Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2017

Apellidos: ______________________ **Nombre:** ______________

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

Cuestiones (2 puntos, 30 minutos, sin apuntes)

$\boxed{\text{A}}$ Se define la función kernel $K(\mathbf{x},\mathbf{y}) = \frac{1}{d(\mathbf{x},\mathbf{y})+1}$ siendo $d(\mathbf{x},\mathbf{y})$ la distancia de Hamming $\left(d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{D} 1 - \delta(x_i, y_i)\right)$, ¿cuál de las siguientes funciones **no** es un kernel?

A) $K(\mathbf{x},\mathbf{y}) = d(\mathbf{x},\mathbf{y}) + 1$
B) $K(\mathbf{x},\mathbf{y}) = \frac{1}{d(\mathbf{x},\mathbf{y})+1} + 1$
C) $K(\mathbf{x},\mathbf{y}) = e^{\frac{1}{d(\mathbf{x},\mathbf{y})+1}}$
D) $K(\mathbf{x},\mathbf{y}) = \frac{1}{2\cdot d(\mathbf{x},\mathbf{y})+2}$

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

x_1	0	1	1	0	0	1	1	1	1	0
x_2	0	0	0	1	0	1	0	0	1	1
x_3	1	1	0	0	1	0	0	0	1	0
c	A	A	A	A	A	B	B	B	B	B

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

A) $\mathbf{p}_A = \left(\frac{3}{5}\,\frac{4}{5}\,\frac{2}{5}\right), \mathbf{p}_B = \left(\frac{4}{5}\,\frac{3}{5}\,\frac{1}{5}\right)$
B) $\mathbf{p}_A = \left(\frac{2}{5}\,\frac{1}{5}\,\frac{3}{5}\right), \mathbf{p}_B = \left(\frac{4}{5}\,\frac{3}{5}\,\frac{1}{5}\right)$
C) $\mathbf{p}_A = \left(\frac{2}{5}\,\frac{1}{5}\,\frac{3}{5}\right), \mathbf{p}_B = \left(\frac{1}{5}\,\frac{2}{5}\,\frac{4}{5}\right)$
D) $\mathbf{p}_A = \left(\frac{7}{15}\,\frac{7}{15}\,\frac{7}{15}\right), \mathbf{p}_B = \left(\frac{8}{15}\,\frac{8}{15}\,\frac{8}{15}\right)$

$\boxed{\text{C}}$ Dados el prototipo multinomial $\hat{p} = (0.3\ 0.0\ 0.4\ 0.0\ 0.5)^t$ estimado a partir de un conjunto de vectores de cuentas, y su correspondiente prototipo multinomial suavizado $\tilde{p} = (0.2\ 0.15\ 0.3\ 0.15\ 0.4)^t$. ¿Qué tipo de suavizado ha sido aplicado?

A) Laplace con $\epsilon = 0.1$
B) Laplace con $\epsilon = 0.2$
C) Descuento absoluto con $\epsilon = 0.1$ aplicando backing-off con distribución uniforme
D) Descuento absoluto con $\epsilon = 0.1$ aplicando interpolación con distribución uniforme

$\boxed{\text{D}}$ ¿Cuál de las siguientes expresiones representa un clasificador basado en el vecino más cercano?

A) $\hat{c}\,(y) = \arg\min_{c} \max_{\mathbf{x}\in X_c} d(\mathbf{x},\mathbf{y})$
B) $\hat{c}\,(y) = \arg\max_{c} \max_{\mathbf{x}\in X_c} d(\mathbf{x},\mathbf{y})$
C) $\hat{c}\,(y) = \arg\max_{c} \min_{\mathbf{x}\in X_c} d(\mathbf{x},\mathbf{y})$
D) $\hat{c}\,(y) = \arg\min_{c} \min_{\mathbf{x}\in X_c} d(\mathbf{x},\mathbf{y})$

B A diferencia de los clasificadores basados en Bayes, el clasificador k-NN:

A) Sólo puede aplicarse a datos no vectoriales.
B) Estima directamente $\hat{P}(c|x)$.
C) Alcanza una cota de error inferior a Bayes.
D) Crea un modelo basado en inferencia sobre los prototipos.

