Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2014

Apellidos: ________ **Nombre:** ________

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

Cuestiones (3 puntos, 30 minutos, sin apuntes)

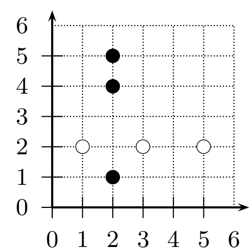
$\boxed{\text{A}}$ ¿Para qué valores de $\mathbf{w}$ la distancia Euclídea ponderada $d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{d} w_i \cdot ((x_i - y_i)^2)^{1/2}$ define una métrica?

A) $w_i > 0 \quad i = 1,\ldots,d$
B) $w_i \geq 0 \quad i = 1,\ldots,d$
C) $w_i \leq 0 \quad i = 1,\ldots,d$
D) Ninguna de las anteriores

$\boxed{\text{D}}$ ¿Cuál de las siguientes instanciaciones de la regla de Bayes para el clasificador de K-vecinos no es equivalente a $\hat{c} = \arg\max_c k_c$?

A) $\hat{c} = \arg\max_c \log k_c + \log K$
B) $\hat{c} = \arg\max_c \log k_c - \log K$
C) $\hat{c} = \arg\max_c \frac{k_c}{K}$
D) Todas las anteriores son equivalentes.

$\boxed{\text{C}}$ La figura de la derecha muestra prototipos bidimensionales de 2 clases, $X = \{x_1 = (1,2,\circ),\ x_2 = (2,1,\bullet),\ x_3 = (2,4,\bullet),\ x_4 = (2,5,\bullet),\ x_5 = (3,2,\circ),\ x_6 = (5,2,\circ)\}$. ¿Cuál es el conjunto de prototipos X' resultante de aplicar el algoritmo de edición de Wilson con el vecino más cercano en distancia Euclídea?



A) $X' = \{x_1, x_2, x_3, x_4, x_5, x_6\}$
B) $X' = \{x_2, x_3, x_4, x_5, x_6\}$
C) $X' = \{x_3, x_4, x_5, x_6\}$
D) $X' = \{x_4, x_5, x_6\}$

$\boxed{\text{B}}$ Considérese el conjunto de prototipos del ejercicio anterior. Si ejecutáramos el algoritmo de condensado de prototipos tras aplicar el algoritmo de edición de Wilson. ¿Qué conjunto de prototipos S obtendríamos como resultado del algoritmo de condensado con el vecino más cercano en distancia Euclídea?

A) $S = \{x_1, x_2, x_3, x_4, x_5, x_6\}$
B) $S = \{x_3, x_5\}$
C) $S = \{x_3, x_4, x_5\}$
D) $S = \{x_3, x_5, x_6\}$

$\boxed{\text{A}}$ 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 $\mathbf{x} = \begin{pmatrix}0\\1\end{pmatrix}$ dados los parámetros anteriores?

A) Clase A
B) Clase B
C) Clase C
D) En cualquiera de las tres

C Dados los parámetros Bernoulli de las clases del ejercicio anterior, ¿qué clases verían modificados los valores de sus parámetros si aplicáramos un suavizado por truncamiento simple con $\epsilon = 0.3$?

A) Sólo la clase A
B) La clase A y la clase B
C) La clase A y la clase C
D) Sólo la clase C

B (**Enunciado corregido**) Dado el siguiente conjunto de vectores enteros bidimensionales:

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

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

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

A En el caso de que la cantidad de muestras de entrenamiento no sea suficiente para estimar de forma robusta los parámetros de una distribución gaussiana, ¿qué tipo de matriz de covarianzas es la más recomendable en este caso?

A) Diagonal
B) Simétrica
C) Completa
D) Invertible

C Considérese una tarea de clasificación en C clases, donde cada clase está representada por una mixtura de gaussianas de I componentes, ¿cuántas matrices de covarianzas distintas es necesario estimar para esta tarea?

