Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2017

Apellidos: ☐ **Nombre:** ☐

Profesor: ☐ **Jorge Civera** ☐ **Carlos Martínez**

Cuestiones (2 puntos, 30 minutos, sin apuntes)

$\boxed{\text{A}}$ La condición de Mercer para caracterizar una función Kernel $K(x, y)$ se puede formular como:

A) $\sum_{i=1}^{n}\sum_{j=1}^{n} K(\mathbf{x}_i, \mathbf{x}_j)\, c_i\, c_j \geq 0 \quad \forall c_i, c_j \in \mathbb{R}$
B) $\mathbf{z}^t\mathbf{K}\mathbf{z} \leq 0 \quad \forall \mathbf{z} \in \mathbb{R}^{n\times 1}$ con $\mathbf{z} \neq \mathbf{0}$ y matriz Gramm $\mathbf{K}$
C) $\sum_{i=1}^{n} K(\mathbf{x}_i, \mathbf{x}_i)\, c_i \geq 0 \quad \forall c_i \in \mathbb{R}$
D) $\sum_{i=1}^{n}\sum_{j=1}^{n} K(\mathbf{x}_i, \mathbf{x}_j)\, c_i \geq 0 \quad \forall c_i \in \mathbb{R}$

$\boxed{\text{C}}$ Dados el prototipo Bernoulli $\hat{p} = (0.85\ 0.05\ 0.5\ 0.15\ 0.95)^t$ estimado a partir de un conjunto de vectores binarios, y su correspondiente prototipo Bernoulli suavizado mediante truncamiento simple $\tilde{p} = (0.85\ 0.1\ 0.5\ 0.15\ 0.9)^t$. ¿Qué valor de ϵ se ha utilizado para obtener dicho prototipo Bernoulli suavizado?

A) $\epsilon = 0.00$
B) $\epsilon = 0.05$
C) $\epsilon = 0.10$
D) $\epsilon = 0.15$

$\boxed{\text{B}}$ Sea el conjunto de datos:

x_1	0	3	0	1	2	4	3	3	4	1
x_2	2	4	3	3	2	1	0	0	0	2
x_3	5	1	4	4	5	2	3	1	1	3
x_4	0	1	0	0	0	6	3	4	5	4
c	A	A	A	A	A	B	B	B	B	B

Indica la estimación por máxima verosimilitud de los parámetros de las distribuciones de Multinomiales para las clases A y B con ese conjunto de datos.

A) $\mathbf{p}_A = \left(\frac{6}{21}\,\frac{14}{17}\,\frac{19}{29}\,\frac{1}{23}\right), \mathbf{p}_B = \left(\frac{15}{21}\,\frac{3}{17}\,\frac{10}{29}\,\frac{22}{23}\right)$
B) $\mathbf{p}_A = \left(\frac{3}{20}\,\frac{7}{20}\,\frac{19}{40}\,\frac{1}{40}\right), \mathbf{p}_B = \left(\frac{3}{10}\,\frac{3}{50}\,\frac{1}{5}\,\frac{11}{25}\right)$
C) $\mathbf{p}_A = \left(\frac{1}{15}\,\frac{7}{45}\,\frac{19}{90}\,\frac{1}{90}\right), \mathbf{p}_B = \left(\frac{1}{6}\,\frac{1}{30}\,\frac{1}{9}\,\frac{11}{45}\right)$
D) $\mathbf{p}_A = \left(\frac{3}{20}\,\frac{7}{20}\,\frac{19}{40}\,\frac{1}{40}\right), \mathbf{p}_B = \left(\frac{7}{10}\,\frac{47}{50}\,\frac{4}{5}\,\frac{14}{25}\right)$

$\boxed{\text{D}}$ Un clasificador gaussiano sobre un espacio vectorial de dimensión D:

A) Es, en general, lineal.
B) Tiene un número de parámetros igual a D.
C) No puede expresarse como una función discriminante.
D) Tiene como parámetros de la probabilidad condicional la media y la matriz de covarianzas.

