Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2013

Apellidos: ☐ **Nombre:** ☐

Profesor: ☐ **Carlos Martínez - Jorge Civera** ☐ **Roberto Paredes**

Cuestiones (3 puntos, 30 minutos, sin apuntes)

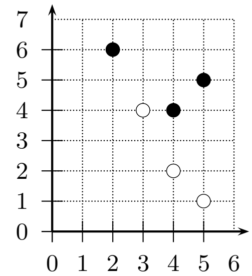
$\boxed{\text{A}}$ Dada una representación vectorial de los objetos, ¿Cuál de las siguientes distancias **no** es una métrica?

A) $d(x,y) = \sum_{i=1}^{d}(x_i - y_i)$.
B) $d(x,y) = \sum_{i=1}^{d}|x_i - y_i|$.
C) $d(x,y) = \sum_{i=1}^{d}(x_i - y_i)^2$.
D) $d(x,y) = (\sum_{i=1}^{d}|x_i - y_i|^3)^{1/3}$.

$\boxed{\text{D}}$ Dado un conjunto de datos de tamaño n, al calcular los K-vecinos más cercanos de $\mathbf{x}$ se obtienen k_c vecinos de la clase c, y esto permite:

A) una estimación directa de $p(\mathbf{x}|c)$, con $p(\mathbf{x}|c) = \frac{k_c}{n}$.
B) una estimación directa de $p(\mathbf{x}|c)$, con $p(\mathbf{x}|c) = \frac{k_c}{K}$.
C) una estimación directa de $p(c|\mathbf{x})$, con $p(c|\mathbf{x}) = \frac{k_c}{n}$.
D) una estimación directa de $p(c|\mathbf{x})$, con $p(c|\mathbf{x}) = \frac{k_c}{K}$.

$\boxed{\text{D}}$ La figura de la derecha muestra prototipos bidimensionales de 2 clases: $\circ$ y $\bullet$. Estima el error del clasificador por el vecino más próximo (en distancia Euclídea) considerando como conjunto de entrenamiento $\{(2,6),(3,4),(4,2)\}$ y conjunto de test $\{(4,4),(5,1),(5,5)\}$.



A) 0
B) 1/3
C) 1/2
D) 2/3

$\boxed{\text{D}}$ Considérese la probabilidad de error asintótica del clasificador por los k vecinos más próximos; esto es, su probabilidad de error cuando el número n de prototipos sobre los que se basa tiende a infinito. ¿Cuál de las siguientes condiciones sobre k es suficiente para garantizar la convergencia de esta probabilidad de error a la del clasificador de Bayes?

A) $k > 2$
B) $k = \frac{N}{10}$
C) $k = 2N$
D) Ninguna de las anteriores.

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

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

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

A) $\hat{p}(1) = \hat{p}(2) = \frac{1}{4}$, $\hat{\mathbf{p}}_1 = \left(\frac{1}{4}, \frac{1}{12}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{4}, \frac{1}{6}\right)^t$
B) $\hat{p}(1) = \hat{p}(2) = \frac{1}{2}$, $\hat{\mathbf{p}}_1 = \left(\frac{2}{3}, \frac{1}{6}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{2}, \frac{1}{3}\right)^t$
C) $\hat{p}(1) = \hat{p}(2) = \frac{1}{4}$, $\hat{\mathbf{p}}_1 = \left(\frac{2}{3}, \frac{1}{6}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{2}, \frac{1}{3}\right)^t$
D) $\hat{p}(1) = \hat{p}(2) = \frac{1}{2}$, $\hat{\mathbf{p}}_1 = \left(\frac{1}{4}, \frac{1}{12}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{4}, \frac{1}{6}\right)^t$

$\boxed{\text{B}}$ *Laplace* es una técnica de suavizado de parámetros que aplicamos al:

A) Clasificador de Bernoulli.
B) Clasificador multinomial.
C) Clasificador Gaussiano.
D) Ninguno de los anteriores.

$\boxed{\text{A}}$ Con respecto a los modelos de mixturas de gaussianas (GMM)

A) La estimación de sus parámetros requiere de procesos iterativos como el basado en el algoritmo EM.
B) La estimación de sus parámetros se hace de forma directa sin proceso iterativo.
C) Son modelos que no reflejan una distribución de probabilidad correcta.
D) Sus parámetros estimables son las medias y las varianzas de cada gaussiana, pero no los pesos asociados

$\boxed{\text{B}}$ El uso de kernels es adecuado cuando:

A) Los objetos son linealmente separables en el espacio de representación escogido.
B) Los objetos no son linealmente separables en el espacio de representación escogido.
C) Queremos proyectar los datos a un espacio de menor dimensionalidad linealmente separable.
D) Queremos garantizar que los datos quedan linealmente separables para cualquier kernel.

