Examen del Bloque 2 de Sistemas Inteligentes

ETSINF, Universitat Politècnica de València, 18 de enero de 2017

Apellidos: ☐ **Nombre:** ☐

Grupo: ☐ **3A** ☐ **3B** ☐ **3C** ☐ **3D** ☐ **3E** ☐ **3F** ☐ **3FLIP**

Marca cada recuadro con una única opción de entre las dadas.

1 $\boxed{\text{C}}$ ¿Cuál de las siguientes expresiones es *incorrecta*?

A) $P(x \mid y) = \dfrac{P(x,y)}{\sum_z P(y \mid z)\ P(z)}$

B) $P(x \mid y) = \dfrac{P(x,y)}{\sum_z P(y,z)}$

C) $P(x \mid y) = \dfrac{\sum_z P(x,z)}{P(y)}$

D) $P(x \mid y) = \dfrac{P(y \mid x)\ P(x)}{P(y)}$

2 $\boxed{\text{B}}$ Se tienen dos bolsas. La primera contiene 3 manzanas de color rojo y 5 de color verde; la segunda, 2 rojas, 2 verdes y 1 amarilla. Se escoge una bolsa al azar y, seguidamente, una manzana al azar de la misma. Supóngase que las bolsas tienen la misma probabilidad de ser escogidas y que, dada una bolsa cualquiera, sus manzanas también tienen idéntica probabilidad de ser escogidas. Si la manzana escogida es roja, ¿cuál es la probabilidad P de que sea de la primera bolsa?

A) $0.00 \leq P < 0.25$

B) $0.25 \leq P < 0.50$

C) $0.50 \leq P < 0.75$

D) $0.75 \leq P$

$$P = P(B=1 \mid C=r) = \frac{P(B=1)P(C=r|B=1)}{P(C=r)}$$
$$= \frac{P(B=1)P(C=r|B=1)}{P(B=1)P(C=r|B=1)+P(B=2)P(C=r|B=2)}$$
$$= \frac{\frac{1}{2}\cdot\frac{3}{8}}{\frac{1}{2}\cdot\frac{3}{8}+\frac{1}{2}\cdot\frac{2}{5}} = \frac{15}{31} = 0.4839$$

3 $\boxed{\text{A}}$ Sea x un objeto (vector de características o cadena de símbolos) a clasificar en una clase de C posibles. Indica cuál de los siguientes clasificadores *no* es de error mínimo:

A) $c(x) = \arg\max\limits_{c=1,\dots,C}\ p(x \mid c)$

B) $c(x) = \arg\max\limits_{c=1,\dots,C}\ p(x,c)$

C) $c(x) = \arg\max\limits_{c=1,\dots,C}\ \log p(x,c)$

D) $c(x) = \arg\max\limits_{c=1,\dots,C}\ P(c \mid x)$

4 $\boxed{\text{A}}$ Para un problema de clasificación de dos clases en $\mathbb{R}^2$ tenemos un clasificador compuesto por dos funciones discriminantes lineales cuyos vectores de pesos en notación homogénea son $\mathbf{a}_\circ = (-1,1,2)^t$ y $\mathbf{a}_\bullet = (1,1,1)^t$. Indica cuáles son las regiones de decisión definidas por el clasificador anterior.

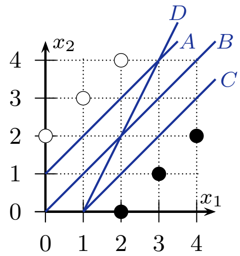
A) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_2 > 2\}$ y $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_2 < 2\}$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \to x_2 = 2 \ \wedge\ g_\circ((0,0)^t) < g_\bullet((0,0)^t)$

B) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_1 > 2\}$ y $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_1 < 2\}$

C) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_1 < 2\}$ y $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_1 > 2\}$

D) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_2 < 2\}$ y $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_2 > 2\}$

5 $\boxed{\text{B}}$ En la figura de la derecha se representan 6 muestras de aprendizaje bidimensionales de 2 clases. Tras aplicar el algoritmo Perceptrón con diferentes valores del parámetro b, se obtienen los siguientes clasificadores caracterizados por sus vectores de pesos, ¿cuál de ellos proporciona una frontera de decisión con mayor holgura ("centradas"), y por tanto de menor error esperado?



