Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2016

Apellidos: ____________ **Nombre:** ____________

Cuestiones (3 puntos, 30 minutos, sin apuntes)

P(A) = P(B) = P(C) = 1/3

$\boxed{\text{C}}$ Sean A, B y C tres clases con la misma probabilidad a priori y f.d. condicional de clase de tipo Bernoulli $\mathbf{p}_A = \begin{pmatrix} 0.1 \\ 0.9 \end{pmatrix}$, $\mathbf{p}_B = \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix}$, $\mathbf{p}_C = \begin{pmatrix} 0.8 \\ 0.5 \end{pmatrix}$. ¿En qué clase sería clasificada una muestra $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ dados los parámetros anteriores?

A) Clase A $c_A = 0.09.$

B) Clase B $c_B = 0.24$

C) Clase C $c_C = 0.4.$

D) En cualquiera de las tres

$$c^*(\mathbf{x}) = \underset{c=1,\dots,C}{\text{argmax}} \ \log P(c) + \log p(\mathbf{x} \mid c)$$

$$= \underset{c=1,\dots,C}{\text{argmax}} \ \log P(c) + \log \prod_{d=1}^{D} p_{cd}^{x_d} (1 - p_{cd})^{(1-x_d)}$$

$$= \underset{c=1,\dots,C}{\text{argmax}} \ \log P(c) + \sum_{d=1}^{D} x_d \log p_{cd} + (1 - x_d) \log(1 - p_{cd})$$

$\boxed{\text{B}}$ Dado el siguiente conjunto de vectores binarios bidimensionales:

n	1	2	3	4	5	6	7	8	9	10	11	12
x_{n1}	1	1	0	1	1	1	0	1	1	0	1	1
x_{n2}	1	0	1	0	1	0	1	1	1	1	1	1
c_n	1	1	1	1	1	1	1	1	2	2	2	2

¿Cuál es la estimación de los parámetros del clasificador Bernoulli más probable?

A) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{6}{11}, \frac{5}{11}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{3}{7}, \frac{4}{7}\right)^t$

B) $\hat{p}(1) = \frac{2}{3}$, $\hat{p}(2) = \frac{1}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{6}{8}, \frac{5}{8}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{3}{4}, \frac{4}{4}\right)^t$

C) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{6}{8}, \frac{5}{8}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{3}{4}, \frac{4}{4}\right)^t$

D) $\hat{p}(1) = \frac{2}{3}$, $\hat{p}(2) = \frac{1}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{6}{11}, \frac{5}{11}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{3}{7}, \frac{4}{7}\right)^t$

$\boxed{\text{B}}$ ¿Cuál de los siguientes valores del parámetro $\mathbf{p}$ no define una distribución multinomial?

A) $\mathbf{p} = \left(\frac{3}{4}, \frac{1}{4}\right)^t$

B) $\mathbf{p} = \left(\frac{3}{4}, \frac{3}{4}\right)^t$

C) $\mathbf{p} = \left(\frac{1}{2}, \frac{1}{2}\right)^t$

D) Todos los valores anteriores del parámetro $\mathbf{p}$ definen una distribución multinomial.

$\boxed{\text{C}}$ Sean A y B dos clases con igual prior y f.d.p. condicionales de clase gaussianas con los siguientes parámetros $\mu_A = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma_A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, $\mu_B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ y $\Sigma_B = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$, ¿qué tipo de frontera de decisión definen?

