Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2018

Apellidos: ☐ **Nombre:** ☐

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

Cuestiones (2 puntos, 30 minutos, sin apuntes)

B Dado un conjunto de entrenamiento $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, una función K es kernel sobre el mismo si la matriz Gramm $\mathbf{K}$ asociada es:

A) Simétrica
B) Semidefinida positiva
C) Triangular superior
D) Singular

D Sea un conjunto de muestras de entrenamiento $X = \{(\mathbf{x}_1, -1), (\mathbf{x}_2, -1), (\mathbf{x}_3, +1)\}$ para la cual se define la función kernel $K(\mathbf{x}, \mathbf{y})$. Si se aplica una iteración del algoritmo Kernel Perceptron sobre X, tras la primera iteración los pesos resultantes son $\alpha = (1, 1, 0)$, ¿cuál es la función discriminante correspondiente a estos pesos?

A) $g(\mathbf{x}) = K(\mathbf{x}_1, \mathbf{x}) + K(\mathbf{x}_2, \mathbf{x}) + 2$
B) $g(\mathbf{x}) = K(\mathbf{x}_1, \mathbf{x}) + K(\mathbf{x}_2, \mathbf{x}) - 2$
C) $g(\mathbf{x}) = -K(\mathbf{x}_1, \mathbf{x}) - K(\mathbf{x}_2, \mathbf{x}) + 2$
D) $g(\mathbf{x}) = -K(\mathbf{x}_1, \mathbf{x}) - K(\mathbf{x}_2, \mathbf{x}) - 2$

C Dado el siguiente prototipo multinomial $\hat{\mathbf{p}} = (0.4\ 0.2\ 0.4\ 0.0)^t$, ¿cuál sería su versión suavizada $\tilde{\mathbf{p}}$ mediante Laplace con $\epsilon = 0.2$?

A) $\tilde{\mathbf{p}} = \left(\frac{5}{10}\ \frac{2}{10}\ \frac{5}{10}\ \frac{1}{10}\right)^t$
B) $\tilde{\mathbf{p}} = \left(\frac{3}{10}\ \frac{1}{10}\ \frac{3}{10}\ \frac{3}{10}\right)^t$
C) $\tilde{\mathbf{p}} = \left(\frac{3}{9}\ \frac{2}{9}\ \frac{3}{9}\ \frac{1}{9}\right)^t$
D) $\tilde{\mathbf{p}} = \left(\frac{2}{9}\ \frac{1}{9}\ \frac{2}{9}\ \frac{4}{9}\right)^t$

A En general, ¿cuál es el principal objetivo del suavizado de la matriz de covarianzas en un clasificador gaussiano?

A) Corregir su estimación por no disponer de una muestra suficientemente representativa de la población
B) Corregir la aparición de valores excesivamente elevados, y por tanto evitar probabilidades cero
C) Evitar valores extremos, tanto cero como uno, en la matriz de covarianzas, y por tanto evitar probabilidades cero
D) Evitar valores cero en la matriz de covarianza, y por tanto evitar probabilidades cero

D Dado el siguiente conjunto de datos en Σ^3, con $\Sigma = \{a, b, c\}$:

x_n	aba	aca	aaa	abb	bba	bcb	cbc	cca	cac
c_n	A	A	A	B	B	B	C	C	C

Indicar la clasificación por vecino más cercano del objeto representado por *cbb* emplando la distancia de Hamming ($d(s,t) = \sum_{i:s_i \neq t_i} 1$)

A) Clase A
B) Clase B
C) Clase C
D) Hay un empate entre las clases B y C, se elegiría una al azar

C Al aplicar edición de prototipos, se espera que:

A) El número de prototipos se reduzca de forma drástica
B) Se den fronteras de decisión lineales
C) Queden regiones de decisión simplemente conexas
D) Desaparezcan todos los puntos más cercanos a las fronteras de decisión iniciales

A Las fuentes de error de un clasificador son *Bias*, *Variance* y *Noise*. En general, ¿qué podrías afirmar de los clasificadores que has estudiado en la asignatura?