D Para que el clasificador k-NN alcance el error de Bayes, siendo n el número de prototipos:

A) Debe usar un valor n potencialmente infinito, independientemente del k usado.
B) Debe usar un valor k potencialmente infinito, independientemente de n.
C) Debe usar un valores k y n potencialmente infinitos, con relación constante entre ellos.
D) Debe usar un valores k y n potencialmente infinitos, pero con k de crecimiento mucho más lento que n.

A La distancia de Mahalanobis-diagonal:

A) Equivale a prenormalizar cada componente por la desviación típica y usar distancia euclídea.
B) Asigna pesos distintos por cada prototipo considerado.
C) Usa las varianzas por clase.
D) Incrementa la medida de distancia para las componentes que presentan una mayor dispersión.

C En general, ¿cuál de las siguientes enumeraciones de clasificadores está ordenada de mayor a menor *variance* (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 - Recuperación 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. **(1.5 puntos)** Sea A y B dos clases con priors $p(A) = 1/4$ y $p(B) = 3/4$, y f.d.p. condicionales de clase gaussianas $p(\mathbf{x} \mid A) \sim N_2(\mu_A, \Sigma_A)$ y $p(\mathbf{x} \mid B) \sim N_2(\mu_B, \Sigma_B)$

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

   a) Calcula funciones discriminantes para A y B. (0.5 puntos)

   b) Calcula la frontera de decisión entre las clases A y B. (0.5 puntos)

   c) Representa gráficamente la frontera y las regiones de decisión. (0.5 puntos)

Solución:
a) En el caso de matrices de covarianza común, la función discriminante se define como:

$$g_c(\mathbf{x}) = \mathbf{w}_c^t \, \mathbf{x} + b_c$$

donde

$$\mathbf{w}_c = \Sigma^{-1}\mu_c \qquad b_c = \log P(c) - \frac{1}{2}\mu_c^t \Sigma^{-1} \mu_c$$

Las funciones discriminantes de las clases serían:

$$\begin{aligned}
g_A(\mathbf{x}) &= \mathbf{w}_A^t \, \mathbf{x} + b_A \\
\mathbf{w}_A &= \Sigma^{-1}\mu_A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
b_A &= \log P(A) - \frac{1}{2}\mu_A^t \Sigma^{-1}\mu_A = \log\frac{1}{4} - \frac{1}{2}\begin{pmatrix} 0 \\ 0 \end{pmatrix}^t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = -\log 4 \\
g_A(\mathbf{x}) &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}^t \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \log 4 = -\log 4
\end{aligned}$$

$$\begin{aligned}
g_B(\mathbf{x}) &= \mathbf{w}_B^t \, \mathbf{x} + b_B \\
\mathbf{w}_B &= \Sigma^{-1}\mu_B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
b_B &= \log P(B) - \frac{1}{2}\mu_B^t \Sigma^{-1}\mu_B = \log\frac{3}{4} - \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}^t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \log\frac{3}{4} - 1 \\
g_B(\mathbf{x}) &= \begin{pmatrix} 1 \\ 1 \end{pmatrix}^t \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \log\frac{3}{4} - 1 = x_1 + x_2 + \log 3 - \log 4 - 1
\end{aligned}$$