A) C
B) I
C) $C \times I$
D) $C + I$

C Los paradigmas de adaptación más habituales en sistemas interactivos son:

A) Aprendizaje on-line, pasivo y por refuerzo
B) Aprendizaje on-line, activo y pasivo
C) Aprendizaje on-line, activo y por refuerzo
D) Aprendizaje on-line y por refuerzo

Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2014

Apellidos: ☐ **Nombre:** ☐

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

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(2 puntos)** Tenemos $N = 12$ 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
x_{n1}	1	2	1	1	2	1	0	0	0	0	0	0
x_{n2}	2	1	0	1	2	0	0	0	0	0	0	0
x_{n3}	0	0	0	0	0	0	1	1	0	1	0	1
x_{n4}	1	0	2	1	0	2	1	2	1	1	2	1
x_{n5}	0	0	0	0	0	0	1	3	1	1	1	1
c_n	1	1	1	1	1	1	2	2	2	2	2	2

   a) Calcula los parámetros del clasificador multinomial más probable respecto a estos datos. (0.4 puntos)

   b) Suaviza los parámetros multinomial mediante Laplace con $\epsilon = 0.2$. (0.4 puntos)

   c) Suaviza los parámetros con descuento absoluto de $\epsilon = 0.1$ e interpolación usando como distribución generalizada, la distribución uniforme. (0.4 puntos)

   d) Suaviza los parámetros con descuento absoluto de $\epsilon = 0.1$ e interpolación usando como distribución generalizada, la distribución $g = (0.1\ 0.2\ 0.4\ 0.2\ 0.1)$, donde g_i es la probabilidad asociada a la dimensión i-ésima de $\mathbf{p}_c$. (0.4 puntos)

   e) Clasifica la muestra de test $y = (1\ 1\ 1\ 1\ 1)^t$ con el clasificador multinomial resultante de aplicar los parámetros suavizados del apartado b). (0.4 puntos)

Solución:

a)

$$p(1) = p(2) = \frac{6}{12} = 0.5$$

$$\hat{\mathbf{p}}_1 = \frac{1}{20}\begin{pmatrix} 1+2+1+1+2+1 \\ 2+1+0+1+2+0 \\ 0+0+0+0+0+0 \\ 1+0+2+1+0+2 \\ 0+0+0+0+0+0 \end{pmatrix} = \frac{1}{20}\begin{pmatrix} 8 \\ 6 \\ 0 \\ 6 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.4 \\ 0.3 \\ 0.0 \\ 0.3 \\ 0.0 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \frac{1}{20}\begin{pmatrix} 0+0+0+0+0+0 \\ 0+0+0+0+0+0 \\ 1+1+0+1+0+1 \\ 1+2+1+1+2+1 \\ 1+3+1+1+1+1 \end{pmatrix} = \frac{1}{20}\begin{pmatrix} 0 \\ 0 \\ 4 \\ 8 \\ 8 \end{pmatrix} = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.2 \\ 0.4 \\ 0.4 \end{pmatrix}$$

b)

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

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

c)

$$\hat{\mathbf{p}}_1 = \begin{pmatrix} 0.4 \\ 0.3 \\ 0.0 \\ 0.3 \\ 0.0 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_1 = \begin{pmatrix} 0.4 - 0.1 + \frac{1}{5} \cdot 0.3 \\ 0.3 - 0.1 + \frac{1}{5} \cdot 0.3 \\ 0.0 \qquad + \frac{1}{5} \cdot 0.3 \\ 0.3 - 0.1 + \frac{1}{5} \cdot 0.3 \\ 0.0 \qquad + \frac{1}{5} \cdot 0.3 \end{pmatrix} = \begin{pmatrix} 0.36 \\ 0.26 \\ 0.06 \\ 0.26 \\ 0.06 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.2 \\ 0.4 \\ 0.4 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_2 = \begin{pmatrix} 0.0 \qquad + \frac{1}{5} \cdot 0.3 \\ 0.0 \qquad + \frac{1}{5} \cdot 0.3 \\ 0.2 - 0.1 + \frac{1}{5} \cdot 0.3 \\ 0.4 - 0.1 + \frac{1}{5} \cdot 0.3 \\ 0.4 - 0.1 + \frac{1}{5} \cdot 0.3 \end{pmatrix} = \begin{pmatrix} 0.06 \\ 0.06 \\ 0.16 \\ 0.36 \\ 0.36 \end{pmatrix}$$