A) Aquellos que tienen un bajo *Bias*, tienen un alto *Variance*
B) Aquellos que tienen un bajo *Bias*, tienen un bajo *Variance*
C) Todos ellos tienen un bajo *Bias*, pero diferentes *Variance*
D) Todos ellos tienen un bajo *Variance*, pero diferentes *Bias*

$\boxed{\text{B}}$ Sobre el algoritmo AdaBoost, ¿cuál de las siguientes afirmaciones es correcta?

A) Existe un peso diferente asociado a cada dimensión de los datos
B) El número de clasificadores débiles disponibles es arbitrario
C) El error máximo de un clasificador débil seleccionado puede superar el 50 %
D) No es posible seleccionar el mismo clasificador débil dos veces

Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2018

Apellidos: ______ **Nombre:** ______

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

Problemas (4 puntos, 90 minutos, con apuntes)

1. (**1 punto**) Sea un conjunto de muestras de entrenamiento $X = \{(\mathbf{x}_1, +1), (\mathbf{x}_2, -1), (\mathbf{x}_3, +1), (\mathbf{x}_4, -1)\}$, del que se ha obtenido la siguiente matriz Gramm:

$$\mathbf{K} = \begin{pmatrix} 1 & 15 & 10 & 6 \\ 15 & 1 & 6 & 10 \\ 10 & 6 & 1 & 15 \\ 6 & 10 & 15 & 1 \end{pmatrix}$$

Se pide aplicar una iteración del algoritmo Kernel Perceptron sobre X tomando como pesos iniciales $\alpha = (0, 0, 0, 0)$.

Solución:

Tomando que $g(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i c_i K(x_i, x) + \alpha_i c_i$, se realizan los cálculos:

- $\mathbf{x}_1$: $g(\mathbf{x}_1) = 0 \leq 0$, $c_1 g(\mathbf{x}_1) = 0 \rightarrow \alpha = (1, 0, 0, 0)$
- $\mathbf{x}_2$: $g(\mathbf{x}_2) = \alpha_1 c_1 K(\mathbf{x}_1, \mathbf{x}_2) + \alpha_1 c_1 = 15 + 1 = 16$, $c_2 g(\mathbf{x}_2) = -16 \leq 0 \rightarrow \alpha = (1, 1, 0, 0)$
- $\mathbf{x}_3$: $g(\mathbf{x}_3) = \alpha_1 c_1 K(\mathbf{x}_1, \mathbf{x}_3) + \alpha_1 c_1 + \alpha_2 c_2 K(\mathbf{x}_2, \mathbf{x}_3) + \alpha_2 c_2 = 10 + 1 - 6 - 1 = 4$, $c_3 g(\mathbf{x}_3) = 4 > 0 \rightarrow \alpha = (1, 1, 0, 0)$
- $\mathbf{x}_4$: $g(\mathbf{x}_4) = \alpha_1 c_1 K(\mathbf{x}_1, \mathbf{x}_4) + \alpha_1 c_1 + \alpha_2 c_2 K(\mathbf{x}_2, \mathbf{x}_4) + \alpha_2 c_2 = 6 + 1 - 10 - 1 = -4$, $c_4 g(\mathbf{x}_4) = 4 > 0 \rightarrow \alpha = (1, 1, 0, 0)$

2. (**1 punto**) Se tienen imágenes binarias de 3×3 píxeles pertenecientes a dos clases, cuyas muestras de entrenamiento se representan por las siguientes imágenes, donde los píxeles blancos se codifican como 1 y los negros como 0:

Clase A    

Clase B    

Se quiere emplear un clasificador de Bernoulli entrenado a partir de dichas imágenes. Así pues, se pide:

a) Calcular los parámetros (probabilidades *a priori* y prototipos Bernoulli $\mathbf{p}_c$) del clasificador de Bernoulli por estimación de máxima verosimilitud sobre la representación directa de las imágenes dadas (**0.5 puntos**).

b) Clasificar la imagen  (**0.25 puntos**).

c) ¿Sería necesario aplicar algún tipo de suavizado? Razona la respuesta (**0.25 puntos**).