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

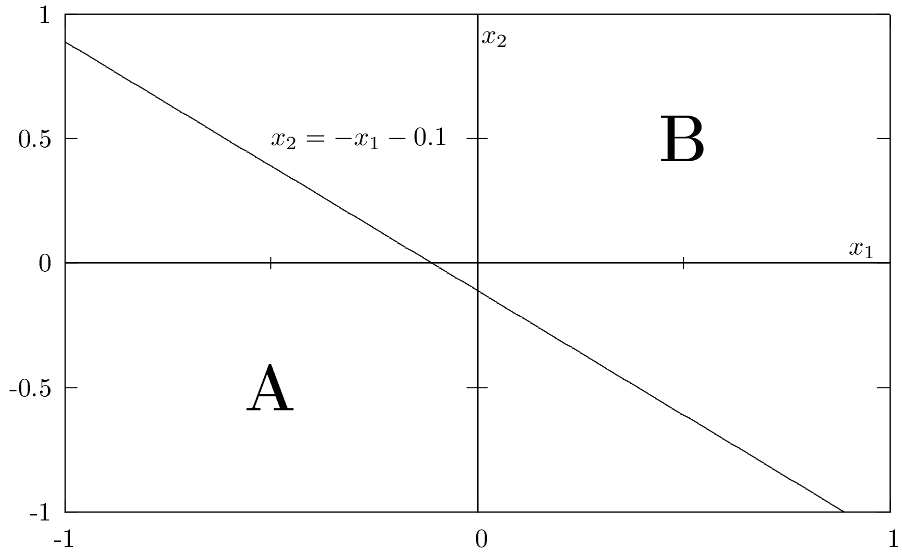
$$g_A(\mathbf{x}) = g_B(\mathbf{x})$$

Por tanto,

$$\begin{aligned}
-\log 4 &= x_1 + x_2 + \log 3 - \log 4 - 1 \\
0 &= x_1 + x_2 + \log 3 - 1 \\
x_2 &= -x_1 - \log 3 + 1 = -x_1 - 0.1
\end{aligned}$$

que es una frontera lineal.

c)



2. (1.5 **puntos**) Sea la función Kernel $K(\mathbf{x},\mathbf{y}) = \frac{1}{d(\mathbf{x},\mathbf{y})+1}$ siendo $d(\mathbf{x},\mathbf{y})$ la distancia de Hamming $\left(d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{D} 1 - \delta(x_i, y_i)\right)$. Sea el conjunto de entrenamiento $\mathbf{x}_1 = (0,0)$, $\mathbf{x}_2 = (0,1)$, $\mathbf{x}_3 = (1,0)$, $\mathbf{x}_4 = (1,1)$, siendo $\mathbf{x}_1, \mathbf{x}_4 \in +1$ y $\mathbf{x}_2, \mathbf{x}_3 \in -1$.

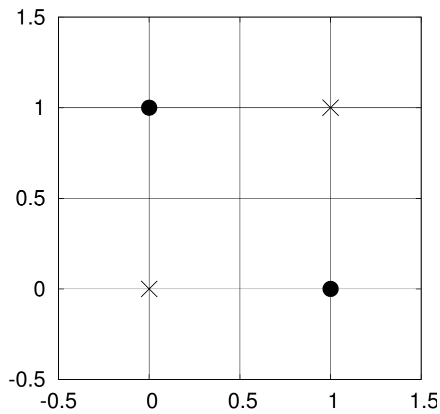
   a) Representa gráficamente el conjunto de entrenamiento y di si es o no linealmente separable. (**0.25 puntos**)

   b) Aplica una iteración de Kernel Perceptron partiendo de $\alpha = (0,0,0,0)$, indicando el valor final de α. (**0.75 puntos**)

   c) ¿Cuál es el error de clasificación de la función $g(x)$ resultante del apartado anterior sobre el conjunto de entrenamiento? (**0.5 puntos**)

Solución:

a) Como se puede observar en la figura el conjunto de entrenamiento no es linealmente separable.



b) La matriz de Gramm sobre el conjunto de entrenamiento para la función Kernel definida en el enunciado es

$$\mathbf{K} = \begin{array}{c|cccc} & (0,0) & (0,1) & (1,0) & (1,1) \\ (0,0) & 1 & 1/2 & 1/2 & 1/3 \\ (0,1) & 1/2 & 1 & 1/3 & 1/2 \\ (1,0) & 1/2 & 1/3 & 1 & 1/2 \\ (1,1) & 1/3 & 1/2 & 1/2 & 1 \end{array}$$

