Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2015

Apellidos: ☐ **Nombre:** ☐

Profesor: ☐ **Jorge Civera** ☐ **Roberto Paredes**

Cuestiones (3 puntos, 30 minutos, sin apuntes)

$\boxed{\text{B}}$ ¿Qué tipo de fronteras de decisión define un clasificador basado en distribuciones de Bernoulli?

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

$\boxed{\text{D}}$ Dado el siguiente conjunto de vectores de cuentas bidimensionales:

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

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

A) $\hat{p}(1) = \frac{1}{2}$, $\hat{p}(2) = \frac{1}{2}$, $\hat{\mathbf{p}}_1 = \left(\frac{4}{7}, \frac{3}{7}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{5}{11}, \frac{6}{11}\right)^t$
B) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{8}{23}, \frac{1}{4}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{15}{23}, \frac{3}{4}\right)^t$
C) $\hat{p}(1) = \frac{1}{2}$, $\hat{p}(2) = \frac{1}{2}$, $\hat{\mathbf{p}}_1 = \left(\frac{11}{23}, \frac{11}{24}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{12}{23}, \frac{13}{24}\right)^t$
D) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{4}{7}, \frac{3}{7}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{5}{11}, \frac{6}{11}\right)^t$

$\boxed{\text{C}}$ Dado los siguientes valores para el parámetro multinomial $\hat{p} = (0.4\ 0.0\ 0.6\ 0.0)^t$, aplicamos un suavizado sobre dicho parámetro con $\epsilon = 0.1$, obteniendo $\tilde{p} = (0.3\ 0.1\ 0.5\ 0.1)^t$, ¿qué tipo de suavizado hemos aplicado?

A) Laplace
B) Descuento absoluto e interpolación
C) Descuento absoluto y backing-off
D) Truncamiento simple

$\boxed{\text{C}}$ Sean A y B dos clases con igual prior y f.d.p. condicionales de clase gaussianas con los siguientes parámetros $\mu_A = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma_A = \begin{pmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}$, $\mu_B = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$ y $\Sigma_B = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$, ¿qué tipo de frontera de decisión definen?

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

$\boxed{B}$ Dadas las matrices de covarianza del apartado anterior, ¿cuales serían las matrices de covarianza resultantes de aplicar *flat smoothing* con $\alpha = \frac{1}{2}$?

A) $\tilde{\Sigma}_A = \begin{pmatrix} \frac{1}{16} & 0 \\ 0 & \frac{1}{8} \end{pmatrix}$ y $\tilde{\Sigma}_B = \begin{pmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}$
B) $\tilde{\Sigma}_A = \begin{pmatrix} \frac{9}{16} & 0 \\ 0 & \frac{5}{8} \end{pmatrix}$ y $\tilde{\Sigma}_B = \begin{pmatrix} \frac{5}{8} & 0 \\ 0 & \frac{3}{4} \end{pmatrix}$
C) $\tilde{\Sigma}_A = \begin{pmatrix} \frac{5}{16} & 0 \\ 0 & \frac{3}{8} \end{pmatrix}$ y $\tilde{\Sigma}_B = \begin{pmatrix} \frac{3}{8} & 0 \\ 0 & \frac{2}{4} \end{pmatrix}$
D) Ninguna de las anteriores

$\boxed{D}$ Respecto a las fronteras de decisión del vecino más cercano (1-nn) en un problema de dos clases:

A) Es lineal solo si tienes una muestra por clase
B) Es lineal siempre
C) Es lineal si tienes más de una muestra por clase
D) Ninguna de las anteriores

$\boxed{A}$ Sea $\mathbf{x}$ la representación de los objetos y c las clases. El clasificador por los k-vecinos obtiene una estimación directa de:

A) $p(c \mid \mathbf{x})$
B) $p(\mathbf{x} \mid c)$
C) $p(\mathbf{x})$
D) Ninguna de las anteriores

$\boxed{A}$ El error de un clasificador tiene tres componentes: Bias, Variance y Noise. Bagging pretende:

