

Departamento de Informática

**Aprendizagem Automática e Data Mining
2011/2012**

**Exame da Época Normal
– com consulta –
Duração: 3h**

29 de Junho de 2012

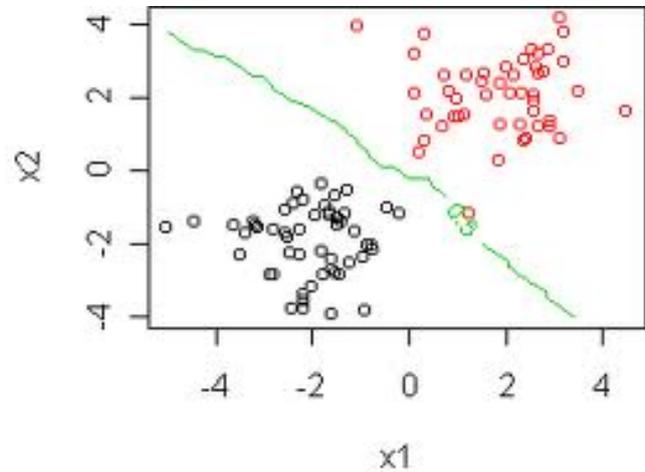
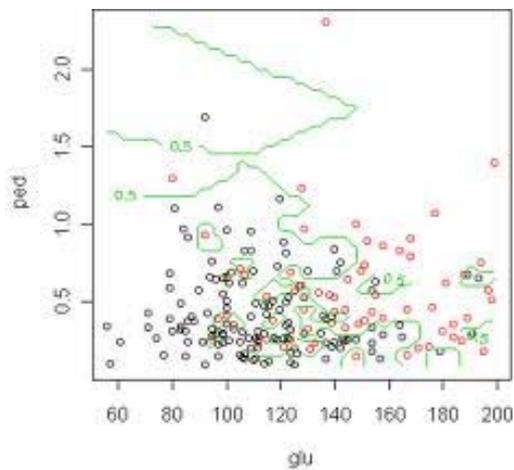
Leia atentamente as seguintes notas. Só depois deve começar o exame.

1. Escreva o seu Nome e Número em todas as folhas de exame, as quais são facultadas pelo docente.
2. Responda a cada questão em FOLHAS SEPARADAS.
3. A simplicidade e clareza das respostas contarão na avaliação.
4. Só pode abandonar a sala **1h** após o início do exame. Tem de entregar sempre as folhas de resposta.
5. Se desistir, deve declarar que desiste e assinar nas folhas de exame.

Bom trabalho!

Questão 1 [1 valor]

Considere as seguintes figuras que representam, segundo 2 atributos, amostras de classes distintas. A cor de cada elemento identifica a classe a que ele pertence.



No contexto da classificação K -NN, quais os valores que escolheria para o parâmetro K de modo a obter a melhor classificação possível para a distribuição representada em cada figura? Justifique.

Questão 2 [2 valores]

Considere a seguinte tabela onde estão registadas amostras correspondentes a dias caracterizados pelos atributos indicados. Um dos atributos (piquenique) representa a classe a que pertence a amostra.

dia	Aspeto (céu)	temperatura	chuva	Vento	piquenique
1	limpo	media	ausente	Fraco	sim
2	limpo	media	Por vezes	Fraco	Não
3	nublado	media	ausente	Fraco	Não
4	limpo	alta	ausente	Forte	Não
5	Pouco nublado	media	ausente	Fraco	sim
6	limpo	media	ausente	Fraco	sim
7	limpo	media	permanente	Ausente	Não
8	limpo	baixa	ausente	Ausente	sim
9	limpo	baixa	ausente	ausente	Não
10	limpo	baixa	ausente	ausente	Não
11	Pouco nublado	baixa	ausente	ausente	Não
12	limpo	media	ausente	forte	Não
13	limpo	media	ausente	ausente	sim
14	nublado	baixa	permanente	ausente	Não
15	Pouco nublado	media	Por vezes	ausente	sim

De acordo com o o classificador Naïve Bayes, qual deverá ser a decisão (piquenique sim/não) a tomar por uma pessoa que de manhã se depara com um dia com as seguintes características:

