Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politécnica de Valéncia, Junio de 2019

Apellidos: | | **Nombre:** | |

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

Cuestiones (2 puntos, 30 minutos, sin apuntes)

$\boxed{\text{C}}$ En el algoritmo Kernel Perceptron, cada componente del vector de salida α_i se interpreta como:

A) El peso que se le da sólo a la componente i-ésima
B) El número de veces que la muestra i-ésima se ha clasificado correctamente en el proceso
C) El número de veces que la muestra i-ésima se ha clasificado incorrectamente en el proceso
D) El peso que se le da sólo al término independiente i-ésimo

$\boxed{\text{B}}$ ¿Qué valor de ϵ en truncamiento simple se ha aplicado al parámetro Bernoulli $\hat{\mathbf{p}} = (0 \;\; 0 \;\; 1 \;\; 1)^t$ para obtener el parámetro suavizado $\tilde{\mathbf{p}} = (0.01 \;\; 0.01 \;\; 0.99 \;\; 0.99)^t$?

A) $\epsilon = 10^{-1}$
B) $\epsilon = 10^{-2}$
C) $\epsilon = 10^{-3}$
D) $\epsilon = 10^{-4}$

$\boxed{\text{D}}$ El clasificador multinomial que sigue $Mult_D(x_+, \mathbf{p}_c)$ se aplica sobre objetos que representan:

A) Valores naturales entre 1 y D
B) Proporciones de valores en una población de elementos en el rango 1-D
C) Secuencias de valores en el rango 1-D
D) Vectores de número de ocurrencias de valores en el rango 1-D en una secuencia

$\boxed{\text{B}}$ En la estimación por máxima verosimilitud de un clasificador gaussiano con matriz de covarianza común se estiman:

A) Probabilidades a priori y matrices de covarianza de cada clase
B) Probabilidades a priori y medias de cada clase, y la matriz de covarianzas del total de datos de entrenamiento
C) Probabilidades a priori de cada clase, y la media y matriz de covarianzas del total de datos de entrenamiento
D) Medias de cada clase y la matriz de covarianzas del total de datos de entrenamiento

$\boxed{\text{A}}$ Al aplicar suavizado por umbralizado de covarianza con $\epsilon > 0$, la matriz de covarianzas que se espera se caracterizará por:

A) Tener los mismos o más ceros fuera de la diagonal que la original
B) Tener los mismos o más ceros en la diagonal que la original
C) Ser diagonal
D) Tener menos ceros que la original

$\boxed{\text{X}}$ En general, la frontera de decisión entre dos clases de un clasificador basado en el vecino más cercano es una frontera lineal definida a trozos, pero que no es lineal globalmente. ¿En cuál de los siguientes casos esto es así? Pregunta cancelada por ser tanto B como C correctas

A) Cuando únicamente se dispone de una muestra de cada clase
B) Cuando las muestras de cada clase se disponen en rectas diferentes y éstas rectas son paralelas
C) Cuando las muestras de cada clase se disponen en rectas diferentes y éstas rectas intersectan
D) En todos los casos anteriores.

$\boxed{\text{D}}$ Cuando se hace aprendizaje de distancias, generalmente se hace:

A) Una selección de los prototipos que mejor contribuyen a la distancia
B) Una proyección sobre un espacio linealmente separable
C) Un cambio del exponente al que se eleva la diferencia entre componentes
D) Un escalado de las distintas componentes de los datos

$\boxed{\text{A}}$ En general, ¿cuál de las siguientes características es propia de un clasificador fuerte?

A) Tienen una precisión alta
B) Necesitan pocos datos de entrenamiento
C) Su sesgo (*bias*) es alto
D) Permiten cancelar el ruido (*noise*)

Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politécnica de Valéncia, Junio de 2019

Apellidos: ☐ **Nombre:** ☐

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

Problemas (4 puntos, 90 minutos, con apuntes)

1. (**0.75 puntos**) Sea la función $K(\mathbf{x}, \mathbf{y}) = \exp\left(-(\mathbf{x}-\mathbf{y})^t(\mathbf{x}-\mathbf{y})\right)$ y el conjunto de muestras

$$X = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$

   a) Calcula la matriz Gramm $\mathbf{K}$ asociada a las muestras de X con la función $K(\mathbf{x}, \mathbf{y})$ (**0.5 puntos**)

   b) Indica si la función $K(\mathbf{x}, \mathbf{y})$ es un Kernel, justificando la respuesta (**0.25 puntos**)

