Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2018

Apellidos: ☐ **Nombre:** ☐

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

Cuestiones (2 puntos, 30 minutos, sin apuntes)

C Ante un conjunto de muestras linealmente separable, ¿qué tipo de kernel es preferible aplicar?

A) Un kernel polinomial
B) Un kernel gaussiano
C) No es necesario aplicar kernels
D) Cualquier kernel

D Dada una función kernel $K(\mathbf{x}, \mathbf{y})$, ¿cuál de las siguientes **no** es una función kernel?

A) $\mathbf{x}^2 K(\mathbf{x}, \mathbf{y})\mathbf{y}^{-1}$
B) $3K(\mathbf{x}, \mathbf{y})^2 + K(\mathbf{x}, \mathbf{y})$
C) $(5 + K(\mathbf{x}, \mathbf{y}))^3$
D) $-K(\mathbf{x}, \mathbf{y})$

B Sean A y B dos clases equiprobables de objetos en $\mathbf{x} = (x_1, x_2)^t \in \{0, 1\}^2$, con f.d. condicional Bernoulli con parámetros:

P(A) = P(B)

$$p_A = \left(\frac{1}{4}, \frac{1}{2}\right) \qquad p_B = \left(\frac{1}{2}, \frac{1}{8}\right)$$

¿Cuál de los siguientes clasificadores realiza la misma asignación de clases que el clasificador Bernoulli dado para cada uno de los objetos binarios en $\mathbb{R}^2$?

A) $g_A(x) = x_1,\ g_B(x) = x_1 + x_2$
B) $g_A(x) = x_2,\ g_B(x) = 1 - x_2$
C) $g_A(x) = x_1,\ g_B(x) = x_2$
D) $g_A(x) = x_1,\ g_B(x) = 1 - x_1$

D El número de parámetros de un clasificador gaussiano general es:

A) Independiente del número de clases.
B) Independiente de la dimensión de los datos.
C) Lineal con la dimensión y cuadrático con el número de clases.
D) Cuadrático con la dimensión y lineal con el número de clases.

A Sea X_c el conjunto de prototipos de la clase c y $X^k(y)$ el conjunto de los k prototipos más cercanos a y. ¿Cuál de las siguientes reglas de clasificación representa la del clasificador por k-vecinos más cercanos?

A) $\hat{c}(y) = \underset{c \in X_c}{\arg\max}\, |X^k(y) \cap X_c|$
B) $\hat{c}(y) = \underset{c \in X_c}{\arg\max}\, |X^k(y) \cup X_c|$
C) $\hat{c}(y) = \underset{c \in X_c}{\arg\min}\, |X^k(y) \cap X_c|$
D) $\hat{c}(y) = \underset{c \in X_c}{\arg\min}\, |X^k(y) \cup X_c|$

B El aprendizaje de distancias adaptadas a un conjunto de prototipos emplea ponderaciones que se basan en:

A) El número de prototipos disponible.
B) Las varianzas de los prototipos.
C) Las medias de los prototipos.
D) El número de vecinos elegido.

C El algoritmo *bagging* consiste en:

A) Una combinación de clasificadores fuertes obtenida por ponderación no homogénea de las muestras del conjunto de entrenamiento.

B) Una combinación de clasificadores fuertes obtenida de un conjunto de entrenamiento no ruidoso.
C) Una combinación de clasificadores fuertes obtenida por particionado del conjunto de entrenamiento.
D) Una combinación de clasificadores débiles obtenida por particionado del conjunto de entrenamiento.

$\boxed{\text{A}}$ Al aplicar *AdaBoost*:

A) Se escogen tanto los clasificadores como el peso de clasificadores y muestras de entrenamiento.
B) Se combinan clasificadores fuertes.
C) Se consigue una reducción de *variance*
D) Se finaliza cuando el error del clasificador escogido en una iteración es lo bastante bajo.

Examen de Teoría de Percepción - 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:

$$K = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{5} \\ \frac{1}{2} & 1 & \frac{1}{5} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{5} & 1 & \frac{1}{2} \\ \frac{1}{5} & \frac{1}{3} & \frac{1}{2} & 1 \end{pmatrix}$$

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:

- $\mathbf{x}_1$: $g(\mathbf{x}_1) = 0 \leq 0$, $c_1 g(\mathbf{x}_1) = 0 \to \alpha = (1, 0, 0, 0)$
- $\mathbf{x}_2$: $g(\mathbf{x}_2) = \alpha_1 c_1 K(\mathbf{x}_1, \mathbf{x}_2) + \alpha_1 c_1 = \frac{1}{2} + 1 = \frac{3}{2}$, $c_2 g(\mathbf{x}_2) = \frac{3}{2} > 0 \to \alpha = (1, 0, 0, 0)$
- $\mathbf{x}_3$: $g(\mathbf{x}_3) = \alpha_1 c_1 K(\mathbf{x}_1, \mathbf{x}_3) + \alpha_1 c_1 = \frac{1}{3} + 1 = \frac{4}{3}$, $c_3 g(\mathbf{x}_3) = -\frac{4}{3} \leq 0 \to \alpha = (1, 0, 1, 0)$
- $\mathbf{x}_4$: $g(\mathbf{x}_4) = \alpha_1 c_1 K(\mathbf{x}_1, \mathbf{x}_4) + \alpha_3 c_3 K(\mathbf{x}_3, \mathbf{x}_4) + \alpha_1 c_1 + \alpha_3 c_3 = \frac{1}{5} - \frac{1}{2} + 1 - 1 = -\frac{3}{10}$, $c_4 g(\mathbf{x}_4) = \frac{3}{10} > 0 \to \alpha = (1, 0, 1, 0)$