A) Lineal

B) Lineal definida a trozos

C) Cuadrática

D) Ninguna de las anteriores

$\boxed{\text{A}}$ La frontera de decisión entre dos clases que se obtiene con el vecino más cercano cuando las clases tienen un único prototipo es:

A) Lineal
B) Lineal a trozos
C) Cuadrática
D) Ninguna de las anteriores

$\boxed{\text{D}}$ Sea la distancia Euclídea ponderada L_w con pesos postivos y no nulos:

A) $L_w(\mathbf{x}, \mathbf{y}) \leq L_2(\mathbf{x}, \mathbf{y})$
B) $L_1(\mathbf{x}, \mathbf{y}) \leq L_w(\mathbf{x}, \mathbf{y})$
C) $L_w(\mathbf{x}, \mathbf{y}) \geq L_0(\mathbf{x}, \mathbf{y})$
D) Ninguna de las anteriores

$\boxed{\text{D}}$ Sea P el error del clasificador de Bayes. El error del vecino más cercano $\hat{P}$ tiene la siguiente propiedad asintótica (cuando $n \to \infty$):

A) $\hat{P} = P$
B) $\hat{P} = 2P$
C) $\hat{P} \leq P$
D) $P \leq \hat{P} \leq 2P$

$\boxed{\text{B}}$ En general, ¿cuál de las siguientes afirmaciones es falsa?

A) El Variance aumenta al escoger clasificadores más fuertes
B) El Variance aumenta al escoger clasificadores más débiles
C) El Bias es menor en clasificadores más fuertes
D) El Bias se reduce empleando Boosting

$\boxed{\text{A}}$ Esencialmente en Boosting:

A) Se combinan diferentes clasificadores sobre el mismo conjunto de aprendizaje
B) Se combinan el mismo clasificador sobre diferentes conjuntos de aprendizaje
C) Se combinan diferentes clasificadores sobre diferentes conjuntos de aprendizaje
D) Ninguna de las anteriores

$\boxed{\text{C}}$ En la teoría interactiva de la decisión, el criterio de decisión de una hipótesis h dada la señal x, la historia h' y la realimentación f es:

A) $\arg\max_h p(h, h', f|x)$
B) $\arg\max_h p(h, h', x|f)$
C) $\arg\max_h p(h|x, h', f)$
D) $\arg\max_h p(f|h, h', x)$

Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2016

Apellidos: **Nombre:**

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(2 puntos)** Tenemos $N = 16$ vectores de cuentas 5-dimensionales aleatoriamente extraídos de $C = 2$ distribuciones multinomiales independientes:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_{n1}	1	2	1	1	2	1	1	0	0	0	0	0	0	0	0	0
x_{n2}	2	1	0	1	2	0	0	3	0	0	0	0	0	0	0	0
x_{n3}	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	1
x_{n4}	1	0	2	1	0	2	3	3	1	2	1	1	2	1	1	3
x_{n5}	0	0	0	0	0	0	0	0	1	3	1	1	1	1	3	1
c_n	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2

   (a) Calcula los parámetros del clasificador multinomial más probable respecto a estos datos. (0.5 puntos)

   (b) Suaviza los parámetros multinomial mediante Laplace con $\epsilon = 0.2$. (0.5 puntos)

   (c) Suaviza los parámetros con descuento absoluto de $\epsilon = 0.05$ e interpolación usando como distribución generalizada, la distribución uniforme. (0.5 puntos)

   (d) Clasifica la muestra de test $y = (1\ 1\ 1\ 1\ 1)^t$ con el clasificador multinomial resultante de aplicar los parámetros suavizados del apartado c). (0.5 puntos)

Solución:

a)

$$p(1) = p(2) = \frac{8}{16} = 0.5$$

$$\hat{\mathbf{p}}_1 = \frac{1}{20}\begin{pmatrix} 1+2+1+1+2+1+1+0 \\ 2+1+0+1+2+0+0+3 \\ 0+0+0+0+0+0+0+0 \\ 1+0+2+1+0+2+3+3 \\ 0+0+0+0+0+0+0+0 \end{pmatrix} = \frac{1}{30}\begin{pmatrix} 9 \\ 9 \\ 0 \\ 12 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.3 \\ 0.0 \\ 0.4 \\ 0.0 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \frac{1}{20}\begin{pmatrix} 0+0+0+0+0+0+0+0 \\ 0+0+0+0+0+0+0+0 \\ 1+1+0+1+0+1+1+1 \\ 1+2+1+1+2+1+1+3 \\ 1+3+1+1+1+1+3+1 \end{pmatrix} = \frac{1}{30}\begin{pmatrix} 0 \\ 0 \\ 6 \\ 12 \\ 12 \end{pmatrix} = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.2 \\ 0.4 \\ 0.4 \end{pmatrix}$$