d)

$$\hat{\mathbf{p}}_1 = \begin{pmatrix} 0.4 \\ 0.3 \\ 0.0 \\ 0.3 \\ 0.0 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_1 = \begin{pmatrix} 0.4 - 0.1 + 0.1 \cdot 0.3 \\ 0.3 - 0.1 + 0.2 \cdot 0.3 \\ 0.0 \qquad + 0.4 \cdot 0.3 \\ 0.3 - 0.1 + 0.2 \cdot 0.3 \\ 0.0 \qquad + 0.1 \cdot 0.3 \end{pmatrix} = \begin{pmatrix} 0.33 \\ 0.26 \\ 0.12 \\ 0.26 \\ 0.03 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.2 \\ 0.4 \\ 0.4 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_2 = \begin{pmatrix} 0.0 \qquad + 0.1 \cdot 0.3 \\ 0.0 \qquad + 0.2 \cdot 0.3 \\ 0.2 - 0.1 + 0.4 \cdot 0.3 \\ 0.4 - 0.1 + 0.2 \cdot 0.3 \\ 0.4 - 0.1 + 0.1 \cdot 0.3 \end{pmatrix} = \begin{pmatrix} 0.03 \\ 0.06 \\ 0.22 \\ 0.36 \\ 0.33 \end{pmatrix}$$

e)

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

$$p(y = (1\ 1\ 1\ 1\ 1) \mid c = 1) = 0.30 \cdot 0.25 \cdot 0.10 \cdot 0.25 \cdot 0.10 = 0.0001875$$
$$p(y = (1\ 1\ 1\ 1\ 1) \mid c = 2) = 0.10 \cdot 0.10 \cdot 0.20 \cdot 0.30 \cdot 0.30 = 0.00018$$

La muestra y se clasifica en la clase 1.

2. **(2 puntos)** Sea A, B y C tres clases con priors $p(A) = 1/2$ y $p(B) = p(C) = 1/4$, y f.d.p. condicionales de clase gaussianas $p(\mathbf{x} \mid A) \sim N_2(\mu_A, \Sigma_A)$, $p(\mathbf{x} \mid B) \sim N_2(\mu_B, \Sigma_B)$ y $p(\mathbf{x} \mid C) \sim N_2(\mu_C, \Sigma_C)$

$$\mu_A = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Sigma_A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \quad \mu_B = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \quad \Sigma_B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \quad \mu_C = \begin{pmatrix} -2 \\ -2 \end{pmatrix} \quad \Sigma_C = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

a) Calcula funciones discriminantes para A, B y C. (1 punto)

b) Calcula la frontera de decisión entre las clases A y B. (0.25 puntos)

c) Calcula la frontera de decisión entre las clases B y C. (0.25 puntos)

d) Clasifica la muestra $\mathbf{y} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. (0.5 puntos)

Nota: $\Sigma_A^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$ y $\Sigma_B^{-1} = \Sigma_C^{-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$

Solución:

a) En general, la función discriminante de un clasificador basado en la regla de Bayes se define como

$$g_c(\mathbf{x}) = p(c) \cdot p(\mathbf{x}|c)$$

o de forma equivalente

$$g_c(\mathbf{x}) = \log p(c) + \log p(\mathbf{x}|c)$$

siendo

$$p(\mathbf{x} \mid c) \sim N_2(\mu_c, \Sigma_c) = (2\pi)^{-1} \cdot \left|\Sigma_c\right|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_\mathbf{c})^t \Sigma_c^{-1} (\mathbf{x} - \mu_\mathbf{c})\right)$$

aplicamos logaritmo neperiano y operamos para simplificar la expresión resultante eliminando la constante $(2\pi)^{-1}$

$$g_c(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^t \Sigma_c^{-1} \mathbf{x} + \mu_c^t \Sigma_c^{-1} \mathbf{x} + \left(\log p(c) - \frac{1}{2} \log \left|\Sigma_c\right| - \frac{1}{2} \mu_c^t \Sigma_c^{-1} \mu_c\right)$$