A) Reducir el variance
B) Reducir el bias
C) Reducir el Noise
D) Ninguna de las anteriores

$\boxed{B}$ En Adaboost, sea C_m el mejor clasificador débil en la iteración m con un error ϵ_m y un $\alpha_m = \frac{1}{2} ln(\frac{1-\epsilon_m}{\epsilon_m})$. Después de M iteraciones obtenemos un clasificador como el siguiente:

A) $G(\mathbf{x}) = \sum_{m=1}^{M} C_m(\mathbf{x})$
B) $G(\mathbf{x}) = \sum_{m=1}^{M} \alpha_m C_m(\mathbf{x})$
C) $G(\mathbf{x}) = \sum_{m=1}^{M} \epsilon_m C_m(\mathbf{x})$
D) Ninguna de las anteriores

$\boxed{D}$ En un contexto de interacción hombre-máquina, la realimentación determinista se caracteriza:

A) Porque el usuario siempre emplea la misma acción
B) Porque el usuario determina la modalidad de entrada entre todas las disponibles
C) Porque la decodificación es multimodal
D) Porque la acción del usuario puede interpretarse de forma no ambigua

Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2015

Apellidos: ☐ **Nombre:** ☐

Profesor: ☐ **Jorge Civera** ☐ **Roberto Paredes**

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(2 puntos)** Tenemos $N = 9$ vectores binarios bidimensionales aleatoriamente extraídos de $C = 2$ distros Bernoulli independientes:

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

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

b) Calcula el error global. (0.5 puntos)

c) Suaviza los parámetros Bernoulli de ambas clases aplicando truncamiento simple con $\epsilon = \frac{1}{4}$. (0.5 puntos)

d) Clasifica el vector $y = (1\ 1)^t$ con el clasificador Bernoulli **suavizado** del apartado anterior. (0.5 puntos)

Solución:

a) La estimación de los parámetros del clasificador Bernoulli es

$$p(1) = \frac{3}{9} = \frac{1}{3} \quad p(2) = \frac{6}{9} = \frac{2}{3}$$

$$\hat{\mathbf{p}}_1 = \frac{1}{3}\begin{pmatrix} 1+0+0 \\ 0+0+1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \frac{1}{6}\begin{pmatrix} 1+1+1+1+1+0 \\ 0+0+0+0+0+1 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{5}{6} \\ \frac{1}{6} \end{pmatrix}$$

b) En general, el error global $p(e)$ se calcula como

$$p(e) = \sum_x p(x,e) = \sum_x p(x)\, p(e \mid x) = \sum_x p(x)\, \left(1 - \max_c p(c \mid x)\right).$$

Aplicamos Bayes y simplificamos la expresión teniendo en cuenta que sólo tenemos dos clases

$$\begin{aligned}
p(e) &= \sum_x p(x)\ \min_c p(c \mid x) \\
&= \sum_x p(x)\ \min_c \frac{p(c)\, p(x \mid c)}{p(x)} \\
&= \sum_x \min_c p(c)\, p(x \mid c) \\
&= \sum_x \min_c p(c) \prod_d\ p_{cd}^{x_d}\, (1-p_{cd})^{(1-x_d)} \\
&= \sum_x \min\left(\frac{1}{3}\cdot\frac{1}{3}^{x_1}\left(1-\frac{1}{3}\right)^{(1-x_1)}\cdot\frac{1}{3}^{x_2}\left(1-\frac{1}{3}\right)^{(1-x_2)}, \frac{2}{3}\cdot\frac{5}{6}^{x_1}\left(1-\frac{5}{6}\right)^{(1-x_1)}\cdot\frac{1}{6}^{x_2}\left(1-\frac{1}{6}\right)^{(1-x_2)}\right)
\end{aligned}$$

