Reconhecimento de Padrões

Engenharia Informática de Gestão– 2001-2002

2º exame - 4 de Fevereiro de 2002

I. Considere o problema de classificação em duas classes dos padrões

$$x1=[0\ 0.5]^T \text{ e } x2=[-.5\ .5]^T$$

dados os conjuntos de treino:

Classe A	Classe B
-0.43 -0.19	-0.43 0.40
-1.67 0.7	-0.83 0.67
0.12 -0.59	-0.32 0.70
0.29 2.18	-0.09 0.68
-1.15 -0.14	-0.67 0.82
1.19 0.11	-0.28 0.67
1.19 1.07	-0.19 0.80
-0.04 0.06	-0.89 0.20
0.33 -0.10	-0.86 0.50
0.17 -0.83	-0.36 0.46

1. Assuma que a classe A tem como distribuição $p(x|\omega_A)=\frac{1}{2\pi}e^{\frac{-(x-\mu_A)^T(x-\mu_A)}{2}}$ e as amostras da classe B são modeladas por $p(x|\omega_B)=\frac{2}{\pi}e^{-2(x-\mu_B)^T(x-\mu_B)}$, em que os parâmetros μ_A e μ_B são desconhecidos. Seja ainda $P(\omega_A)=P(\omega_B)=0.5$.

(a) (2 valores) Determine as estimativas de máxima verosimilhança para os parâmetros μ_A e μ_B usando os conjuntos de treino acima gerados de forma independente. Justifique todos os seus cálculos.

(b) Seja o classificador de Bayes, com custos associados às decisões dados por: $c_{11}=c_{22}=0$, $c_{12}=3$, $c_{21}=1$. (Se não fez a alínea anterior, considere $\mu_A=[0\ \ 0]^T$ e $\mu_B=[-0.5\ \ 0.5]^T$).

(i) (2 valores) Determine a regra de decisão de Bayes, identificando funções discriminantes a associar a cada classe.

(ii) (1 valor) Determine as regiões de decisão do classificador.

(iii) (2 valores) Classifique as amostras x_1 e x_2 de acordo com o classificador de Bayes. Indique a probabilidade de erro de classificação.

2. Considere agora desconhecidas as distribuições das classes A e B.

(a) (1 valor) Classifique as amostras x_1, x_2, x_3 pelo método do vizinho mais próximo.

(b) (1 valor) Determine a partição do espaço de características proporcionado pelo classificador 1-NN (o vizinho mais próximo).

II. **(2 valores)** Prove que o estimador bayesiano de erro quadrático médio é um estimador de variância mínima na situação em que é um estimador não polarizado (considere o caso escalar).

III. Considere as seguintes sequências de símbolos:

S1=abbcbbba,	S2=acba,	S3=abcbba
S4=bacb,	S5=bcaccb,	S6=bccacccb

Usando como medida de distância entre strings o menor número de operações de edição de string (eliminação, inserção e substituição), obtém-se a seguinte matriz de distâncias entre as strings:

	1	2	3	4	5	6
1	0.0	.50	.25	.66	.72	.77
2	.50	0.0	.33	.40	.57	.67
3	.25	.33	0.0	.57	.67	.72
4	.67	.4	.57	0.0	.33	.50
5	.72	.57	.67	.33	0.0	.25
6	.77	.66	.72	.50	.25	0.0

1. **(2 valores)** Usando o algoritmo de hierárquico aglomerativo de ligação simples (*clustering Single link*) baseado na matriz de distâncias acima, proponha uma partição para as sequências acima. Desenhe o dendrograma correspondente. Justifique todos os passos usados para a resolução do problema.

2. Considere agora apenas as sequências S1, S2 e S3.

(a) **(2 valores)** Assumindo que estas strings são amostras da linguagem $L = ab^n cb^{n+1}a, n \geq 0$, indique uma gramática independente do contexto, estocástica, que modele esta linguagem.

(b) **(0.5 valores)** Poderia encontrar uma gramática regular para descrever esta linguagem? Justifique.

(c) **(1.5 valores)** Determine a árvore de derivação para a string S1 de acordo com a gramática que desenhou em (a). (Caso não tenha resolvido a alínea (a), use a seguinte gramática:

$$G = (V_N, V_T, R, \sigma) \qquad V_N = \{\sigma, B\} \qquad V_T = \{a, b, c\}$$

$$R: \quad \sigma \rightarrow aBcBa \qquad B \rightarrow b \qquad B \rightarrow bB$$

(d) **(3 valores)** Infira agora uma gramática para modelar S1, S2 e S3 usando o método das derivativas formais. Comente a adequação da gramática e proponha possíveis refinamentos e/ou soluções alternativas.