Las funciones discriminantes de las clases serían:

$$\begin{aligned} g_A(\mathbf{x}) =& -\frac{1}{2} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^t \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}^t \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &+ \left(\log \frac{1}{2} - \frac{1}{2} \log \left|\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\right| - \frac{1}{2} \begin{pmatrix} 0 \\ 0 \end{pmatrix}^t \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) \\ =& -\frac{1}{4}x_1^2 - \frac{1}{4}x_2^2 + \log \frac{1}{2} - \frac{1}{2} \log 4 \end{aligned}$$

$$\begin{aligned} g_B(\mathbf{x}) =& -\frac{1}{2} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^t \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \end{pmatrix}^t \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &+ \left(\log \frac{1}{4} - \frac{1}{2} \log \left|\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\right| - \frac{1}{2} \begin{pmatrix} 2 \\ 2 \end{pmatrix}^t \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 2 \end{pmatrix}\right) \\ =& -x_1^2 - x_1 x_2 - \frac{1}{2}x_2^2 + 6x_1 + 4x_2 + \log \frac{1}{4} - 10 \end{aligned}$$

$$\begin{aligned} g_C(\mathbf{x}) =& -\frac{1}{2} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^t \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} -2 \\ -2 \end{pmatrix}^t \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &+ \left(\log \frac{1}{4} - \frac{1}{2} \log \left|\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\right| - \frac{1}{2} \begin{pmatrix} -2 \\ -2 \end{pmatrix}^t \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} -2 \\ -2 \end{pmatrix}\right) \\ =& -x_1^2 - x_1 x_2 - \frac{1}{2}x_2^2 - 6x_1 - 4x_2 + \log \frac{1}{4} - 10 \end{aligned}$$

b) La frontera de decisión entre las clases A y B se obtiene igualando sus respectivas funciones discriminantes

$$g_A(\mathbf{x}) = g_B(\mathbf{x})$$

Por tanto,

$$-\frac{1}{4}x_1^2 - \frac{1}{4}x_2^2 + \log \frac{1}{2} - \frac{1}{2} \log 4 = -x_1^2 - x_1 x_2 - \frac{1}{2}x_2^2 + 6x_1 + 4x_2 + \log \frac{1}{4} - 10$$

que resulta en una curva cuadrática

$$\frac{3}{4}x_1^2 + x_1 x_2 + \frac{1}{4}x_2^2 - 6x_1 - 4x_2 + 10 = 0$$

c) Al igual que en el apartado b) igualamos las funciones discriminantes, pero en este caso las de las clase B y C

$$-x_1^2 - x_1x_2 - \frac{1}{2}x_2^2 + 6x_1 + 4x_2 + \log\frac{1}{4} - 10 = -x_1^2 - x_1x_2 - \frac{1}{2}x_2^2 - 6x_1 - 4x_2 + \log\frac{1}{4} - 10$$

que como esperábamos definen una frontera de decisión lineal

$$3x_1 + 2x_2 = 0 \rightarrow x_2 = -\frac{3}{2}x_1$$

dado que la matriz de covarianza es común a ambas clases.

d) Aplicamos la regla de clasificación a la muestra y

$$\hat{c}\,(y) = \arg\max_c g_c(y)$$

Esto es

$$\begin{aligned}
g_A(y) &= -\frac{1}{4}1^2 - \frac{1}{4}1^2 + \log\frac{1}{2} - \frac{1}{2}\log 4 = -1.89 \\
g_B(y) &= -1^2 - 1 - \frac{1}{2}1^2 + 6 + 4 + \log\frac{1}{4} - 10 = -3.89 \\
g_C(y) &= -1^2 - 1 - \frac{1}{2}1^2 - 6 - 4 + \log\frac{1}{4} - 10 = -23.89
\end{aligned}$$

Por tanto, $\hat{c}\,(y) = A$.