Solución:

a) Las codificaciones de las imágenes de entrenamiento serían:

Clase A	0 0 0 1 1 1 0 0 0	0 1 0 1 1 1 0 0 1	1 0 0 1 1 1 0 0 1	0 0 0 0 1 1 0 1 0
Clase B	1 1 1 0 0 0 1 1 1	1 1 1 0 1 0 1 1 1	1 0 1 0 0 1 1 1 1	0 1 1 0 0 0 1 1 0

Al haber igualdad de muestras de entrenamiento de cada clase, tendremos $P(A) = P(B) = \frac{1}{2}$, y los prototipos Bernoulli de cada clase serán:

$$\mathbf{p}_A = \left(\frac{1}{4}, \frac{1}{4}, 0, \frac{3}{4}, 1, 1, 0, \frac{1}{4}, \frac{1}{2}\right)$$

$$\mathbf{p}_B = \left(\frac{3}{4}, \frac{3}{4}, 1, 0, \frac{1}{4}, \frac{1}{4}, 1, 1, \frac{3}{4}\right)$$

b) Al haber componentes de los prototipos Bernoulli con valor 0, ambos clasificadores darían un valor 0, con lo que no es posible clasificar la muestra.

c) Sí, es necesario emplear suavizado para evitar situaciones como las que están ocurriendo en la clasificación de la muestra de test dada.

3. (**1 punto**) Se tiene el siguiente conjunto de prototipos en $\mathbb{R}^2$ de dos clases:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Prototipo	(1,1)	(2,1)	(2,2)	(3,2)	(3,3)	(4,3)	(3,4)	(5,3)
Clase	A	B	A	B	A	B	A	B

Se pide:

a) Clasificar la muestra (3,1) por vecino más cercano usando distancia euclídea (**0.25 puntos**).

b) Aplicar el algoritmo de edición de Wilson sobre el conjunto de prototipos por orden de índice creciente empleando vecino más cercano. En caso de empate por distancias, se clasifica en la clase correcta (**0.5 puntos**).

c) Clasificar la misma muestra de test del apartado a) usando el conjunto de prototipos editados (**0.25 puntos**).

Solución:

a) La distancia euclídea (al cuadrado) de la muestra (3,1) a cada prototipo es:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Prototipo	(1,1)	(2,1)	(2,2)	(3,2)	(3,3)	(4,3)	(3,4)	(5,3)
d^2	4	1	2	1	4	5	9	8

Por tanto, se clasifica en la clase B (que es la del prototipo (2,1)).

b) Llamando a cada prototipo x_i, siendo i el índice en la tabla de prototipos, la matriz de distancia euclídea (al cuadrado) entre los prototipos es:

d^2	x_1						
x_2	1	x_2					
x_3	2	1	x_3				
x_4	5	2	1	x_4			
x_5	8	5	2	1	x_5		
x_6	13	5	5	2	1	x_6	
x_7	13	10	5	4	1	2	x_7
x_8	20	13	10	5	4	1	5

Aplicamos edición de Wilson con $k = 1$:

- $x_1 \to x_2$: Error, $X = \{x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$
- $x_2 \to x_3$: Error, $X = \{x_3, x_4, x_5, x_6, x_7, x_8\}$
- $x_3 \to x_4$: Error, $X = \{x_4, x_5, x_6, x_7, x_8\}$
- $x_4 \to x_5$: Error, $X = \{x_5, x_6, x_7, x_8\}$
- $x_5 \to x_6/x_7$: Acierto, $X = \{x_5, x_6, x_7, x_8\}$
- $x_6 \to x_5/x_8$: Acierto, $X = \{x_5, x_6, x_7, x_8\}$
- $x_7 \to x_5$: Acierto, $X = \{x_5, x_6, x_7, x_8\}$
- $x_8 \to x_6$: Acierto, $X = \{x_5, x_6, x_7, x_8\}$