C Sea A y B dos clases con igual probabilidad a priori y probabilidades condicionales de clase gaussianas:

$$p(\mathbf{x} \mid A) \sim N_2(\mu_A, \Sigma_A) \qquad \text{y} \qquad p(\mathbf{x} \mid B) \sim N_2(\mu_B, \Sigma_B)$$

con

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

¿Cuál de los siguientes pares de funciones discriminantes **no** define un clasificador equivalente al clasificador Gaussiano dado?

A) $g_A(x) = x_1 + \log \frac{1}{2} - \frac{1}{2}$ $\quad g_B(x) = \frac{1}{2}x_2 + \log \frac{1}{2} - \frac{1}{4}$
B) $g_A(x) = x_1 - \frac{1}{4}$ $\quad g_B(x) = \frac{1}{2}x_2$
C) $g_A(x) = 4x_1 + 2x_2 - 1$ $\quad g_B(x) = 0$
D) Todos los anteriores pares de funciones discriminantes son equivalentes al clasificador Gaussiano dado

B Sea X^k el conjunto de los $k \in \mathbb{N}^+$ prototipos más próximos a $\mathbf{y}$, ¿cuál de las siguientes expresiones representa un clasificador basado en los k-vecinos más cercanos?

A) $\hat{c}\,(y) = \arg\min_c |X^k \cap X_c|$
B) $\hat{c}\,(y) = \arg\max_c |X^k \cap X_c|$
C) $\hat{c}\,(y) = \arg\min_c \min_{\mathbf{x} \in X_c} d(\mathbf{x}, \mathbf{y})$
D) $\hat{c}\,(y) = \arg\min_c \max_{\mathbf{x} \in X_c} d(\mathbf{x}, \mathbf{y})$

A Con un algoritmo de edición se consigue:

A) Eliminar prototipos que, generalmente, provocan errores de clasificación.
B) Reducir considerablemente el coste computacional de la clasificación.
C) Convertir las fronteras de decisión entre clases, usando NN, en lineales.
D) Seleccionar aquellos prototipos más cercanos a las fronteras de decisión.

C En general, ¿cuál de las siguientes enumeraciones de clasificadores está ordenada de menor a mayor *bias* (de izquierda a derecha)?

A) k-NN, Multinomial, Gaussiano
B) Gaussiano, k-NN, Multinomial
C) k-NN, Gaussiano, Multinomial
D) Gaussiano, Multinomial, k-NN

Examen de Teoría de Percepción - Segundo Parcial

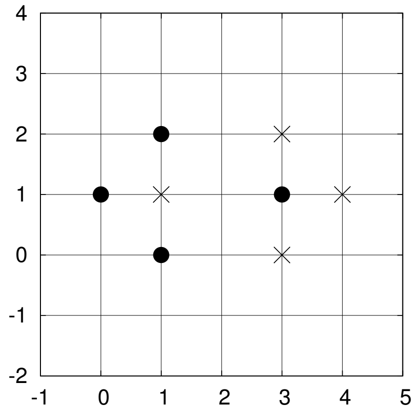
ETSINF, Universitat Politècnica de València, Junio de 2017

Apellidos: ______ **Nombre:** ______

Profesor: ☐ **Jorge Civera** ☐ **Carlos Martínez**

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(2 puntos)**
La figura de la derecha muestra prototipos bidimensionales de 2 clases, $X = \{x_1 = (1, 1, \times), x_2 = (3, 1, \bullet), x_3 = (3, 0, \times), x_4 = (1, 0, \bullet), x_5 = (4, 1, \times), x_6 = (0, 1, \bullet), x_7 = (3, 2, \times), x_8 = (1, 2, \bullet)\}$. Considera la utilización de un clasificador de vecino más cercano en distancia L_1. Se pide:



   a) Representa gráficamente la frontera y regiones de decisión.