$\boxed{\text{C}}$ Esencialmente el algoritmo Kernel Perceptron lo que hace es:

A) Incrementar la importacia (peso) de las muestras correctamente clasificadas.
B) Decrementar la importacia (peso) de las muestras correctamente clasificadas.
C) Incrementar la importacia (peso) de las muestras incorrectamente clasificadas.
D) Decrementar la importacia (peso) de las muestras incorrectamente clasificadas.

$\boxed{\text{C}}$ El protocolo de interacción izquierda-derecha

A) Es un protocolo exclusivamente activo
B) Es un protocolo sólo aplicable en sistemas adaptativos
C) Es el protocolo más apropiado ante entradas secuenciales
D) Es un protocolo que no requiere de entrada

Examen de Teoría de Percepción - Segundo Parcial

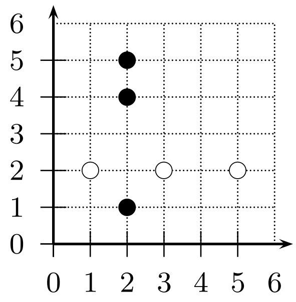
ETSINF, Universitat Politècnica de València, Junio de 2013

Apellidos: **Nombre:**

Profesor: ☐ **Carlos Martínez - Jorge Civera** ☐ **Roberto Paredes**

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(1 punto)** La figura de la derecha muestra prototipos bidimensionales de 2 clases: $x_1 = (1, 2, \circ), x_2 = (2, 1, \bullet), x_3 = (3, 2, \circ), x_4 = (2, 4, \bullet), x_5 = (5, 2, \circ), x_6 = (2, 5, \bullet)$. Aplica el algoritmo de edición de Wilson sobre el conjunto de prototipos por orden de índice utilizando vecino más próximo ($k = 1$) en distancia:



   a) Euclídea.

   b) Mahalanobis-diagonal por clase.

   Nota: Si es necesario, asume que $\infty \cdot 0 = 0$.

   Solución:

   a) El algoritmo de Wilson se reduce a recorrer iterativamente los prototipos tomando uno de ellos como muestra de test y el resto como muestras de entrenamiento, eliminando los prototipos mal clasificados. En este caso es el clasificador de vécino más próximo ($k = 1$) con distancia Euclídea convencional:

	Clase estimada	Conjunto resultante
$x_1 = (1, 2, \circ)$	$\bullet$	$\{x_2, x_3, x_4, x_5, x_6\}$
$x_2 = (2, 1, \bullet)$	$\circ$	$\{x_3, x_4, x_5, x_6\}$
$x_3 = (3, 2, \circ)$	$\circ$	$\{x_3, x_4, x_5, x_6\}$
$x_4 = (2, 4, \bullet)$	$\bullet$	$\{x_3, x_4, x_5, x_6\}$
$x_5 = (5, 2, \circ)$	$\circ$	$\{x_3, x_4, x_5, x_6\}$
$x_6 = (2, 5, \bullet)$	$\bullet$	$\{x_3, x_4, x_5, x_6\}$

   b) En este caso el algoritmo de Wilson utiliza el clasificador de vécino más próximo ($k = 1$) con distancia Mahalanobis-diagonal por clase:

$$d(\mathbf{y}, \mathbf{p}) = \sqrt{\sum_{i=1}^{D} \frac{1}{\sigma_{ic}^2}(y_i - p_i)^2}$$

   donde c es la clase de $\mathbf{p}$, y σ_{ic}^2 es la varianza de la componente i-ésima en la clase c.

   Aplicamos una primera iteración del algoritmo de Wilson. Para calcular la distancia de Mahalanobis-diagonal por clase entre cada prototipo $\mathbf{p}$ y un prototipo excluido $\mathbf{y}$, necesitamos calcular previamente las varianzas por clase y componente que en todos los casos son:

σ_{ic}^2	$i = 1$	$i = 2$
$c = \circ$	$\alpha > 0$	0
$c = \bullet$	0	$\beta > 0$

   La distancia entre un prototipo $\mathbf{p}$ de la clase $\circ$ y un prototipo excluido $\mathbf{y}$ de la clase $\bullet$ resulta en:

$$d(\bullet, \circ) = \sqrt{\frac{1}{\sigma_{1\circ}^2}(y_1 - p_1)^2 + \frac{1}{\sigma_{2\circ}^2}(y_2 - p_2)^2} = \sqrt{\frac{1}{\alpha}(y_1 - p_1)^2 + \frac{1}{0}(y_2 - p_2)^2} = +\infty$$

