

Examen del Bloque 2 de Sistemas Inteligentes (tipo D)

ETSINF, UPV, 10 de diciembre de 2018. Puntuación: numaciertos - numerroses/3.

- 1 **D** ¿Cuál de las siguientes distribuciones de probabilidad *no puede* deducirse a partir de la prob. conjunta $P(x, y, z)$?:
- A) $P(x | y)$.
 - B) $P(z | x, y)$.
 - C) $P(z)$.
 - D) Toda distribución en la que intervenga cualquier combinación de estas variables puede deducirse de $P(x, y, z)$.

- 2 **C** Sea un problema de clasificación en cuatro clases, $C = \{a, b, c, d\}$, donde las cuatro clases son equiprobables, y sea y un hecho o dato. La decisión óptima de clasificación para y es la clase a con una probabilidad a posteriori de 0.30. ¿Cuál de las siguientes afirmaciones es correcta?
- A) La probabilidad de error es menor que 0.50.
 - B) $P(C = a | Y = y) > P(C = b | Y = y) + P(C = c | Y = y) + P(C = d | Y = y)$.
 - C) $P(Y = y | C = a) = 0.3 \cdot P(Y = y) / 0.25$.
 - D) Ninguna de las anteriores.

- 3 **C** Supóngase que tenemos dos cajas con 40 galletas cada una. La primera caja contiene 10 galletas de chocolate y 30 sin chocolate. La segunda caja contiene 20 galletas de cada tipo. Ahora supóngase que se escoge una caja al azar, y luego una galleta al azar de la caja escogida. Si la galleta escogida no es de chocolate, la probabilidad P de que proceda de la primera caja es:

- A) $0/4 \leq P < 1/4$.
- B) $1/4 \leq P < 2/4$.
- C) $2/4 \leq P < 3/4$.
- D) $3/4 \leq P \leq 4/4$.

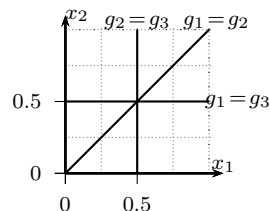
$$P(C = 1 | G = c) = \frac{P(C = 1) P(G = c | C = 1)}{P(C = 1) P(G = c | C = 1) + P(C = 2) P(G = c | C = 2)} = 0.6$$

- 4 **A** Sea $\mathbf{x} = (x_1, \dots, x_D)^t$, $D > 1$, un objeto representado mediante un vector de características D -dimensional a clasificar en una de C clases. Indica cuál de los siguientes clasificadores *no* es de error mínimo:

- A) $c(\mathbf{x}) = \arg \max_{c=1, \dots, C} p(x_1 | c) p(x_2, \dots, x_D | x_1, c)$
- B) $c(\mathbf{x}) = \arg \max_{c=1, \dots, C} p(c) p(x_1, \dots, x_D | c)$
- C) $c(\mathbf{x}) = \arg \max_{c=1, \dots, C} p(c | x_1) p(x_2, \dots, x_D | x_1, c)$
- D) $c(\mathbf{x}) = \arg \max_{c=1, \dots, C} p(x_1, c) p(x_2, \dots, x_D | x_1, c)$

- 5 **B** En la figura de la derecha se representan las fronteras de decisión de un clasificador en 3 clases. ¿Cuales de los siguientes vectores de pesos definen dichas fronteras?

- A) $\mathbf{w}_1 = (0, 0, 1)^t$ $\mathbf{w}_2 = (0, 1, 0)^t$ y $\mathbf{w}_3 = (1, 0, 0)^t$
- B) $\mathbf{w}_1 = (0, 0, 1)^t$ $\mathbf{w}_2 = (0, 1, 0)^t$ y $\mathbf{w}_3 = (0.5, 0, 0)^t$
- C) $\mathbf{w}_1 = (0.5, 0, 0)^t$ $\mathbf{w}_2 = (0, 1, 0)^t$ y $\mathbf{w}_3 = (0, 0, 1)^t$
- D) $\mathbf{w}_1 = (0, 0, 1)^t$ $\mathbf{w}_2 = (1, 0, 0)^t$ y $\mathbf{w}_3 = (0, 1, 0)^t$