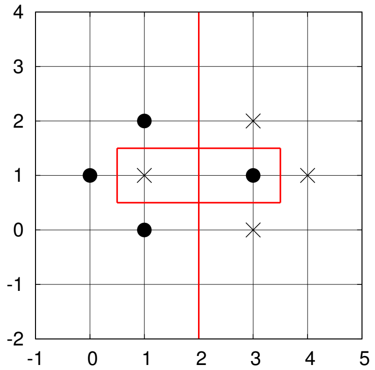
   b) Aplica el algoritmo de edición de Wilson visitando los prototipos por valor de índice creciente y representa gráficamente la frontera y regiones de decisión resultantes. En caso de empate, clasifica en la clase correcta.

   c) Sobre el conjunto de prototipos resultantes del apartado anterior, aplica el algoritmo de condensado de Hart visitando los prototipos por valor de índice creciente y representa gráficamente la frontera y regiones de decisión resultantes. En caso de empate, clasifica en la clase correcta.

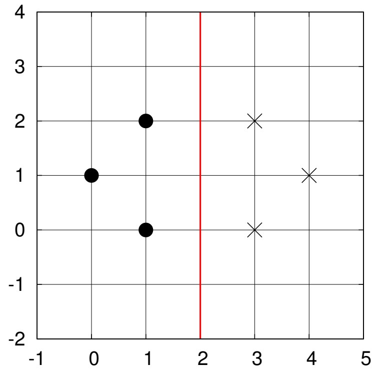
   d) Aplica de nuevo el algoritmo de edición de Wilson, pero siendo los prototipos x_1 y x_2 los últimos en ser visitados. Compara la frontera y regiones de decisión resultantes con las del apartado anterior.

Solución:

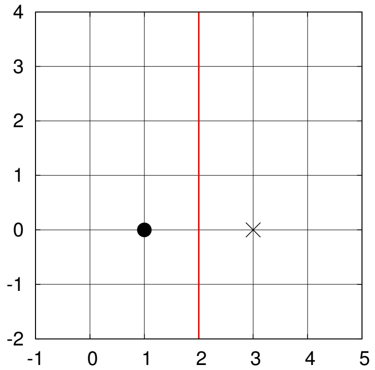
a)



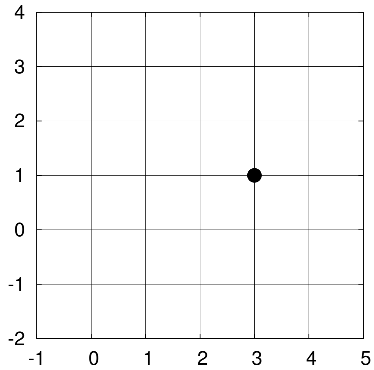
b)



c)



d)



2. **(1 punto)** Sea la función Kernel $K(\mathbf{x},\mathbf{y}) = \exp(\mathbf{x}^t \cdot \mathbf{y})$ y sea el conjunto de entrenamiento $\mathbf{x}_1 = (1,0)$, $\mathbf{x}_2 = (-1,0)$, $\mathbf{x}_3 = (-1,1)$, $\mathbf{x}_4 = (0,-1)$, siendo $\mathbf{x}_1, \mathbf{x}_3 \in -1$ y $\mathbf{x}_2, \mathbf{x}_4 \in +1$.

   a) Aplicar una iteración de Kernel perceptron partiendo de $\alpha = (0,0,0,0)$, indicando el valor final de α. (**0.75 puntos**)

   b) ¿Tiene sentido aplicar Kernels en este conjunto de entrenamiento? (**0.25 puntos**)