Por tanto, el conjunto de prototipos final es $X = \{(3,3), (4,3), (3,4), (5,3)\}$

c) Al aplicar vecino más cercano se clasifica en la clase A (la del prototipo (3,3))

4. (**1 punto**) Sean las siguientes muestras y clasificadores:

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

$$g_0(\mathbf{z}) = \begin{cases} +1 & z_1 \geq 0.5 \\ -1 & z_1 < 0.5 \end{cases} \quad g_1(\mathbf{z}) = \begin{cases} -1 & z_1 \geq 0.5 \\ +1 & z_1 < 0.5 \end{cases} \quad g_2(\mathbf{z}) = \begin{cases} +1 & z_2 \geq 0.5 \\ -1 & z_2 < 0.5 \end{cases} \quad g_3(\mathbf{z}) = \begin{cases} -1 & z_2 \geq 0.5 \\ +1 & z_2 < 0.5 \end{cases}$$

Tras aplicar una primera iteración de AdaBoost se elige $C_1 = g_3$, con $\alpha_1 = \frac{1}{2}\ln 3$, y los pesos se actualizan a $w^{(2)} = \left(\frac{1}{6}, \frac{3}{6}, \frac{1}{6}, \frac{1}{6}\right)$. Se pide aplicar una segunda iteración de AdaBoost para ese conjunto de datos y clasificadores indicando:

a) Clasificador escogido C_2.

b) Valor de ϵ_2.

c) Valor de α_2.

d) Actualización de los pesos para la siguiente iteración ($w^{(3)}$).

Solución:

	$\epsilon^{(2)}$
	$\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3\ \mathbf{x}_4$
g_0	$0 + 0 + \frac{1}{6} + 0 = \frac{1}{6}$
g_1	$\frac{1}{6} + \frac{3}{6} + 0 + \frac{1}{6} = \frac{5}{6}$
g_2	$0 + \frac{3}{6} + 0 + 0 = \frac{3}{6}$
g_3	$\frac{1}{6} + 0 + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}$

$$m = 2 \to C_2 = g_0 \quad \epsilon_2 = \frac{1}{6} \quad \alpha_2 = \frac{1}{2}\ \ln\left(\frac{1-\frac{1}{6}}{\frac{1}{6}}\right) = \frac{1}{2}\ \ln 5$$

$$G(\mathbf{x}) = \alpha_1\, C_1(\mathbf{x}) + \alpha_2\, C_2(\mathbf{x}) = \frac{1}{2}\ln 3 \cdot g_3(\mathbf{x}) + \frac{1}{2}\ \ln 5 \cdot g_0(\mathbf{x}) \ \to \ \text{Error} = 0.25$$

	$w^{(2)} \cdot \exp(-y_i \alpha_2 C_2(x_i))$	$w^{(3)}$
$\mathbf{x}_1$	$\frac{1}{6} \cdot e^{-\frac{1}{2}\ \ln 5} = \frac{1}{6\sqrt{5}}$	$\frac{1}{10}$
$\mathbf{x}_2$	$\frac{3}{6} \cdot e^{-\frac{1}{2}\ \ln 5} = \frac{3}{6\sqrt{5}}$	$\frac{3}{10}$
$\mathbf{x}_3$	$\frac{1}{6} \cdot e^{\frac{1}{2}\ \ln 5} = \frac{5}{6\sqrt{5}}$	$\frac{5}{10}$
$\mathbf{x}_4$	$\frac{1}{6} \cdot e^{-\frac{1}{2}\ \ln 5} = \frac{1}{6\sqrt{5}}$	$\frac{1}{10}$
Suma	$\frac{10}{6\sqrt{5}}$	