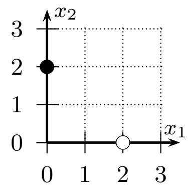
A) $\mathbf{a}_\circ = (-1, 1, 2)^t$ y $\mathbf{a}_\bullet = (\ 0, 2, 1)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \rightarrow x_2 = x_1 + 1$

B) $\mathbf{a}_\circ = (\ 1, 1, 2)^t$ y $\mathbf{a}_\bullet = (\ 1, 2, 1)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \rightarrow x_2 = x_1$

C) $\mathbf{a}_\circ = (\ 1, 1, 2)^t$ y $\mathbf{a}_\bullet = (\ 0, 2, 1)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \rightarrow x_2 = x_1 - 1$

D) $\mathbf{a}_\circ = (\ 1, 1, 1)^t$ y $\mathbf{a}_\bullet = (-1, 3, 0)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \rightarrow x_2 = 2x_1 - 2$

6 $\boxed{\text{C}}$ En la figura de la derecha se representan dos muestras de aprendizaje bidimensionales de 2 clases: ($\mathbf{x}_1$,$\circ$) y ($\mathbf{x}_2$,$\bullet$). Dados los pesos $\mathbf{a}_\circ = (0, 1, 0)^t$ y $\mathbf{a}_\bullet = (1, 0, 0)^t$, si aplicamos una iteración del algoritmo Perceptrón procesando únicamente la muestra $\mathbf{x}_1$, ¿cuál es el valor mínimo del margen b con el que se actualizan los vectores de pesos?



A) $b = 0.5$

B) $b = 1.0$

C) $b = 1.5$ $\quad g_\circ(\mathbf{x}_1) = 2 \quad g_\bullet(\mathbf{x}_1) = 1 \quad$ if $(g_\bullet(\mathbf{x}_1) + b > g_\circ(\mathbf{x}_1))$

D) Ninguno de los anteriores

7 $\boxed{\text{D}}$ Se tiene un problema de clasificación para el cual se ha aprendido un clasificador. El intervalo de confianza al 95 % para la probabilidad de error de dicho clasificador se ha estimado empíricamente, a partir de un cierto conjunto de muestras de test. Indica cuál de las siguientes opciones permitiría reducir el tamaño del intervalo estimado:

A) Reducir significativamente el conjunto de test.

B) Mantener el conjunto de test y re-entrenar el clasificador con el algoritmo C-medias de Duda y Hart.

C) Mantener el conjunto de test y re-entrenar el clasificador con el algoritmo C-medias convencional ("popular").

D) Aumentar significativamente el conjunto de test.

8 $\boxed{\text{A}}$ Sea un problema de clasificación en 3 clases (A, B y C) para el que se dispone de 6 datos de aprendizaje representados mediante vectores de características tridimensionales (ver tabla a la derecha). Si deseamos aplicar el algoritmo de aprendizaje de árboles de clasificación con dichos datos, ¿cuál es el número N de particiones diferentes que se pueden generar en el nodo raíz del árbol? Nota: no se tengan en cuenta las particiones que dan lugar a nodos vacíos.

n	1	2	3	4	5	6
x_{n1}	0	1	0	1	0	1
x_{n2}	1	1	2	2	3	3
x_{n3}	0	2	0	3	2	3
c_n	A	A	B	B	C	C

A) $0 \leq N \leq 5$ $\quad \{(1,0), (2,1), (2,2), (3,0), (3,2)\}$

B) $5 < N \leq 10$

C) $10 < N \leq 20$

D) Se pueden generar infinitas particiones.

9 $\boxed{\text{D}}$ Supóngase que estamos aplicando el algoritmo de aprendizaje de árboles de clasificación para un problema de cuatro clases; esto es, $\mathcal{C} = \{1, 2, 3, 4\}$. El algoritmo ha alcanzado un nodo t que incluye ocho datos: 4 de la clase 1, 2 de la 2, 1 de la 3 y 1 de la 4. La impureza de t, $\mathcal{I}(t)$, medida como la entropía de la distribución empírica de las probabilidades a posteriori de las clases en t, es:

A) $0.00 \leq \mathcal{I}(t) < 0.25$

B) $0.25 \leq \mathcal{I}(t) < 0.50$

C) $0.50 \leq \mathcal{I}(t) < 0.75$

D) $0.75 \leq \mathcal{I}(t)$ $\quad \mathcal{I}(t) = -\sum_{c=1}^{4} \hat{P}(c \mid t) \log_2 \hat{P}(c \mid t) = -\frac{4}{8}\log_2\frac{4}{8} - \frac{2}{8}\log_2\frac{2}{8} - 2\,\frac{1}{8}\log_2\frac{1}{8} = \frac{7}{4} = 1.75$

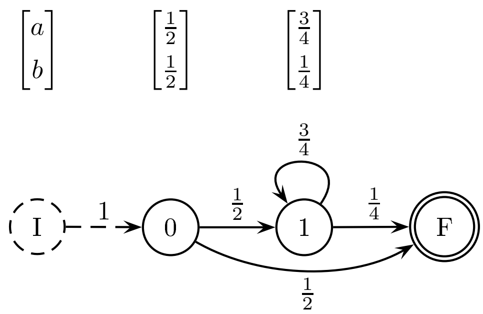
10 $\boxed{\text{B}}$ Indica cuál de las siguientes afirmaciones sobre aprendizaje supervisado (AS) y no-supervisado (ANS) es correcta:

A) Tanto en ANS como en AS se requieren datos de entrenamiento sin etiqueta de clase.
B) En ANS se requieren datos de entrenamiento sin etiqueta de clase; en AS, con etiqueta.
C) En ANS se requieren datos de entrenamiento con etiqueta de clase; en AS, sin etiqueta.
D) Tanto en ANS como en AS se requieren datos de entrenamiento con etiqueta de clase.

11 $\boxed{\text{D}}$ Considérese el algoritmo C-medias en su versión convencional o "popular" (CM), así como en su versión de Duda y Hart (DH). Aunque ambas optimizan la *suma de errores cuadráticos* (SEC), sus resultados pueden diferir pues:

A) DH mimimiza la SEC y CM la maximiza.
B) DH maximiza la SEC y CM la minimiza.
C) Ambas maximizan la SEC, si bien DH puede alcanzar mejores soluciones que CM.
D) Ninguna de las anteriores.

12 $\boxed{\text{C}}$ Dado el modelo oculto de Markov M que se muestra en la figura de la derecha en el que $P_M(a) = P_M(b) = \frac{1}{4}$, ¿cuál es el valor $S = \sum_x P_M(x)$ donde x es cualquier posible cadena formada por dos o más símbolos?

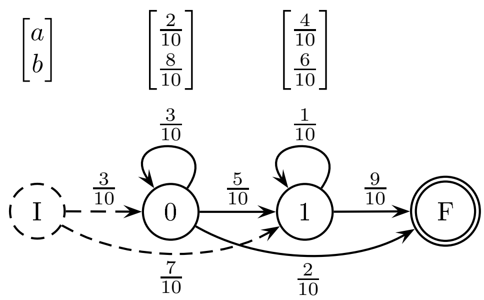


A) $0 \leq S < \frac{1}{4}$.
B) $\frac{1}{4} \leq S < \frac{2}{4}$.
C) $\frac{2}{4} \leq S < \frac{3}{4}$.
D) $\frac{3}{4} \leq S \leq 1$.

13 $\boxed{\text{B}}$ Siendo M un modelo oculto de Markov y x una cadena tal que $P_M(x) > 0$, siempre se cumple que:

A) La secuencia de estados de M que genera la cadena x con máxima probabilidad es única.
B) La aproximación de Viterbi a $P_M(x)$ es única.
C) La secuencia de estados de M que genera la cadena x con máxima probabilidad no es única.
D) La aproximación de Viterbi a $P_M(x)$ no es única.

14 $\boxed{\text{D}}$ Se tiene un problema de clasificación en dos clases (A y B) de objetos representados mediante cadenas de símbolos en el alfabeto $\Sigma = \{a, b\}$. Las funciones de probabilidad condicional de las clases vienen caracterizadas por los modelos ocultos de Markov M_A y M_B. Supóngase que $P(A) = 0.45$, $P(ba \mid A) = P_{M_A}(ba) = 0.0612$ y $P(ba \mid B) = P_{M_B}(ba)$, siendo M_B el modelo representado en la figura de la derecha. ¿A qué clase se asignaría la cadena "`ba`" por mínima probabilidad de error?:



A) Con los datos aportados no se puede determinar.
B) Indistintamente en A ó B ya que $P_{M_A}(ba) = P_{M_B}(ba)$.
C) En la clase A.
D) En la clase B.

$\hat{c} = \arg\max_c P(c)P(ba \mid c)$
$P(A)P(ba \mid A) = 0.45 \cdot 0.0612$
$P(B)P(ba \mid B) = 0.55 \cdot 0.0612$
$\hat{c} = B$

15 $\boxed{\text{A}}$ Dado el modelo oculto de Markov M_B de la pregunta anterior, tras *una* iteración de re-estimación por Viterbi a partir de las cadenas de entrenamiento "`ba`", "`b`" y "`aa`", indica cuál de los siguientes resultados es cierto:

A) $A_{01} = A_{1F} = 1$
B) $B_{0a} = B_{1a} = \frac{1}{2}$
C) $\pi_0 = \frac{1}{3}$
D) $\pi_1 = \frac{2}{3}$

A	0	1	F
0	0	1	0
1	0	0	1

B	a	b
0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{2}{3}$	$\frac{1}{3}$

$\pi_0 = \frac{2}{3}$ $\pi_1 = \frac{1}{3}$