{x}_3$, $\mathbf{m}_3 = \mathbf{x}_5$. Para eventuais empates na afetação das entidades aos clusters favorecem-se os centróides da esquerda/baixo.

Indique, justificando, quais os clusters encontrados, bem como os correspondentes protótipos.

Questão 3 [2 + 2 valores]

3.1. Enuncie as restrições que definem uma partição difusa. Explique o seu significado.

3.2. Considere o seguinte conjunto de pontos $X = \{\mathbf{x}_1 = (0,0), \mathbf{x}_2 = (2,0), \mathbf{x}_3 = (0,3), \mathbf{x}_4 = (2,3)\}$. Correu-se o algoritmo Fuzzy c -means com parâmetros $c=2$; e $m=2.0$. O algoritmo converge para uma partição difusa com protótipos $\mathbf{v}_1 = (1,0)$ e $\mathbf{v}_2 = (1,3)$.

(i) Calcule os valores de pertença u_{ik} das entidades aos clusters.

(ii) Discuta vantagens do algoritmo Fuzzy c-means face à partição encontrada.

Questão 4 [4 + 2 valores]

Considere a seguinte matriz de dissimilaridade entre 5 pontos:

	p1	p2	p3	p4	p5
p1	0.0	1.2	3.0	3.7	4.2
p2	1.2	0.0	2.5	3.2	3.9
p3	3.0	2.5	0.0	1.8	2.0
p4	3.7	3.2	1.8	0.0	1.5
p5	4.2	3.9	2.0	1.5	0.0

4.1. Calcule o resultado da aplicação do algoritmo *Single Link* e apresente-o na forma de um dendograma

4.2. Indique qual a partição obtida pelo algoritmo para $K=3$.