

Examen del bloque 2 de SIN (tipo B)
ETSINF, Universitat Politècnica de València, 14 de enero de 2020

Apellidos:

Nombre:

Grupo: ☐ 3A ☐ 3B ☐ 3C ☐ 3D ☐ 3E ☐ 3F ☐ 3G ☐ 4IA

Test (1,75 puntos)

Marca cada recuadro con una única opción. Puntuación: $\max(0, (\text{aciertos} - \text{errores} / 3) \cdot 1,75 / 9)$.

- 1 ☐ D Sea \mathbf{x} un objeto a clasificar en una clase de C posibles. Indica cuál de los siguientes clasificadores *no* es de error mínimo (o escoge la última opción si los tres son de error mínimo):

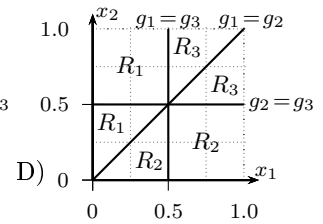
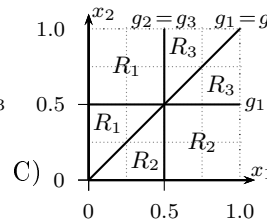
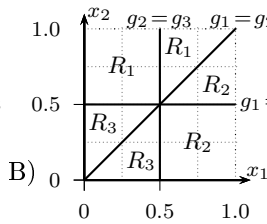
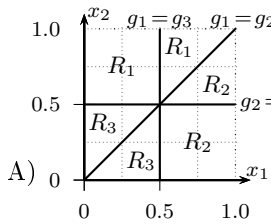
A) $c(\mathbf{x}) = \arg \max_{c=1, \dots, C} \log p(\mathbf{x}, c)$.

B) $c(\mathbf{x}) = \arg \max_{c=1, \dots, C} p(c | \mathbf{x})^2$.

C) $c(\mathbf{x}) = \arg \max_{c=1, \dots, C} \sqrt{p(\mathbf{x}, c)} / p(\mathbf{x})$.

D) Los tres clasificadores anteriores son de error mínimo.

- 2 ☐ B Sea un clasificador en tres clases basado en las funciones discriminantes lineales bidimensionales de vectores de pesos: $\mathbf{w}_1 = (0, 0, 1)^t$, $\mathbf{w}_2 = (0, 1, 0)^t$ y $\mathbf{w}_3 = (0.5, 0, 0)^t$. Indica cuál de las figuras dadas a continuación es coherente con las fronteras y regiones de decisión que define dicho clasificador.



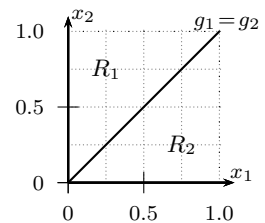
- 3 ☐ C Dado el clasificador en dos clases definido por su frontera y regiones de decisión de la figura de la derecha, ¿cuál de los siguientes vectores de pesos *no* define un clasificador equivalente al dado?

A) $\mathbf{w}_1 = (0, 0, 1)^t$ y $\mathbf{w}_2 = (0, 1, 0)^t$.

B) $\mathbf{w}_1 = (0, -1, 0)^t$ y $\mathbf{w}_2 = (0, 0, -1)^t$.

C) $\mathbf{w}_1 = (0, 1, 0)^t$ y $\mathbf{w}_2 = (0, 0, 1)^t$.

D) Todos los vectores de pesos anteriores definen clasificadores equivalentes.



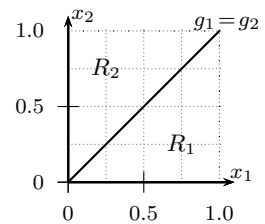
- 4 ☐ C Durante la aplicación del algoritmo Perceptrón ($\alpha = 1.0$ y $b = 0$) en un problema de clasificación en dos clases, se han obtenido los vectores de pesos $\mathbf{w}_1 = (-1, 1, 0)^t$ y $\mathbf{w}_2 = (1, 0, 1)^t$. Supón que el siguiente paso en la aplicación de Perceptrón consiste en procesar una cierta muestra de entrenamiento \mathbf{x} de clase c . Indica cuál de las siguientes opciones daría como resultado un conjunto de pesos que define la frontera y regiones de decisión de la figura de la derecha.

A) $\mathbf{x} = (-1, 1)^t$ y $c = 2$.

B) $\mathbf{x} = (0, 0)^t$ y $c = 2$.

C) $\mathbf{x} = (0, 0)^t$ y $c = 1$.

D) $\mathbf{x} = (-1, 1)^t$ y $c = 1$.