La traza del algoritmo Kernel Perceptron es la siguiente:

	$g(x)$	$g(x_i)$	c_i	error	α
x_1	0	0	$+1$	Sí	$(1,0,0,0)$
x_2	$K(x_1,x)+1$	$3/2$	-1	Sí	$(1,1,0,0)$
x_3	$K(x_1,x)+1-K(x_2,x)-1$	$1/6$	-1	Sí	$(1,1,1,0)$
x_4	$K(x_1,x)+1-K(x_2,x)-1-K(x_3,x)-1$	$-5/3$	$+1$	Sí	$(1,1,1,1)$

c) Considerando el vector de pesos $\alpha = (1,1,1,1)$ obtenidos, la función discriminante es:

$$\begin{aligned} g(x) &= K(x_1,x)+1-K(x_2,x)-1-K(x_3,x)-1+K(x_4,x)+1 \\ &= K(x_1,x)-K(x_2,x)-K(x_3,x)+K(x_4,x) \end{aligned}$$

	$g(x_i)$					c_i	error
x_1	1	-1/2	-1/2	+1/3 =	1/3	$+1$	No
x_2	1/2	-1	-1/2	+1/2 =	-1/2	-1	No
x_3	1/2	-1/2	-1	+1/2 =	-1/2	-1	No
x_4	1/3	-1/2	-1/2	+1 =	1/3	$+1$	No

Por tanto, el error de clasificación es cero y la función Kernel empleada proyecta el conjunto de entrenamiento a un espacio donde son linealmente separables.

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

$\mathbf{x}_1 = (0,0) \in +1$ $\quad$ $\mathbf{x}_2 = (2,2) \in -1$ $\quad$ $\mathbf{x}_3 = (1,2) \in +1$ $\quad$ $\mathbf{x}_4 = (0,1) \in -1$ $\quad$ $\mathbf{x}_5 = (-1,1) \in +1$

$$g_1(\mathbf{z}) = \begin{cases} +1 & z_1 > 0 \\ -1 & z_1 \leq 0 \end{cases} \quad g_2(\mathbf{z}) = \begin{cases} +1 & z_2 > 1 \\ -1 & z_2 \leq 1 \end{cases} \quad g_3(\mathbf{z}) = \begin{cases} +1 & z_2 - z_1 > 0 \\ -1 & z_2 - z_1 \leq 0 \end{cases} \quad g_4(\mathbf{z}) = \begin{cases} +1 & z_1 + z_2 \leq 3 \\ -1 & z_1 + z_2 > 3 \end{cases}$$

Aplica 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	X	X	✓
$\mathbf{x}_2$	X	X	✓	✓
$\mathbf{x}_3$	✓	✓	✓	✓
$\mathbf{x}_4$	✓	✓	X	X
$\mathbf{x}_5$	X	X	✓	✓

Pesos iniciales: $w^{(1)} = \left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$

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

g_1	g_2	g_3	g_4
$\frac{3}{5}$	$\frac{3}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

$C_1 = g_4$
$\epsilon_1 = \frac{1}{5}$
$\alpha_1 = \frac{1}{2}\ln 4 = \ln 2$

	$w^{(1)} \exp(-y_i \alpha_1 C_1(x_i))$
$\mathbf{x}_1$	$\frac{1}{10}$
$\mathbf{x}_2$	$\frac{1}{10}$
$\mathbf{x}_3$	$\frac{1}{10}$
$\mathbf{x}_4$	$\frac{2}{5} = \frac{4}{10}$
$\mathbf{x}_5$	$\frac{1}{10}$
Suma total	$\frac{4}{5} = \frac{8}{10}$

$w^{(2)} = \left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2}, \frac{1}{8}\right)$