   Lo mismo ocurre cuando calculamos la distancia Mahalanobis-diagonal por clase entre un prototipo $\mathbf{p}$ de la clase $\bullet$ y un prototipo excluido $\mathbf{y}$ de la clase $\circ$. Sin embargo, no ocurre lo mismo cuando calculamos esta distancia entre prototipos de la misma clase. Por ejemplo, para dos prototipos de la clase $\circ$:

$$d(\circ, \circ) = \sqrt{\frac{1}{\sigma_{1\circ}^2}(y_1 - p_1)^2 + \frac{1}{\sigma_{2\circ}^2}(y_2 - p_2)^2} = \sqrt{\frac{1}{\alpha}(y_1 - p_1)^2 + \frac{1}{0}(0)^2} = \sqrt{\frac{1}{\alpha}(y_1 - p_1)^2}$$

   Por tanto, todos los prototipos son correctamente clasificados y el algoritmo de Wilson no elimina ninguno.

2. **(1 punto)** Dados los parámetros de un clasificador multinomial $\hat{\mathbf{p}}_1 = (0.5\ 0.3\ 0.2\ 0.0\ 0.0)$ y $\hat{\mathbf{p}}_2 = (0.0\ 0.0\ 0.3\ 0.3\ 0.4)$ con priors idénticas, clasifica la muestra de test $y = (1\ 1\ 1\ 1\ 1)$ tras aplicar los siguientes suavizados:

   a) Laplace con $\epsilon = 0.1$

   b) Descuento absoluto con $\epsilon = 0.05$ y backing-off utilizando como distribución generalizada, la distribución uniforme.

   Solución:

   a)

$$\hat{\mathbf{p}}_1 = \begin{pmatrix} 0.5 \\ 0.3 \\ 0.2 \\ 0.0 \\ 0.0 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_1 = \frac{1}{1.5}\begin{pmatrix} 0.5+0.1 \\ 0.3+0.1 \\ 0.2+0.1 \\ 0.0+0.1 \\ 0.0+0.1 \end{pmatrix} = \frac{1}{1.5}\begin{pmatrix} 0.6 \\ 0.4 \\ 0.3 \\ 0.1 \\ 0.1 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.3 \\ 0.3 \\ 0.4 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_2 = \frac{1}{1.5}\begin{pmatrix} 0.0+0.1 \\ 0.0+0.1 \\ 0.3+0.1 \\ 0.3+0.1 \\ 0.4+0.1 \end{pmatrix} = \frac{1}{1.5}\begin{pmatrix} 0.1 \\ 0.1 \\ 0.4 \\ 0.4 \\ 0.5 \end{pmatrix}$$

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

$$p(y = (1\ 1\ 1\ 1\ 1) \mid c = 1) = \frac{0.6}{1.5} \cdot \frac{0.4}{1.5} \cdot \frac{0.3}{1.5} \cdot \frac{0.1}{1.5} \cdot \frac{0.1}{1.5} = 0.00009$$
$$p(y = (1\ 1\ 1\ 1\ 1) \mid c = 2) = \frac{0.1}{1.5} \cdot \frac{0.1}{1.5} \cdot \frac{0.4}{1.5} \cdot \frac{0.4}{1.5} \cdot \frac{0.5}{1.5} = 0.0001$$

   La muestra y se clasifica en la clase 2.

   b)

$$\hat{\mathbf{p}}_1 = \begin{pmatrix} 0.5 \\ 0.3 \\ 0.2 \\ 0.0 \\ 0.0 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_1 = \begin{pmatrix} 0.5 - 0.05 & \\ 0.3 - 0.05 & \\ 0.2 - 0.05 & \\ 0.0 & + \frac{1}{2} \cdot 0.15 \\ 0.0 & + \frac{1}{2} \cdot 0.15 \end{pmatrix} = \begin{pmatrix} 0.45 \\ 0.25 \\ 0.15 \\ 0.075 \\ 0.075 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.3 \\ 0.3 \\ 0.4 \end{pmatrix} \rightarrow \hat{\mathbf{p}}_2 = \begin{pmatrix} 0.0 & + \frac{1}{2} \cdot 0.15 \\ 0.0 & + \frac{1}{2} \cdot 0.15 \\ 0.3 - 0.05 & \\ 0.3 - 0.05 & \\ 0.4 - 0.05 & \end{pmatrix} = \begin{pmatrix} 0.075 \\ 0.075 \\ 0.25 \\ 0.25 \\ 0.35 \end{pmatrix}$$

$$p(y = (1\ 1\ 1\ 1\ 1) \mid c = 1) = 0.45 \cdot 0.25 \cdot 0.15 \cdot 0.075 \cdot 0.075 = 0.00009$$
$$p(y = (1\ 1\ 1\ 1\ 1) \mid c = 2) = 0.075 \cdot 0.075 \cdot 0.25 \cdot 0.25 \cdot 0.35 = 0.0001$$

   La muestra y se clasifica en la clase 2.