- 5 ☐ C Sea un problema de clasificación en tres clases para objetos del tipo $\mathbf{x} = (x_1, x_2)^t \in \{0, 1\}^2$, con las distribuciones de probabilidad de la derecha. ¿Cuál es el error de Bayes, ϵ^* , en este problema?

A) $\epsilon^* < 0.2$.

B) $0.2 \leq \epsilon^* < 0.4$.

C) $0.4 \leq \epsilon^* < 0.7$. $.2 \cdot .4 + .3 \cdot .2 + .2 \cdot .5 + .3 \cdot 2/3 = .44$

D) $0.7 \leq \epsilon^*$.

\mathbf{x}		$P(c \mathbf{x})$			$P(\mathbf{x})$
x_1	x_2	$c=1$	$c=2$	$c=3$	
0	0	0.6	0.2	0.2	0.2
0	1	0.1	0.1	0.8	0.3
1	0	0.3	0.5	0.2	0.2
1	1	1/3	1/3	1/3	0.3

- 6 **D** Se tiene un problema de clasificación para el cual se ha aprendido un clasificador. Asimismo, se tiene un conjunto de $M = 100$ muestras de test con el cual se ha estimado:

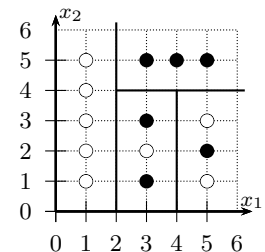
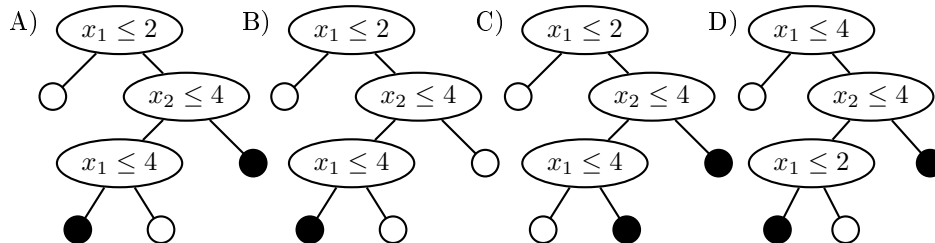
- La probabilidad de error del clasificador aprendido, $\hat{p} = 0.10 = 10\%$.
- Un intervalo de confianza al 95 % para dicha probabilidad de error, $\hat{I} = [0.04, 0.16] = [4\%, 16\%]$.

Se considera que la probabilidad de error estimada es razonable y que la misma no variará significativamente aunque usemos muchas más muestras de test. Ahora bien, el intervalo de confianza (al 95 %) estimado, $\hat{I} = 10\% \pm 6\%$, nos parece un poco amplio y nos preguntamos si es posible reducir su amplitud mediante el uso de más de $M = 100$ muestras de test. Además, si ello fuera posible, nos preguntamos si sería posible reducir dicha amplitud a la mitad o menos; esto es, tal que $\hat{I} = 10\% \pm \hat{R}$ con $\hat{R} \leq 3\%$. En relación con estas cuestiones, indica cuál de las siguientes afirmaciones es correcta.

- A) No es posible reducir la amplitud de \hat{I} ya que hemos considerado que \hat{p} no variará significativamente y, siendo así, la amplitud de \hat{I} tampoco puede variar significativamente.
- B) En general, no es posible reducir la amplitud de \hat{I} pues \hat{I} no depende significativamente de M .
- C) Sí es posible reducir la amplitud de \hat{I} , a la mitad o menos, si doblamos M al menos ($M \geq 200$).
- D) Sí es posible reducir la amplitud de \hat{I} , a la mitad o menos, si empleamos al menos cuatro veces más muestras de test aproximadamente ($M \geq 400$).

$$1.96 \cdot \sqrt{(0.1 \cdot 0.9)/M} \leq 0.03 \rightarrow M \geq 385$$

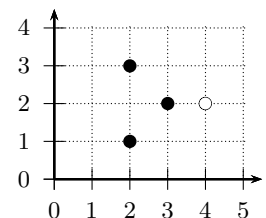
- 7 **A** Dado el conjunto de muestras de 2 clases (\circ y \bullet) de la figura de la derecha, ¿cuál de los siguientes árboles de clasificación es coherente con la partición representada?