   Solución:

   a) $g(\mathbf{x}_1) = 0$, $c_1 g(\mathbf{x}_1) = 0 \leq 0$, $\alpha = (1,0,0,0)$
   $g(\mathbf{x}_2) = 1 \cdot (-1)\exp(\mathbf{x}_1^t \cdot \mathbf{x_2}) + 1 \cdot (-1) = -\exp(-1) - 1$, $c_2 g(\mathbf{x}_2) = -(\exp(-1)+1) \leq 0$, $\alpha = (1,1,0,0)$
   $g(\mathbf{x}_3) = 1 \cdot (-1)\exp(\mathbf{x}_1^t \cdot \mathbf{x_3}) + 1 \cdot (+1)\exp(\mathbf{x}_2^t \cdot \mathbf{x_3}) + 1 \cdot (-1) + 1 \cdot (+1) = -\exp(-1) + \exp(1)$, $c_3 g(\mathbf{x}_3) = \exp(-1) - \exp(1) \leq 0$, $\alpha = (1,1,1,0)$
   $g(\mathbf{x}_4) = 1 \cdot (-1)\exp(\mathbf{x}_1^t \cdot \mathbf{x_4}) + 1 \cdot (+1)\exp(\mathbf{x}_2^t \cdot \mathbf{x_4}) + 1 \cdot (-1)\exp(\mathbf{x}_3^t \cdot \mathbf{x_4}) + 1 \cdot (-1) + 1 \cdot (+1) + 1 \cdot (-1) = -\exp(0) + \exp(0) - \exp(-1) - 1 = -(\exp(-1)+1)$, $c_4 g(\mathbf{x}_4) = -(\exp(-1)+1) \leq 0$, $\alpha = (1,1,1,1)$
   Resultado final: $\alpha = (1,1,1,1)$

   b) No sería necesario, pues las muestras ya son linealmente separables.

3. **(1 punto)** Sean las siguientes muestras y clasificadores:

   $\mathbf{x}_1 = (1,1) \in +1 \qquad \mathbf{x}_2 = (2,1) \in -1 \qquad \mathbf{x}_3 = (2,2) \in +1 \qquad \mathbf{x}_4 = (1,3) \in -1$

   $$g_1(\mathbf{z}) = \begin{cases} +1 & z_1 \leq 1.5 \\ -1 & z_1 > 1.5 \end{cases} \quad g_2(\mathbf{z}) = \begin{cases} +1 & z_2 \leq 1.5 \\ -1 & z_2 > 1.5 \end{cases} \quad g_3(\mathbf{z}) = \begin{cases} +1 & z_1 + z_2 > 4 \\ -1 & z_1 + z_2 \leq 4 \end{cases} \quad g_4(\mathbf{z}) = \begin{cases} +1 & z_2 - z_1 \leq 1 \\ -1 & z_2 - z_1 > 1 \end{cases}$$

   Aplicar una iteración de AdaBoost para ese conjunto de datos y clasificadores indicando:

   a) Clasificador escogido C_1.

   b) Valor de ϵ_1.

   c) Valor de α_1.

   d) Actualización de los pesos para la siguiente iteración ($w^{(2)}$).

   Solución:

   Tabla de acierto/fallo:

	g_1	g_2	g_3	g_4
$\mathbf{x}_1$	✓	✓	X	✓
$\mathbf{x}_2$	✓	X	✓	X
$\mathbf{x}_3$	X	X	X	✓
$\mathbf{x}_4$	X	✓	✓	✓

   Pesos iniciales: $w^{(1)} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$

   Sumatorio de los $w^{(1)}$ de las muestras incorrectas:

g_1	g_2	g_3	g_4
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$

   $C_1 = g_4$
   $\epsilon_1 = \frac{1}{4}$
   $\alpha_1 = \frac{1}{2}\ln 3$

	$w^{(1)}\exp(-y_i \alpha_1 C_1(x_i))$
$\mathbf{x}_1$	$\frac{1}{4\sqrt{3}}$
$\mathbf{x}_2$	$\frac{\sqrt{3}}{4}$
$\mathbf{x}_3$	$\frac{1}{4\sqrt{3}}$
$\mathbf{x}_4$	$\frac{1}{4\sqrt{3}}$
Suma total	$\frac{\sqrt{3}}{2}$

   $w^{(2)} = \left(\frac{1}{6}, \frac{1}{2}, \frac{1}{6}, \frac{1}{6}\right)$