b)

$$\hat{\mathbf{p}}_1 = \begin{pmatrix} 0.3 \\ 0.3 \\ 0.0 \\ 0.4 \\ 0.0 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_1 = \frac{1}{2.0}\begin{pmatrix} 0.3+0.2 \\ 0.3+0.2 \\ 0.0+0.2 \\ 0.4+0.2 \\ 0.0+0.2 \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0.25 \\ 0.10 \\ 0.30 \\ 0.10 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.2 \\ 0.4 \\ 0.4 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_2 = \frac{1}{2.0}\begin{pmatrix} 0.0+0.2 \\ 0.0+0.2 \\ 0.2+0.2 \\ 0.4+0.2 \\ 0.4+0.2 \end{pmatrix} = \begin{pmatrix} 0.10 \\ 0.10 \\ 0.20 \\ 0.30 \\ 0.30 \end{pmatrix}$$

c)

$$\hat{\mathbf{p}}_1 = \begin{pmatrix} 0.4 \\ 0.3 \\ 0.0 \\ 0.3 \\ 0.0 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_1 = \begin{pmatrix} 0.4 - 0.05 + \frac{1}{5} \cdot 0.15 \\ 0.3 - 0.05 + \frac{1}{5} \cdot 0.15 \\ 0.0 \qquad + \frac{1}{5} \cdot 0.15 \\ 0.3 - 0.05 + \frac{1}{5} \cdot 0.15 \\ 0.0 \qquad + \frac{1}{5} \cdot 0.15 \end{pmatrix} = \begin{pmatrix} 0.38 \\ 0.28 \\ 0.03 \\ 0.28 \\ 0.03 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.2 \\ 0.4 \\ 0.4 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_2 = \begin{pmatrix} 0.0 \qquad + \frac{1}{5} \cdot 0.15 \\ 0.0 \qquad + \frac{1}{5} \cdot 0.15 \\ 0.2 - 0.05 + \frac{1}{5} \cdot 0.15 \\ 0.4 - 0.05 + \frac{1}{5} \cdot 0.15 \\ 0.4 - 0.05 + \frac{1}{5} \cdot 0.15 \end{pmatrix} = \begin{pmatrix} 0.03 \\ 0.03 \\ 0.18 \\ 0.38 \\ 0.38 \end{pmatrix}$$

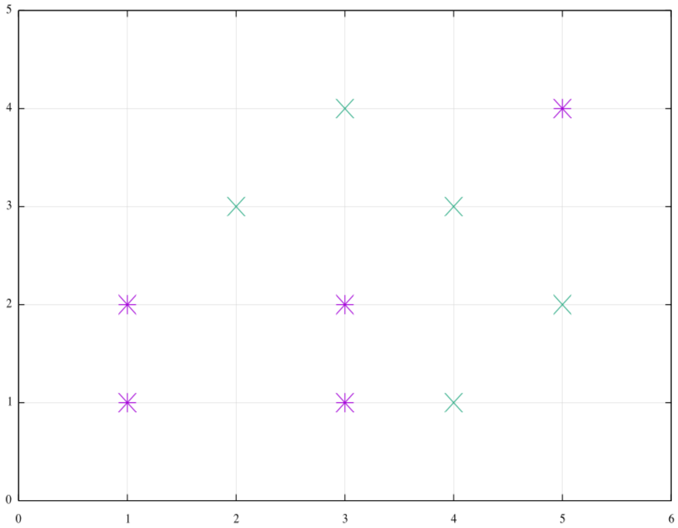
d)

En nuestro caso, dado que las priors son idénticas, la regla de clasificación se reduce a: $\hat{c}(y) = \arg\max_c p(y \mid c)$

$$p(y = (1\ 1\ 1\ 1\ 1) \mid c = 1) = 0.38 \cdot 0.28 \cdot 0.03 \cdot 0.28 \cdot 0.03 = 2.7 \cdot 10^{-5}$$
$$p(y = (1\ 1\ 1\ 1\ 1) \mid c = 2) = 0.03 \cdot 0.03 \cdot 0.18 \cdot 0.38 \cdot 0.38 = 2.3 \cdot 10^{-5}$$

La muestra y se clasifica en la clase 1.