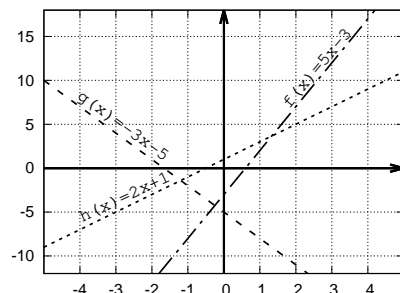


- 6 **A** Sea un clasificador lineal para dos clases, \circ y \bullet , de vectores de pesos $\mathbf{a}_\circ = (2, -5, 4)^t$ y $\mathbf{a}_\bullet = (5, 1, 1)^t$, respectivamente. ¿Cuál de las siguientes afirmaciones es correcta?

- A) Los vectores de pesos $\mathbf{a}_\circ = (3, 4, 1)^t$ y $\mathbf{a}_\bullet = (2, 2, 2)^t$ definen la misma frontera de decisión que los del enunciado.
- B) Los vectores de pesos $\mathbf{a}_\circ = (-2, 5, -4)^t$ y $\mathbf{a}_\bullet = (-5, -1, -1)^t$ definen un clasificador equivalente al del enunciado.
- C) El punto $\mathbf{x}' = (1, 2)^t$ pertenece a la clase \circ .
- D) El punto $\mathbf{x}' = (-2, 0)^t$ se encuentra en la frontera de decisión.

- 7 **C** En la figura de la derecha se muestran las funciones discriminantes lineales resultantes de entrenar un clasificador con el algoritmo Perceptrón con un conjunto de puntos de \mathbb{R} . Las funciones obtenidas son: $g(x) = -3x - 5$, $h(x) = 2x + 1$ y $f(x) = 5x - 3$. Indica cuáles son las fronteras de decisión correctas entre $g(x)$ y $h(x)$, y entre $h(x)$ y $f(x)$:

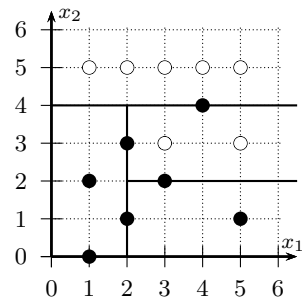
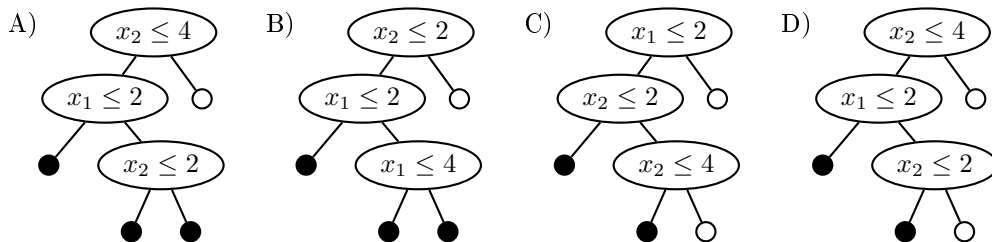
- A) $x = -5/3$ y $x = 3/5$.
- B) $x = -1/2$ y $x = 3/5$.
- C) $x = -6/5$ y $x = 4/3$.
- D) $x = -5/3$ y $x = 4/3$.



- 8 **A** Indica cuál de las siguientes afirmaciones referentes al algoritmo Perceptrón (al que llamaremos P) es *cierta* cuando se aplica al aprendizaje con una muestra de vectores etiquetados S :

- A) Si la muestra de aprendizaje es linealmente separable, P termina tras un número finito de iteraciones y los pesos resultantes permiten clasificar S sin errores.
- B) El número de vectores de S bien clasificados con los pesos obtenidos en cada iteración de P es mayor que el número de vectores bien clasificados en la iteración anterior.
- C) P siempre converge en un número finito de iteraciones, aunque es posible que los pesos finalmente obtenidos no clasifiquen correctamente a todos los vectores de S .
- D) Cuanto más grande es S , mayor es el número de iteraciones que necesita P para converger.

- 9 [D] Dado el conjunto de muestras bidimensionales 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?



- 10 [C] 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 *splits diferentes* que hay que explorar en el nodo raíz del árbol? Nota: no deben tenerse en cuenta los *splits* que dan lugar a nodos vacíos.

