Examen del Bloque 2 de Sistemas Inteligentes

ETSINF, Universitat Politècnica de València, 15 de enero de 2014

Apellidos: ______________________ **Nombre:** ____________

Grupo: ☐ **3A** ☐ **3B** ☐ **3C** ☐ **3D** ☐ **3E** ☐ **RE1** ☐ **RE2**

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

1 $\boxed{\text{B}}$ Dada la probabilidad conjunta de dos variables aleatorias X y Y, la probabilidad condicional $P(Y = y \mid X = x)$ se puede obtener mediante:

A) $P(y \mid x) \;=\; 1 \;/\; P(x, y)$
B) $P(y \mid x) \;=\; P(x, y) \;/\; \sum_{y'} P(x, y')$
C) $P(y \mid x) \;=\; \sum_{x'} P(x', y) \;/\; \sum_{y'} P(x, y')$
D) $P(y \mid x) \;=\; \sum_{x'} P(x', y) \cdot \sum_{y'} P(x, y')$

2 $\boxed{\text{A}}$ En un problema de decisión binario ($D = \{0, 1\}$), sea y un hecho o dato y $d^\star(y) = 0$ la decisión de mínimo error para ese y. Identifica cuál de las siguientes expresiones determina *incorrectamente* la mínima probabilidad de error para dicho y:

A) $P_\star(\text{error} \mid Y = y) \;=\; 1 - P(D = 1 \mid Y = y)$
B) $P_\star(\text{error} \mid Y = y) \;=\; 1 - P(D = 0 \mid Y = y)$
C) $P_\star(\text{error} \mid Y = y) \;=\; P(D = 1 \mid Y = y)$
D) $P_\star(\text{error} \mid Y = y) \;=\; 1 - \text{máx}_d\, P(D = d \mid Y = y)$

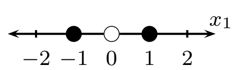
3 $\boxed{\text{D}}$ En un problema de diagnóstico diferencial entre *Gripe* y *Resfriado*, se sabe que la incidencia relativa de la *Gripe* con respecto al *Resfriado* es del 30 % y se conocen las siguientes distribuciones de temperaturas corporales:

$t(^oC)$	36	37	38	39	40
$P(T = t \mid D = \text{GRIPE})$	0.05	0.10	0.20	0.30	0.35
$P(T = t \mid D = \text{RESFR})$	0.10	0.30	0.40	0.15	0.05

La probabilidad a posteriori de que un paciente con 38º de fiebre tenga *Gripe* es:

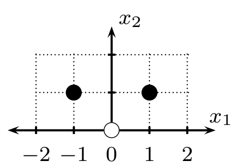
A) mayor que 0.8
B) menor que 0.1
C) entre 0.3 y 0.6
D) menor que la probabilidad de que con esa temperatura tenga *Resfriado*

4 $\boxed{\text{B}}$ En la figura de la derecha se representan tres muestras de aprendizaje unidimensionales de 2 clases: ○ y ●. ¿Cuál será el número de errores de clasificación cometidos por un clasificador lineal de mínimo error?



A) 0
B) 1
C) 2
D) 3

5 $\boxed{\text{A}}$ Supongamos que en el ejercicio anterior añadimos un nueva característica x_2 que se define como $x_2 = x_1^2$. De esta forma las tres muestras de aprendizaje pasan a ser bidimensionales como se observa en la figura de la derecha. En este caso, ¿cuál será el número de errores de clasificación cometidos por un clasificador lineal de mínimo error?



A) 0
B) 1
C) 2
D) 3

6 $\boxed{\text{A}}$ Sea un problema de clasificación en 2 clases, $c = 1, 2$, para objetos representados mediante vectores de características bidimensionales. Se tienen 2 muestras de entrenamiento: $\mathbf{x}_1 = (0, 0)^t$ de la clase $c_1 = 1$, y $\mathbf{x}_2 = (1, 1)^t$ de $c_2 = 2$. Asimismo, se tiene un clasificador lineal definido por los vectores de pesos: $\mathbf{w}_1 = (w_{10}, w_{11}, w_{12}) = (1, -1, -1)^t$ y $\mathbf{w}_2 = (w_{20}, w_{21}, w_{22}) = (-1, 1, 1)^t$. Si aplicamos una iteración del algoritmo Perceptrón a partir de estos vectores de pesos, con factor de aprendizaje $\alpha = 1$ y margen $b = 0.1$, entonces:

A) No se modificará ningún vector de pesos.
B) Se modificará el vector de pesos de la clase 1.
C) Se modificará el vector de pesos de la clase 2.
D) Se modificarán los vectores de pesos de ambas clases.