3. **(1 punto)** Sea la mixtura de gaussianas dada por las siguientes dos distribuciones gaussianas y sus correspondientes pesos:

$$\mathcal{N}_1(w_1, \mu_1, \Sigma_1) \quad \mu_1 = (2\ 4)^t \quad \Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \quad w_1 = \frac{1}{3}$$

$$\mathcal{N}_2(w_2, \mu_2, \Sigma_2) \quad \mu_2 = (5\ \ -1)^t \quad \Sigma_2 = \begin{pmatrix} \frac{9}{2} & 0 \\ 0 & 2 \end{pmatrix} \quad w_2 = \frac{2}{3}$$

Obtener la probabilidad de que la mixtura genere el dato $x = (3\ 0)^t$.

Solución:

$$p(x|\mathcal{N}_1) = \frac{1}{|\Sigma_1|^{\frac{1}{2}}(2\pi)^{\frac{d}{2}}} e^{-\frac{1}{2}(x-\mu_1)^t \Sigma_1^{-1}(x-\mu_1)} = \frac{1}{2 \cdot (2\pi)} e^{-\frac{1}{2}(1\ -4)\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{pmatrix}(1\ -4)^t} = \frac{1}{4\pi} e^{-\frac{1}{2}(1\ -1)(1\ -4)^t} = \frac{1}{4\pi} e^{-\frac{5}{2}}$$

$$p(x|\mathcal{N}_2) = \frac{1}{|\Sigma_2|^{\frac{1}{2}}(2\pi)^{\frac{d}{2}}} e^{-\frac{1}{2}(x-\mu_2)^t \Sigma_2^{-1}(x-\mu_2)} = \frac{1}{3 \cdot (2\pi)} e^{-\frac{1}{2}(-2\ 1)\begin{pmatrix} \frac{2}{9} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}(-2\ 1)^t} = \frac{1}{6\pi} e^{-\frac{1}{2}(-\frac{4}{9}\ \frac{1}{2})(-2\ 1)^t} =$$
$$\frac{1}{6\pi} e^{-\frac{1}{2}\frac{25}{18}} = \frac{1}{6\pi} e^{-\frac{25}{36}}$$

$$p(x|\{\mathcal{N}_1, \mathcal{N}_2\}) = w_1 \cdot p(x|\mathcal{N}_1) + w_2 \cdot p(x|\mathcal{N}_2) = \frac{1}{3} \cdot \frac{1}{4\pi} e^{-\frac{5}{2}} + \frac{2}{3} \cdot \frac{1}{6\pi} e^{-\frac{25}{36}} = \frac{1}{12\pi} e^{-\frac{5}{2}} + \frac{1}{9\pi} e^{-\frac{25}{36}} \approx$$
$$0.0022 + 0.0177 = 0.0198$$

4. **(1 punto)** Sea la siguiente función kernel, $K(x,y) = (x \cdot y + 1)^d$, con $d = 2$. Sea el siguiente conjunto de aprendizaje $X = \{(x_1, +1), (x_2, +1), (x_3, +1), (x_4, -1), (x_5, -1)\}$, con: $x_1 = (1,1)$, $x_2 = (1,0)$, $x_3 = (0,1)$, $x_4 = (2,1)$, $x_5 = (1,2)$.

   Se tienen los siguientes pesos α_i: $\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\alpha_4 = \alpha_5 = 0.5$.

   Clasifica la muestra de test $y = (0,2)$.

   Solución:

$$g(y) = \sum_{i=1}^{n} \alpha_i c_i K(y, x_i) + \sum_{i=1}^{n} \alpha_i c_i =$$

$$1 \cdot (+1) \cdot K(y,x_1) + 1 \cdot (+1) \cdot K(y,x_2) + 1 \cdot (+1) \cdot K(y,x_3) + (0.5) \cdot (-1) \cdot K(y,x_4) + (0.5) \cdot (-1) \cdot K(y,x_5)$$

$$+1 \cdot (+1) + 1 \cdot (+1) + 1 \cdot (+1) + (0.5) \cdot (-1) + (0.5) \cdot (-1) =$$

$$(+1) \cdot 9 + (+1) \cdot 1 + (+1) \cdot 9 + (-0.5) \cdot 9 + (-0.5) \cdot 25 + (+1) + (+1) + (+1) + (-0.5) + (-0.5) = 19 - 17 + 2 = 4$$

   Por tanto, al ser $g(x) > 0$, se clasifica en $+1$.