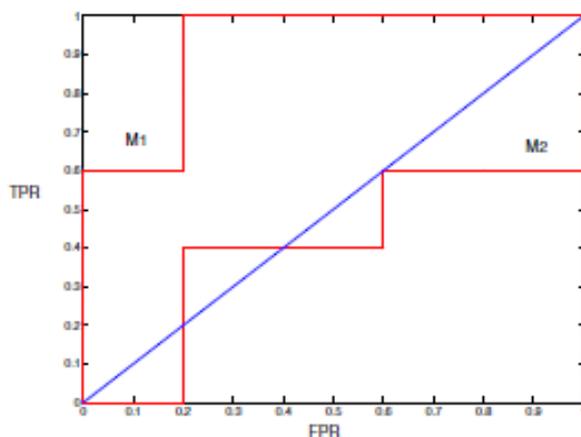
Aspeto (céu)	temperatura	chuva	vento
Pouco nublado	media	ausente	forte

Questão 3 [1.5 valores]

Pretende-se avaliar a performance de dois modelos de classificação M1 e M2. Para isso, escolheu-se um conjunto de teste com dez entidades. Cada entidade é descrita por 26 atributos codificados de 'A' a 'Z' e uma classe binária com valores '+'/'-'. A tabela seguinte mostra as probabilidades *a posteriori* resultado da aplicação de cada um dos classificadores ao conjunto de teste (note que só se apresentam os resultados para a classe '+').

Entidade	Classe original	M1 $P(+ A, B, \dots Z)$	M2 $P(+ A, B, \dots Z)$
1	+	0.73	0.61
2	+	0.69	0.03
3	-	0.44	0.68
4	-	0.55	0.31
5	+	0.67	0.45
6	+	0.47	0.09
7	-	0.08	0.38
8	-	0.15	0.05
9	+	0.45	0.01
10	-	0.35	0.04

A curva ROC obtida na análise de performance de cada um dos classificadores está indicada na figura abaixo.



Considere que se definiu o limiar de cutoff $\tau = 0.5$ para classificar uma entidade como sendo '+' (i.e. se a correspondente probabilidade *a posteriori* for superior a τ).

- 3.1 Calcule as medidas de 'precision', 'recall' e 'F-measure' para o classificador M1.
- 3.2 Calcule as medidas anteriores para o classificador M2 assumindo o mesmo valor de τ .
- 3.3 Qual é o melhor modelo de classificação em termos da medida 'F-measure'? Estes resultados são consistentes com os que se obtêm por análise da curva ROC? Justifique, referindo-se aos valores de TPR e FPR.

Questão 4 [2.5 valores]

Considere o seguinte conjunto de pontos reais, X classificados por uma classe binária com valores '+' e '-', como representado na figura.

X	Y
-0.1	-
0.7	+
1.0	+
1.6	-
2.0	+
2.5	+
3.2	-
3.5	-
4.1	+
4.9	+



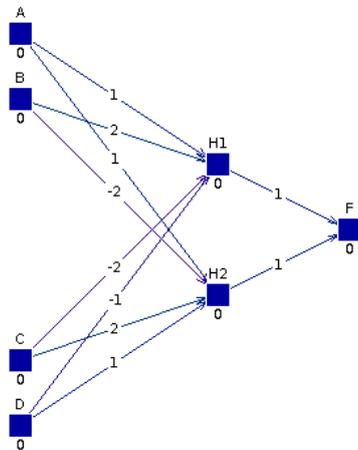
Com este conjunto de dados treinou-se um classificador K -NN com distância Euclideana não pesada.

4.1. Calcule o erro de validação cruzada *leave-one-out* para o K -NN com $K=1$. Apresente o resultado como o número de casos mal classificados. Justifique a sua resposta, apresentando para cada ponto o valor da classe predicta, numa tabela como a original.

4.2 Qual é o erro de validação cruzada *leave-one-out* para o K -NN com $K=3$? Justifique a sua resposta como na alínea anterior.