7 $\boxed{\text{C}}$ El algoritmo Perceptrón está gobernado por dos parámetros que denominamos *velocidad de aprendizaje*, α, y *margen*, b, siendo ambos valores reales. En caso de que no supiéramos si las muestras de aprendizaje son linealmente separables, ¿qué valores de los parámetros α y b proporcionan mayores garantías de obtener fronteras de decisión de mejor calidad?

A) $\alpha = 0.1$ y $b = 0.0$.
B) $\alpha = 0.0$ y $b = 0.0$.
C) $\alpha = 0.1$ y $b = 1.0$.
D) $\alpha = 0.0$ y $b = 1.0$.

8 $\boxed{\text{C}}$ Sea un problema de clasificación en C clases, $c = 1, \ldots, C$, para el que se ha aprendido un árbol de clasificación T. Sea t un nodo de T cuya impureza viene dada mediante la entropía, $H(t)$, asociada a las probabilidades a posteriori de las clases en t, $P(1 \mid t), \ldots, P(C \mid t)$. El nodo t será máximamente puro cuando:

A) Las clases sean equiprobables; esto es, $P(1 \mid t) = \cdots = P(C \mid t) = \frac{1}{C}$.
B) Exista una clase c^* de mayor probabilidad que el resto; esto es, $P(c^* \mid t) > P(c \mid t)$ para todo $c \neq c^*$.
C) Exista una clase c^* de probabilidad 1; esto es, tal que $P(c^* \mid t) = 1$.
D) Ninguna de las anteriores.

9 $\boxed{\text{A}}$ Sea un problema de clasificación en 2 clases, $c = 1, 2$, para objetos representados mediante vectores de características reales bidimensionales; esto es, de la forma $\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2$. Se tienen 4 muestras de entrenamiento: $\mathbf{y_1} = (1, 0.2)^t$, perteneciente a la clase 1; y $\mathbf{y_2} = (2, 0.2)^t$, $\mathbf{y_3} = (3, 0.8)^t$ y $\mathbf{y_4} = (1, 0.8)^t$, pertenecientes a la clase 2. Queremos construir un árbol de clasificación empleando el decremento de impureza (medida en términos de entropía) para medir la calidad de una partición de un nodo. En el caso del nodo raíz y considerando sólo la característica y_1, ¿cuál de las siguientes afirmaciones es *cierta*? (Nota: $\log_2(1/3) = -1.585$ y $\log_2(2/3) = -0.585$).

A) La mejor partición es $y_1 \leq 1$.
B) La mejor partición es $y_1 \leq 2$.
C) La mejor partición es $y_1 \leq 3$.
D) Ninguna de las anteriores.

10 $\boxed{\text{C}}$ Sea un problema de clasificación en C clases, $c = 1, \ldots, C$, para el que se ha aprendido un árbol de clasificación T. Sea t un nodo terminal de T en el que se han estimado las probabilidades a posteriori de las clases $\hat{P}(1 \mid t), \ldots, \hat{P}(C \mid t)$. Un criterio simple y eficaz para asignar una etiqueta de clase a t es:

A) La de una clase de probabilidad a posteriori mínima.
B) La de una clase de probabilidad a posteriori próxima a la media (i.e. $\frac{1}{C}$).
C) La de una clase de probabilidad a posteriori máxima.
D) Ninguna de las anteriores.

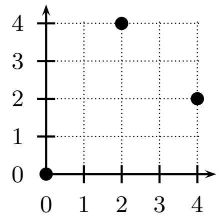
11 $\boxed{\text{D}}$ Indica cuál de la siguientes afirmaciones sobre *Clustering* es correcta:

A) Se suele emplear el algoritmo *Perceptrón* a partir de datos de entrenamiento *con* etiquetas de clase.
B) Se suele emplear el algoritmo *Perceptrón* a partir de datos de entrenamiento *sin* etiquetas de clase.
C) Se suele emplear el algoritmo *C-Medias* a partir de datos de entrenamiento *con* etiquetas de clase.
D) Se suele emplear el algoritmo *C-Medias* a partir de datos de entrenamiento *sin* etiquetas de clase.