$\mathbf{x}$		$p(c)\, p(x \mid c)$		
x_1	x_2	$c=1$	$c=2$	$\min_c$
0	0	$\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}$	$\frac{2}{3}\cdot\frac{1}{6}\cdot\frac{5}{6}$	$\frac{10}{108}$
0	1	$\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}$	$\frac{2}{3}\cdot\frac{1}{6}\cdot\frac{1}{6}$	$\frac{2}{108}$
1	0	$\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}$	$\frac{2}{3}\cdot\frac{5}{6}\cdot\frac{5}{6}$	$\frac{2}{27}$
1	1	$\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}$	$\frac{2}{3}\cdot\frac{5}{6}\cdot\frac{1}{6}$	$\frac{1}{27}$

Sumando el mínimo de $p(c)\,p(x \mid c)$ para cada x:

$$p(e) = \frac{10}{108} + \frac{2}{108} + \frac{2}{27} + \frac{1}{27} = \frac{2}{9}$$

c)

$$\tilde{\mathbf{p}}_1 = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$$
$$\tilde{\mathbf{p}}_2 = \begin{pmatrix} \frac{3}{4} \\ \frac{1}{4} \end{pmatrix}$$

d) Aplicamos la regla de clasificación de Bayes:

$$\hat{c}(y) = \arg\max_c p(c)\,p(y \mid c) = p(c) \prod_d p_{cd}^{x_d}\,(1 - p_{cd})^{(1-x_d)}$$

$$p(c=1)\,p(y = (1\;1)^t \mid c=1) = \frac{1}{3} \cdot \frac{1}{3}^{x_1} \left(1 - \frac{1}{3}\right)^{(1-x_1)} \cdot \frac{1}{3}^{x_2} \left(1 - \frac{1}{3}\right)^{(1-x_2)} = \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{27}$$
$$p(c=2)\,p(y = (1\;1)^t \mid c=2) = \frac{2}{3} \cdot \frac{3}{4}^{x_1} \left(1 - \frac{3}{4}\right)^{(1-x_1)} \cdot \frac{1}{4}^{x_2} \left(1 - \frac{1}{4}\right)^{(1-x_2)} = \frac{2}{3} \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{1}{8}$$

La muestra y se clasifica en la clase 2.

2. **(2 puntos)** Dado un problema de clasificación en dos clases $\{+1, -1\}$ donde los objetos a clasificar $\mathbf{x} \in \Re^2$. Se decide emplear Adaboost con los siguientes clasificadores débiles:

$$g_i(\mathbf{z}) = \begin{cases} s & \textit{si } \mathbf{z}_1 > t \\ -s & \textit{en otro caso} \end{cases} \qquad \#\mathbf{z}_1 \textit{ es la primera componente de } \mathbf{z}$$

$$0 \leq i < 8$$

donde el punto de corte $t = \lfloor i/2 \rfloor + 0.5$, y

$$s = \begin{cases} 1 & \textit{si } i \textit{ es par} \\ -1 & \textit{en otro caso} \end{cases}$$

A partir de unos datos de aprendizaje se ejecuta Adaboost con $M = 3$ y se obtienen los siguientes clasificadores débiles con su error asociado en cada iteración:

Iteración	Mejor clasificador	Error
1	g_0	1/4
2	g_4	1/6
3	g_3	1/5

Se pide:

a) Calcular los α asociados a cada clasificador débil (0.5 puntos)

b) Clasificar la muestra $\mathbf{x} = (2, 3)$ con el clasificador resultante (1 punto)

c) Responde brevemente cómo son las fronteras de decisión del clasificador resultante (0.5 puntos)

Solución:

a) $\alpha_1 = 0.549306$, $\alpha_2 = 0.804719$, $\alpha_3 = 0.693147$

b) Clasificador resultante $G(\mathbf{x}) = \sum_{m=1}^{M} \alpha_m C_m(\mathbf{x})$
$G(\mathbf{x}) = 0.549306 g_0(\mathbf{x}) \ + \ 0.804719 g_4(\mathbf{x}) \ + \ 0.693147 g_3(\mathbf{x})$
$G(\ (2,3)\ ) = 0.549306(+1) \ + \ 0.804719(-1) \ + \ 0.693147(-1) = -0,948 < 0 \rightarrow$ clase -1

c) Son fronteras de decisión lineales paralelas a la segunda dimensión (eje y en dos dimensiones)