Solución:

   a)

$$\mathbf{K} = \begin{pmatrix} 1 & \exp(-1) & \exp(-4) & \exp(-1) \\ \exp(-1) & 1 & \exp(-5) & \exp(-2) \\ \exp(-4) & \exp(-5) & 1 & \exp(-1) \\ \exp(-1) & \exp(-2) & \exp(-1) & 1 \end{pmatrix}$$

   b) Sí es un Kernel, pues es un Kernel gaussiano con $\sigma^2 = \frac{1}{2}$

2. **(1.5 puntos)** Se dispone de secuencias de ADN provenientes de tres clases, cuyas muestras de entrenamiento por clase se representan por el vector de frecuencias de su bases:

Nº muestra	1	2	3	4	5	6	7	8	9
A	4	5	3	1	2	3	2	1	2
C	3	2	3	3	2	1	1	2	2
G	1	2	1	1	1	1	3	3	4
T	2	1	1	2	2	2	5	5	5
Clase	A	A	A	B	B	B	C	C	C

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 A T G G C T C A T" (**0.4 puntos**)

   c) Suaviza las probabilidades del prototipo multinomial de la clase A utilizando descuento absoluto ($\epsilon = \frac{1}{7}$) mediante interpolación, usando como distribución generalizada la distribución multinomial incondicional (independiente de la clase) estimada a partir de las muestras de entrenamiento (**0.5 puntos**)

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

Solución:

   a) Las probabilidades a priori son $P(A) = P(B) = P(C) = \frac{1}{3}$, y los prototipos multinomiales de cada clase:

$$\hat{\mathbf{p}}_A = \left(\frac{12}{28}, \frac{8}{28}, \frac{4}{28}, \frac{4}{28}\right) = \left(\frac{3}{7}, \frac{2}{7}, \frac{1}{7}, \frac{1}{7}\right)$$
$$\hat{\mathbf{p}}_B = \left(\frac{6}{21}, \frac{6}{21}, \frac{3}{21}, \frac{6}{21}\right) = \left(\frac{2}{7}, \frac{2}{7}, \frac{1}{7}, \frac{2}{7}\right)$$
$$\hat{\mathbf{p}}_C = \left(\frac{5}{35}, \frac{5}{35}, \frac{10}{35}, \frac{15}{35}\right) = \left(\frac{1}{7}, \frac{1}{7}, \frac{2}{7}, \frac{3}{7}\right)$$

   b) La representación *bag-of-words* de la secuencia a clasificar es $x = (2\ 2\ 2\ 4)$ y clasificación:

$$g_A(x) = \frac{1}{3} \cdot \left(\frac{3}{7}\right)^2 \cdot \left(\frac{2}{7}\right)^2 \cdot \left(\frac{1}{7}\right)^2 \cdot \left(\frac{1}{7}\right)^4 = \frac{36}{3 \cdot 7^{10}}$$

$$g_B(x) = \frac{1}{3} \cdot \left(\frac{2}{7}\right)^2 \cdot \left(\frac{2}{7}\right)^2 \cdot \left(\frac{1}{7}\right)^2 \cdot \left(\frac{2}{7}\right)^4 = \frac{256}{3 \cdot 7^{10}}$$
$$g_C(x) = \frac{1}{3} \cdot \left(\frac{1}{7}\right)^2 \cdot \left(\frac{1}{7}\right)^2 \cdot \left(\frac{2}{7}\right)^2 \cdot \left(\frac{3}{7}\right)^4 = \frac{324}{3 \cdot 7^{10}}$$
$$\hat{c}(x) = \arg\max_{A,B,C} \{g_A(x), g_B(x), g_C(x)\} = C$$

c) La distribución multinomial incondicional (independiente de la clase) estimada a partir de las muestras de entrenamiento se calcula como:

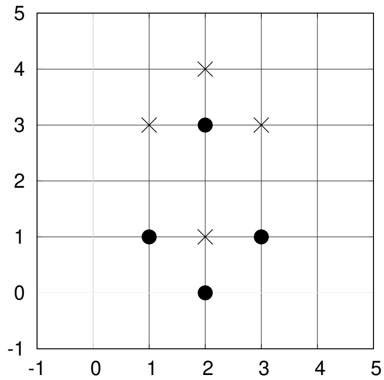
$$\mathbf{p}_g = \left(\frac{23}{84}, \frac{19}{84}, \frac{17}{84}, \frac{25}{84}\right)$$

Como podemos descontar ϵ en todos los casos, la masa de probabilidad descontada para la clase A y que repartimos según $\mathbf{p}_g$ será $\frac{4}{7}$. La estimación del prototipo multinomial de la clase A será:

$$\tilde{\mathbf{p}}_A = \hat{\mathbf{p}}_A - \begin{pmatrix} \frac{1}{7} \\ \frac{1}{7} \\ \frac{1}{7} \\ \frac{1}{7} \end{pmatrix} + \frac{4}{7} \cdot \mathbf{p}_g = \left(\frac{65}{147}, \frac{40}{147}, \frac{17}{147}, \frac{25}{147}\right)^t$$

d) *Respuesta negativa:* No es necesario porque no hay probabilidades cero.
Respuesta afirmativa: Sí, es recomendable porque la estimación de los prototipos multinomiales se realiza únicamente a partir de 3 muestras.

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



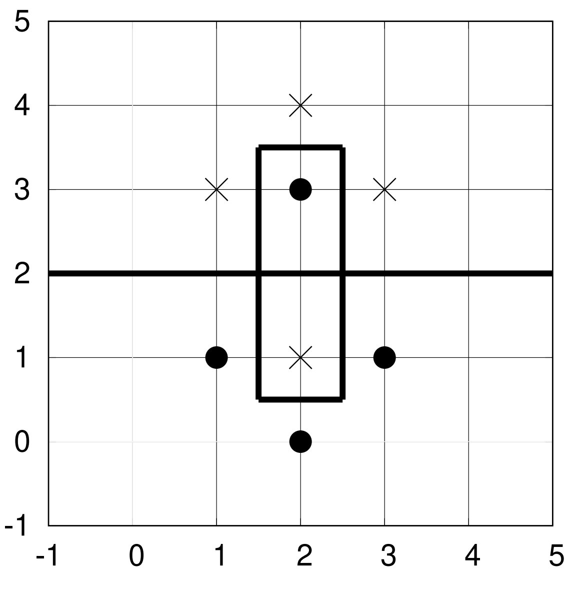
a) Representa gráficamente la frontera y regiones de decisión. (**0.2 puntos**)

b) Aplica el algoritmo de edición de Wilson visitando los prototipos por valor de índice creciente. Representa gráficamente la frontera y regiones de decisión. En caso de empate, clasifica en la clase correcta. (**0.4 puntos**)

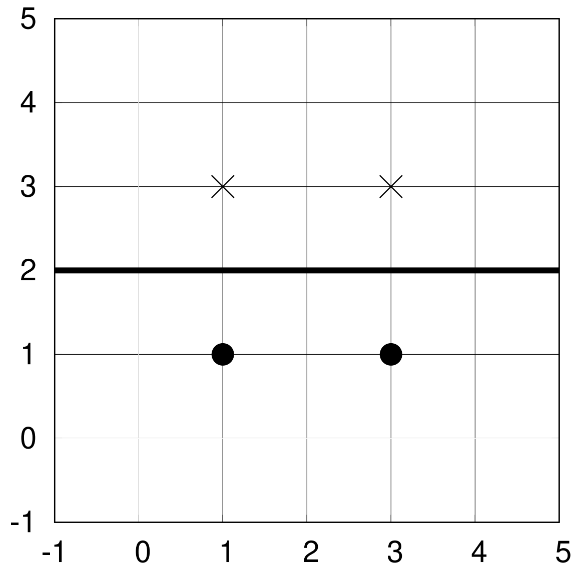
c) A partir de la aplicación del algoritmo de Wilson, aplica el algoritmo de condensado de Hart visitando los prototipos por valor de índice creciente. Representa gráficamente la frontera y regiones de decisión. En caso de empate, clasifica en la clase correcta. (**0.4 puntos**)