12 $\boxed{\text{D}}$ El criterio de clustering particional "Suma de Errores Cuadráticos" es apropiado cuando los datos forman clústers:

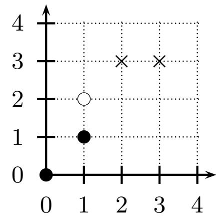
A) No alargados.
B) Alargados y de cualquier tamaño.
C) Alargados y de tamaño similar.
D) Ninguna de las anteriores.

13 $\boxed{\text{B}}$ La menor suma de errores cuadráticos con la que pueden agruparse en dos clústers los puntos a la derecha es un valor:



A) Entre 0 y 3.
B) Entre 3 y 6. $J = 4$
C) Entre 6 y 9.
D) Mayor que 9.

14 $\boxed{\text{B}}$ La figura a la derecha muestra una partición de 5 puntos bidimensionales en 3 clústers (representados mediante los símbolos $\bullet$, $\circ$ y $\times$). Considera todas las posibles transferencias de clúster de cada punto (en un clúster no unitario). En términos de suma de errores cuadráticos (J):

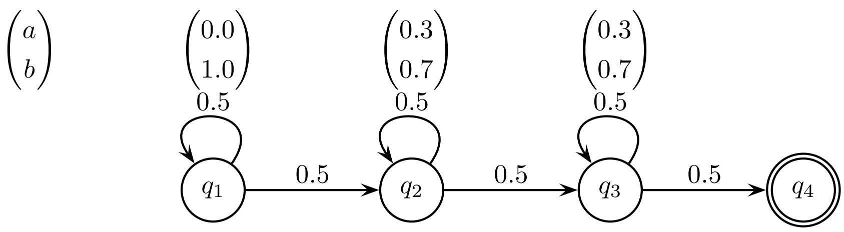


A) Ninguna transferencia permite mejorar J.
B) Sólo se puede mejorar J transfiriendo $(1, 1)^t$ del clúster $\bullet$ al $\circ$.
C) Sólo se puede mejorar J transfiriendo $(2, 3)^t$ del clúster $\times$ al $\circ$.
D) Las dos transferencias anteriores permiten mejorar J.

15 $\boxed{\text{C}}$ Dado un Modelo Oculto de Markov Θ y una cadena y reconocida por dicho modelo, indica cuál de las siguientes afirmaciones es cierta:

A) Siempre se cumple que $P(y|\Theta) = \widetilde{P}(y|\Theta)$.
B) Siempre se cumple que $P(y|\Theta) \leq \widetilde{P}(y|\Theta)$.
C) Siempre se cumple que $P(y|\Theta) \geq \widetilde{P}(y|\Theta)$.
D) Siempre se cumple que $P(y|\Theta) \neq \widetilde{P}(y|\Theta)$.

16 $\boxed{\text{C}}$ Dado el Modelo Oculto de Markov Θ



con $\pi_{q_1} = 1, \pi_{q_2} = \pi_{q_3} = \pi_{q_4} = 0$ y las cadenas $y_1 =$ "babb" y $y_2 =$ "aaaa", indica cuál de las siguientes afirmaciones es cierta:

A) $P(y_1|\Theta) = P(y_2|\Theta)$.
B) $P(y_1|\Theta) < P(y_2|\Theta)$.
C) $P(y_1|\Theta) > P(y_2|\Theta)$.
D) $P(y_1|\Theta) = P(y_2|\Theta) = 0$.

17 $\boxed{\text{D}}$ Dado el Modelo Oculto de Markov Θ de la pregunta anterior, si lo estimamos con una sola iteración con la muestra $Y = \{baba, abab\}$ utilizando el algoritmo de Viterbi indica cuál de las siguientes afirmaciones es cierta:

A) El modelo estimado tiene todos los parámetros a 0.0.
B) Ninguna probabilidad de transición entre estados en el modelo estimado toma valor 0.0.
C) El modelo obtenido tiene todos los parámetros igual al modelo incial.
D) El modelo obtenido queda con varios parámetros con valor 0.

18 $\boxed{\text{C}}$ En relación al algoritmo *forward* definido para Modelos Ocultos de Markov indica cuál de la siguientes afirmaciones es verdadera:

A) Calcula la probabilidad de una cadena teniendo en cuenta sólo la secuencia de análisis de máxima probabilidad.
B) Calcula la probabilidad de una cadena sin incluir la probabilidad de la secuencia de estados más probable.
C) Calcula la probabilidad de una cadena incluyendo todas las secuencias de estados.
D) Nunca calcula la probabilidad de una cadena.