2. (**1 punto**) Se dispone de secuencias de ADN provenientes de dos clases, cuyas 5 muestras de entrenamiento por clase se representan por sus secuencias de bases:

Clase A	A C C G G T	A A C G G T	A A C G T T	A A C C G T	A C G G T T
Clase B	T T G C A A	T T G G C A	T G G C C A	T G G C A A	T G C C A A

Se quiere emplear un clasificador de multinomial entrenado a partir de su representación *bag-of-words*. Así pues:

a) Calcular los parámetros (probabilidades *a priori* y prototipos multinomiales $\mathbf{p}_c$) del clasificador multinomial por estimación de máxima verosimilitud sobre los datos dados. (**0.4 puntos**)

b) Clasificar la secuencia T T G C C A (**0.2 puntos**)

c) ¿Sería necesario aplicar algún tipo de suavizado? Razona la respuesta. (**0.2 puntos**)

d) ¿Cómo podrías reducir la tasa de error del clasificador? (**0.2 puntos**)

Solución:

a) La representación *bag-of-words* del conjunto de entrenamiento sería:

	A C G T	A C G T	A C G T	A C G T	A C G T
Clase A	1 2 2 1	2 1 2 1	2 1 1 2	2 2 1 1	1 1 2 2
Clase B	2 1 1 2	2 2 1 1	1 2 2 1	2 1 2 1	1 1 2 2

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

$$\mathbf{p}_A = \mathbf{p}_B = \left(\frac{8}{30}, \frac{7}{30}, \frac{8}{30}, \frac{7}{30}\right)$$

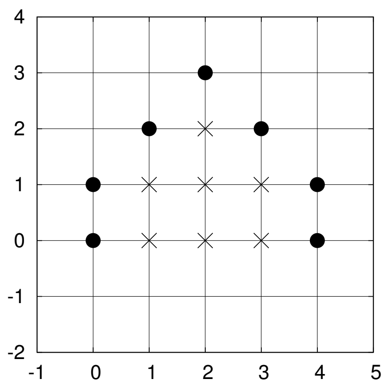
b) La clasificación de cualquier secuencia resulta siempre en empate entre las dos clases, porque ambas clases poseen las mismas probabilidades a priori y prototipos multinomiales.

c) No es necesario porque no hay probabilidades cero.

d) Como se puede apreciar en el conjunto de entrenamiento, las secuencias de bases en las clases A y B son diferentes, pero los vectores de contadores son los mismos. Por ello, para mejorar la tasa de error del clasificador será necesario que éste capture la información secuencial. Se podrían plantear dos aproximaciones. Por una parte utilizar una representación que capture esta información como por ejemplo la representación basada en n-gramas. Por otra parte, utilizar un modelo (clasificador) que aprenda la información secuencial, el más conocido es un modelo oculto de Markov.

3. **(1 punto)**

La figura de la derecha muestra prototipos bidimensionales de 2 clases, $X = \{x_1 = (0,0,\bullet),\ x_2 = (1,0,\times),\ x_3 = (1,2,\bullet),\ x_4 = (1,1,\times),\ x_5 = (0,1,\bullet),\ x_6 = (2,2,\times),\ x_7 = (2,3,\bullet),\ x_8 = (3,2,\bullet),\ x_9 = (3,0,\times),\ x_{10} = (4,0,\bullet),\ x_{11} = (3,1,\times),\ x_{12} = (4,1,\bullet),\ x_{13} = (2,0,\times),\ x_{14} = (2,1,\times)\}$. Considera la utilización de un clasificador de vecino más cercano en distancia L_1. Se pide:



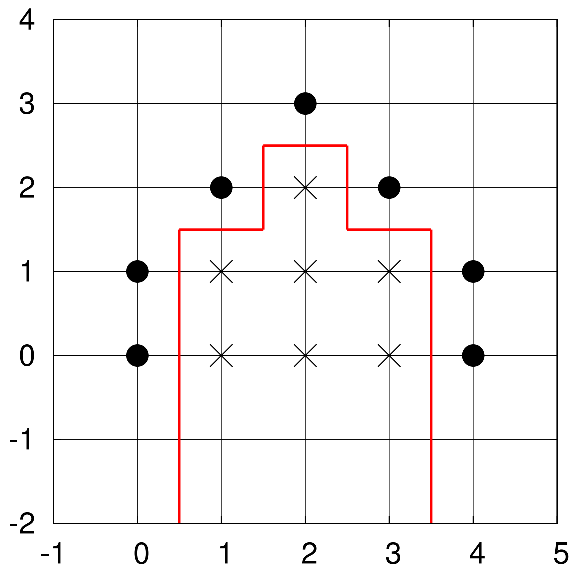
a) Representa gráficamente la frontera y regiones de decisión. (**0.25 puntos**)

b) Aplica el algoritmo de condensado de Hart visitando los prototipos por valor de índice creciente. En caso de empate, clasifica en la clase incorrecta. (**0.5 puntos**)