- 8 **C** La figura a la derecha muestra una partición de 4 puntos bidimensionales en 2 clústers (representados mediante los símbolos \bullet y \circ). La transferencia del punto $(3, 2)^t$ del cluster \bullet al cluster \circ :

- A) produce un incremento en la Suma de Errores Cuadráticos (SEC).
- B) no altera la SEC.
- C) produce un decremento en la SEC.
- D) produce una SEC negativa.

$$\Delta J = 0.5 - 0.67335 = -0.17335$$



- 9 **C** En relación al cálculo de la probabilidad $P(y | M)$ con la que un modelo de Markov M genera una cadena de símbolos y , indica qué afirmación es cierta:

- A) La única forma de calcular $P(y | M)$ consiste en generar explícitamente todas las secuencias de estados, calcular la probabilidad de que cada secuencia de estados haya generado y y posteriormente sumar todas las probabilidades obtenidas.
- B) Una forma eficiente computacionalmente de calcular $P(y | M)$ consiste en aplicar el algoritmo de Viterbi.
- C) Una forma eficiente computacionalmente de calcular $P(y | M)$ consiste en aplicar el algoritmo *Forward*.
- D) La única forma de calcular $P(y | M)$ consiste en generar explícitamente todas las secuencias de estados mediante el algoritmo de Viterbi, calcular la probabilidad de que cada secuencia haya generado y y sumar todas las probabilidades obtenidas.

Problema (2 puntos)

Sea un modelo de Markov de conjunto de estados $Q = \{1, 2, F\}$ y conjunto de símbolos $\Sigma = \{a, b\}$. Se pide:

- a) (1 punto) Sean el vector de probabilidades iniciales (π), matriz de transición entre estados (A) y matriz de generación de símbolos (B):

π	1	2
	0.6	0.4

A	1	2	F
1	0.6	0.3	0.1
2	0.3	0.4	0.3

B	a	b
1	0.3	0.7
2	0.8	0.2

Realiza una traza del algoritmo de Viterbi para la cadena $y = aab$ obteniendo la mejor secuencia de estados.

- b) (1 punto) Sean las tres cadenas de símbolos: $y_1 = bbaa$, $y_2 = abab$ y $y_3 = aabbb$. Al aplicar el algoritmo de Viterbi con un cierto modelo de Markov M , se obtienen, respectivamente, las siguientes secuencias óptimas de estados: $1122F$, $2121F$ y $22111F$. A partir de dichas cadenas y sus respectivas secuencias óptimas de estados, re-estima las probabilidades iniciales (π), de transición (A) y de emisión (B) de M (del mismo modo que se hace en una iteración del algoritmo de re-estimación de Viterbi).

- a) Traza del algoritmo de Viterbi para la cadena $y = aab$:

V	a	a	b	
1	$0.6 \cdot 0.3 = 0.18$	$0.18 \cdot 0.6 \cdot 0.3 = 0.0324$ $0.32 \cdot 0.3 \cdot 0.3 = 0.0288$ $0.0324 > 0.0288$ (de 1)	$0.0324 \cdot 0.6 \cdot 0.7 = 0.0136$ $0.1024 \cdot 0.3 \cdot 0.7 = 0.0215$ $0.0136 < 0.0215$ (de 2)	
2	$0.4 \cdot 0.8 = 0.32$	$0.18 \cdot 0.3 \cdot 0.8 = 0.0432$ $0.32 \cdot 0.4 \cdot 0.8 = 0.1024$ $0.0432 < 0.1024$ (de 2)	$0.0324 \cdot 0.3 \cdot 0.2 = 0.0019$ $0.1024 \cdot 0.4 \cdot 0.2 = 0.0082$ $0.0019 < 0.0082$ (de 2)	
F	—	—	—	$0.0215 \cdot 0.1 = 0.0022$ $0.0082 \cdot 0.3 = 0.0025$ $0.0022 < 0.0025$ (de 2)

La secuencia óptima de estados es: $222F$

- b) La estimación de π , A y B para las cadenas de entrenamiento $y_1 = bbaa$, $y_2 = abab$ y $y_3 = aabbb$ es

π : El estado 1 se ha utilizado una vez como estado inicial y el estado 2 dos veces.

A : ■ La transición 1-1 3 veces, la 1-2 2 veces, la 1-F dos veces

■ La transición 2-1 3 veces, la 2-2 2 veces, la 2-F una vez

B : ■ El símbolo a se ha emitido 0 vez en el estado 1 y 6 veces en el estado 2

■ El símbolo b se ha emitido 7 veces del estado 1 y 0 veces del estado 2

Normalizando

π	1	2
	$\frac{1}{3}$	$\frac{2}{3}$

A	1	2	F
1	$\frac{3}{7}$	$\frac{2}{7}$	$\frac{2}{7}$
2	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

B	a	b
1	0.0	1.0
2	1.0	0.0