Questão 5 [1+2+1+1+1 valores]

Considere o seguinte MLP:



Considere ainda: os valores sob as unidades indicam os valores de *bias*, as unidades H₁, H₂ e

F têm função de ativação $\sigma(x) = \frac{1}{1 + e^{-x}}$ e o seguinte conjunto de dados (sendo F a saída do MLP):

A	B	C	D	F
0.0	5.0	2.5	0.0	5.0
0.0	0.0	2.5	0.0	0.0
0.0	0.0	0.0	0.0	5.0
0.0	5.0	0.0	0.0	0.0

5.1 Calcule a saída em F, quando é apresentada a entrada (A,B,C,D)=(1,1,1,1).

5.2 Para o mesmo padrão de entrada (A,B,C,D)=(1,1,1,1), Qual o valor de atualização do valor de bias de H1 por aplicação do algoritmo standard error backpropagation, quando se pretende que F(1,1,1,1)=5,82 (Indique e assumo os valores habituais para os parâmetros deste algoritmo, se necessário, efetuando o arredondamento final às décimas).

5.3. Imagine que as unidades H1 e H2 representavam os únicos neurónios num SOM, utilizando o conjunto de dados indicado no ponto 3 e após normalização min-max, qual é a BMU do padrão (A,B,C,D)=(2.5, 2.5, 0, 0)?

NOTA: no seu interesse, pf. tente apresentar uma respostas sucintas às questões que se seguem: i.e., evite usar mais de meia página A4 nas respostas às questões 4 e 5!

5.4. Com estes dados considera que uma rede onde se tivesse removido o neurónio H1 conseguiria aprender melhor que a rede original?

5.5. Considera possível à rede apresentada, após aprendizagem, atingir um SSE de 0.0 no conjunto de dados apresentado?

Questão 6 [2.0+ 1.0 + 0.5 valores]

Considere a aplicação do algoritmo Fuzzy C-Means (FCM) ao conjunto de 6 pontos do espaço bidimensional: $\mathbf{a}_1=(2,2)$, $\mathbf{a}_2=(4,10)$, $\mathbf{a}_3=(9,5)$, $\mathbf{a}_4=(14,8)$, $\mathbf{a}_5=(18,10)$, $\mathbf{a}_6=(21,7)$. O algoritmo é aplicado com parâmetros $C=2$, $m=2.0$ e $\varepsilon=0.01$, tendo-se escolhido como protótipos iniciais os pontos \mathbf{a}_5 e \mathbf{a}_6 .

6.1 Calcule a matriz de pertinência difusa $U^{(1)}$ e os correspondentes protótipos $\mathbf{v}_1^{(1)}$ e $\mathbf{v}_2^{(1)}$ após uma iteração do algoritmo. Apresente todos os cálculos necessários. Apresente os valores finais com 2 casas decimais.

6.2. Enuncie as restrições que definem uma partição difusa. Explique e justifique o seu significado.

6.3. Apresente um exemplo de uma matriz de pertinência difusa (pode recorrer à obtida em 6.1) e a correspondente matriz de pertinência *hard* obtida por defusificação. Qual das restrições enunciadas em 6.2 se altera? Defina essa restrição.

Questão 7 [1.5 valores]

Comente a seguinte frase:

no contexto do classificador SVM, se as classes não forem linearmente separáveis, considerando os atributos usados, não é possível garantir a existência de hiperplanos entre cada par de classes, o que impede à partida uma classificação sem erros.

Questão 8 [2 valores]

No contexto da publicação *Mining Quantitative Association Rules in Large Relational Tables* dada nas aulas teóricas, considere a seguinte tabela contendo dados sobre pessoas:

Registo nº	Sexo	vencimento	Niv. Escolaridade
1	F	650	Básico
2	M	700	Básico
3	M	1200	Secundário
4	F	1150	Secundário
5	M	1250	Secundário
6	F	1180	Secundário
7	M	1700	Superior
8	M	1800	Superior
9	F	1850	Superior
10	M	5000	Superior

Considere ainda os parâmetros *Minimum Support* e *Minimum Confidence* com valores de 20% e 60% respectivamente e extraia as regras de associação (*Association Rules*) possíveis, tendo em conta apenas estes dois parâmetros e os dados da tabela. Para tal, assuma as partições que lhe pareçam adequadas relativamente aos atributos que entender.