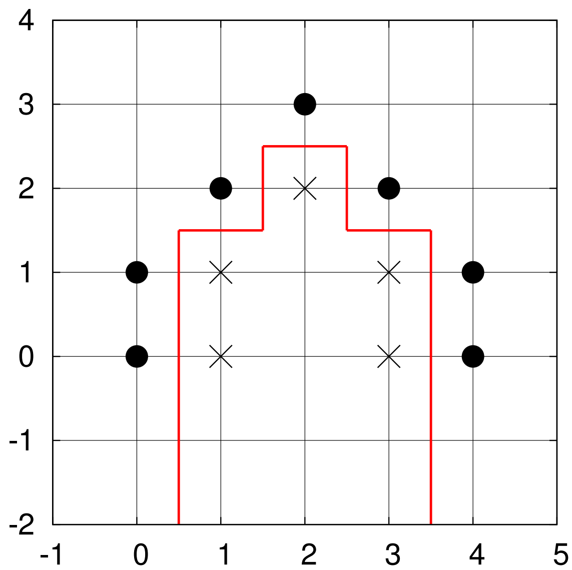
c) ¿Es posible eliminar más prototipos del conjunto S resultante del apartado anterior sin alterar la frontera y regiones de decisión iniciales? Si es así, indica cuáles. (**0.25 puntos**)

Solución:

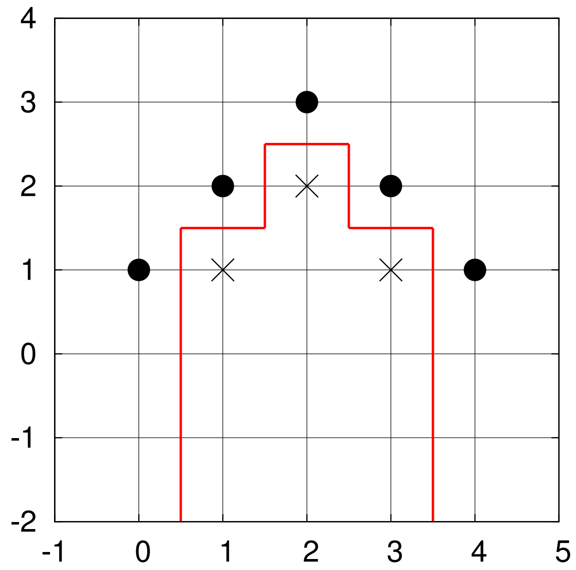
a)



b) $S = X - \{x_{13}, x_{14}\}$



c) Sí, es posible eliminar los prototipos x_1, x_2, x_9 y x_{10}.



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

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

   $$g_1(\mathbf{z}) = \begin{cases} +1 & z_1 \geq 0 \\ -1 & z_1 < 0 \end{cases} \quad g_2(\mathbf{z}) = \begin{cases} +1 & z_2 \geq 0 \\ -1 & z_2 < 0 \end{cases} \quad g_3(\mathbf{z}) = \begin{cases} +1 & z_1 + z_2 > 2 \\ -1 & z_1 + z_2 \leq 2 \end{cases} \quad g_4(\mathbf{z}) = \begin{cases} +1 & z_2 - z_1 \geq 0 \\ -1 & z_2 - z_1 < 0 \end{cases}$$

   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:
   Tabla de acierto/fallo:

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

Pesos: $w^{(2)} = \left(\frac{1}{6}, \frac{3}{6}, \frac{1}{6}, \frac{1}{6}\right)$
Error de clasificación ponderado por $w^{(2)}$:

g_1	g_2	g_3	g_4
$\frac{4}{6}$	$\frac{5}{6}$	$\frac{3}{6}$	$\frac{1}{6}$

$C_2 = g_4$
$\epsilon_2 = \frac{1}{6}$
$\alpha_2 = \frac{1}{2} \ln 5$

	$w^{(1)} \exp(-y_i \alpha_2 C_2(x_i))$
$\mathbf{x}_1$	$\frac{1}{6\sqrt{5}}$
$\mathbf{x}_2$	$\frac{3}{6\sqrt{5}}$
$\mathbf{x}_3$	$\frac{1}{6\sqrt{5}}$
$\mathbf{x}_4$	$\frac{\sqrt{5}}{6}$
Suma total	$\frac{\sqrt{5}}{3}$

$w^{(3)} = \left(\frac{1}{10}, \frac{3}{10}, \frac{1}{10}, \frac{5}{10}\right)$