2. **(2 puntos)** Dado el conjunto de aprendizaje $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8, \mathbf{x}_9, \mathbf{x}_{10}\}$ con la distribución de clases que muestra la figura. Realiza una ejecución del algoritmo de condensado de Hart con distancia euclídea y $k = 1$.



Clase A

$\mathbf{x}_1 = (1, 1)$
$\mathbf{x}_3 = (1, 2)$
$\mathbf{x}_5 = (3, 2)$
$\mathbf{x}_9 = (5, 4)$
$\mathbf{x}_{10} = (3, 1)$

Clase B

$\mathbf{x}_2 = (4, 1)$
$\mathbf{x}_4 = (5, 2)$
$\mathbf{x}_6 = (2, 3)$
$\mathbf{x}_7 = (4, 3)$
$\mathbf{x}_8 = (3, 4)$

El orden de recorrido de los prototipos en el algoritmo de Hart es **decreciente** con el índice de los mismo: $\mathbf{x}_{10} \ldots \mathbf{x}_1$.

En caso de **empate** de distancias de prototipos de clases diferentes se clasifica en la clase incorrecta.

Solución:

Primera parte, construcción del conjunto S y G:

$$\begin{aligned}
&\mathbf{x}_{10} \rightarrow S \\
&\mathbf{x}_9, \text{Acierto} \rightarrow G \\
&\mathbf{x}_8, \text{Error} \rightarrow S \\
&\mathbf{x}_7, \text{Acierto} \rightarrow G \\
&\mathbf{x}_6, \text{Acierto} \rightarrow G \\
&\mathbf{x}_5, \text{Acierto} \rightarrow G \\
&\mathbf{x}_4, \text{Error} \rightarrow S \\
&\mathbf{x}_3, \text{Acierto} \rightarrow G \\
&\mathbf{x}_2, \text{Error} \rightarrow S \\
&\mathbf{x}_1, \text{Acierto} \rightarrow G
\end{aligned}$$

$S = \{\mathbf{x}_{10}, \mathbf{x}_8, \mathbf{x}_4, \mathbf{x}_2\}$ y $G = \{\mathbf{x}_9, \mathbf{x}_7, \mathbf{x}_6, \mathbf{x}_5 \mathbf{x}_3, \mathbf{x}_1\}$

Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_9, \text{Error} \rightarrow S \\
&\mathbf{x}_7, \text{Error} \rightarrow S \\
&\mathbf{x}_6, \text{Acierto, no mover a } S \\
&\mathbf{x}_5, \text{Acierto, no mover a } S \\
&\mathbf{x}_3, \text{Acierto, no mover a } S \\
&\mathbf{x}_1, \text{Acierto, no mover a } S \\
&error = 1 \rightarrow \text{volver a recorrer } G
\end{aligned}$$

Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_6, \text{Acierto, no mover a } S \\
&\mathbf{x}_5, \text{Acierto, no mover a } S \\
&\mathbf{x}_3, \text{Acierto, no mover a } S \\
&\mathbf{x}_1, \text{Acierto, no mover a } S \\
&error = 0 \rightarrow \text{Acabar}
\end{aligned}$$

Acaba con el conjunto reducido: $S = \{\mathbf{x}_{10}, \mathbf{x}_9, \mathbf{x}_8, \mathbf{x}_7, \mathbf{x}_4, \mathbf{x}_2\}$