Solución:

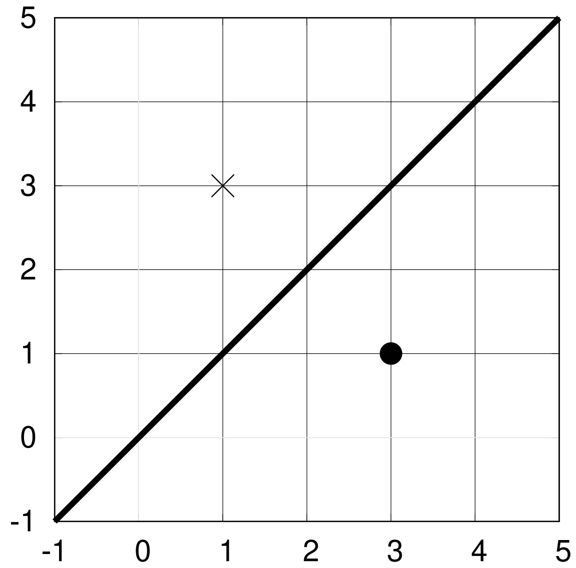
a)



b) $X' = X - \{x_1, x_2, x_3, x_4\}$



c) $S = \{x_5, x_7\}$.



4. **(0.75 puntos)** Se tiene el siguiente conjunto de datos y clasificadores:

$$X = \{((1,1),+1),((-1,1),-1),((-1,-1),+1),((1,-1),-1)\}$$

$$g_1(\mathbf{z}) = \begin{cases} +1 & z_1 + z_2 > -\frac{1}{2} \\ -1 & z_1 + z_2 \leq -\frac{1}{2} \end{cases} \qquad g_2(\mathbf{z}) = \begin{cases} +1 & z_1 + z_2 > \frac{1}{2} \\ -1 & z_1 + z_2 \leq \frac{1}{2} \end{cases}$$

a) Aplica dos iteraciones de AdaBoost a los datos de X con los clasificadores dados, incluyendo el clasificador escogido, los valores de error relativo (ϵ) y peso del clasificador (α) en cada iteración, así como el conjunto de pesos de las muestras usado en esa iteración (**0.6 puntos**)

b) ¿Tendría sentido aplicar una tercera iteración? Razona la respuesta (**0.15 puntos**)

Solución:

a) Tabla de acierto/fallo:

	x_1	x_2	x_3	x_4
g_1	✓	X	X	X
g_2	✓	✓	X	✓

Iteración 1:

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

Error de clasificación ponderado: $g_1 \rightarrow \frac{3}{4}$, $g_2 \rightarrow \frac{1}{4}$

Clasificador escogido: $C_1 = g_2$

Error relativo y peso del clasificador: $\epsilon_1 = \frac{1}{4}$, $\alpha_1 = \frac{1}{2}\ln 3$

Iteración 2:

Recálculo de pesos:

	$w^{(1)} \exp(-c_i \alpha_1 C_1(x_i))$	$w^{(2)}$
x_1	$\frac{1}{4}\frac{1}{\sqrt{3}}$	$\frac{1}{6}$
x_2	$\frac{1}{4}\frac{1}{\sqrt{3}}$	$\frac{1}{6}$
x_3	$\frac{1}{4}\sqrt{3}$	$\frac{3}{6}$
x_4	$\frac{1}{4}\frac{1}{\sqrt{3}}$	$\frac{1}{6}$
Suma	$\frac{2\sqrt{3}}{4}$	

Pesos: $w^{(2)} = \left\{\frac{1}{6}, \frac{1}{6}, \frac{3}{6}, \frac{1}{6}\right\}$

Error de clasificación ponderado: $g_1 \to \frac{5}{6}$, $g_2 \to \frac{3}{6}$

Clasificador escogido: $C_2 = g_2$

Error relativo y peso del clasificador: $\epsilon_2 = \frac{3}{6}$, $\alpha_2 = 0$

b) No, pues al ser $\alpha_2 = 0$ el conjunto de pesos no cambia, con lo que no va a haber cambios en posteriores iteraciones