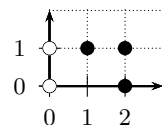
n	1	2	3	4	5	6
x_{n1}	0	1	0	1	0	1
x_{n2}	1	1	2	2	3	3
x_{n3}	0	0	2	3	2	3
c_n	A	A	B	B	C	C

- A) $0 \leq N < 2$.
 B) $2 \leq N < 4$.
 C) $4 \leq N < 6$. $\{(1,0), (2,1), (2,2), (3,2)\}$
 D) $6 \leq N$.
- 11 [B] Tenemos un problema de clasificación en tres clases, $C = \{a, b, c\}$ para objetos representados en un espacio de dos dimensiones ($\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2$). Tenemos las siguientes cuatro muestras: $\mathbf{y}_1 = (4, 1)^t$, pertenece a la clase a ; $\mathbf{y}_2 = (1, 2)^t$ y $\mathbf{y}_3 = (2, 3)^t$ pertenecen a la clase b ; y $\mathbf{y}_4 = (5, 1)^t$ pertenece a la clase c . Queremos construir un árbol de clasificación y el algoritmo ha alcanzado un nodo t que incluye los 4 datos mencionados. Utilizando la reducción de la impureza (en términos de entropía) para medir la calidad de un *split*, indica cuál de las siguientes afirmaciones es correcta:
- A) $\Delta \mathcal{I}(1, 2, t) > \Delta \mathcal{I}(2, 1, t)$.
 B) $\Delta \mathcal{I}(1, 2, t) > \Delta \mathcal{I}(2, 2, t)$.
 C) $\Delta \mathcal{I}(2, 2, t) > \Delta \mathcal{I}(2, 1, t)$.
 D) $\Delta \mathcal{I}(1, 4, t) = 0$.

- 12 [B] Sea T un árbol de clasificación construido mediante el algoritmo ADC a partir de una muestra de vectores etiquetados S . Indica cuál de las siguientes afirmaciones es *falsa*:
- A) Si el parámetro ϵ es suficientemente pequeño, el número de vectores de S que T clasifica incorrectamente puede ser tan pequeño como se quiera.
 B) Para todo nodo t de T , su impureza es igual a la suma de las impurezas de sus nodos hijos, t_R y t_L .
 C) El número de *splits* posibles en cualquier nodo de T es siempre menor o igual que $D \cdot |S|$, donde D es la dimensión de los vectores de S .
 D) Aunque T suele ser un árbol aproximadamente bien equilibrado, su altura puede ser mayor que $\log_2 |S|$.

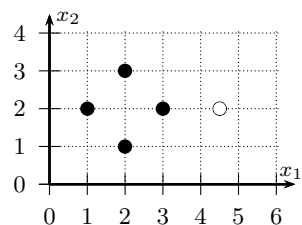
- 13 [C] En la figura de la derecha se muestra una partición de 5 puntos bidimensionales de 2 clusters. La transferencia del punto $(1, 1)^t$ del cluster \bullet al cluster \circ

- A) produce un incremento en la SEC.
 B) produce un decremento en la SEC.
 C) no altera la SEC.
 D) produce una SEC negativa.



- 14 [C] En la figura de la derecha se representan 5 muestras bidimensionales particionadas inicialmente en dos clústers (\bullet y \circ). ¿Cuál sería el resultado de la aplicación de una iteración del algoritmo C -medias en su versión convencional?, ¿y en su versión Duda y Hart (D&H)?

- A) Ambas versiones transfieren la muestra $(3, 2)$.
 B) Sólo la versión convencional transfiere la muestra $(3, 2)$.
 C) Sólo la versión D&H transfiere la muestra $(3, 2)$.
 D) Ninguna de las dos versiones transfiere la muestra $(3, 2)$.



- 15 [A] Se aplica el algoritmo C -medias de Duda y Hart a un conjunto de N vectores no etiquetados y se obtiene una partición de dicho conjunto en C subconjuntos disjuntos cuya suma de errores cuadráticos, SEC, es R . ¿Cuál de las siguientes afirmaciones es *falsa*?:

- A) Si $C \geq N/2$, $R = 0$.
 B) Si $C = N$, $R = 0$.
 C) Si $C \leq N$, C -medias termina en un número finito de iteraciones y R es un mínimo local de la SEC.
 D) Ninguna de las anteriores.