Examen Final de Sistemas Inteligentes: Bloque 2

ETSINF, Universitat Politècnica de València, 26 de enero de 2015

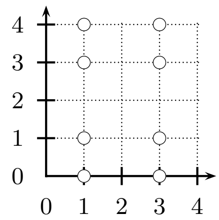
Apellidos: ____________ **Nombre:** ____________

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

Cuestiones (2 puntos; tiempo estimado: 30 minutos)

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

1 $\boxed{\text{B}}$ La figura a la derecha muestra 8 puntos bidimensionales. La menor "Suma de Errores Cuadráticos", J, con la que pueden agruparse estos puntos en dos clústers es:



A) $0 \leq J \leq \ 7$

B) $7 < J \leq 14$ $\quad J = 10$

C) $14 < J \leq 21$

D) $21 < J$

2 $\boxed{\text{D}}$ Sean X, Y y Z tres variables aleatorias. Se dice que X e Y son *condicionalmente independientes* dada Z si y solo si

$$P(X=x, Y=y \mid Z=z) \;=\; P(X=x \mid Z=z)\, P(Y=y \mid Z=z) \qquad \text{para todo } x,\ y \text{ y } z.$$

Si se cumple esta igualdad, podemos calcular $P(Z=z \mid X=x, Y=y)$ como sigue:

A) $P(Z=z \mid X=x, Y=y) \;=\; \dfrac{P(X=x, Y=y, Z=z)}{P(X=x, Y=y)}$

B) $P(Z=z \mid X=x, Y=y) \;=\; \dfrac{P(Z=z)\, P(X=x, Y=y \mid Z=z)}{P(X=x, Y=y)}$

C) $P(Z=z \mid X=x, Y=y) \;=\; \dfrac{P(Z=z)\, P(X=x \mid Z=z)\, P(Y=y \mid Z=z)}{P(X=x, Y=y)}$

D) De las tres maneras anteriores.

3 $\boxed{\text{C}}$ Se quiere construir un sistema de reconocimiento de formas para dígitos manuscritos representados mediante cadenas de contorno de 4 direcciones; esto es, mediante cadenas de símbolos en el alfabeto $\Sigma = \{1, 2, 3, 4\}$. Dada una secuencia de cadenas de entrenamiento con sus correspondientes etiquetas de clase, construiremos el sistema como sigue:

A) Emplearemos el algoritmo Perceptrón y obtendremos un clasificador lineal.

B) Aprenderemos un Árbol de Decisión y Clasificación mediante el algoritmo ADC.

C) Diseñaremos un clasificador basado en modelos de Markov aplicando el algoritmo de re-estimación por Viterbi.

D) Las tres opciones anteriores son válidas.

4 $\boxed{\text{C}}$ En un problema de clasificación en tres clases ($C = \{a, b, c\}$), en el que se dispone de 100 muestras de la clase a, 100 muestras de la clase b y 100 muestras de la clase c, sea y un hecho o dato. La decisión óptima de clasificación para y es la clase a con una probabilidad a posteriori de 0.50. ¿Cuál de las siguientes afirmaciones es correcta?

A) $P(C=a \mid Y=y) \;>\; P(C=b \mid Y=y) + P(C=c \mid Y=y)$

B) $P(Y=y \mid C=a) \;=\; \dfrac{0.5\ P(C=a)}{P(Y=y)}$

C) $P(Y=y \mid C=a) \;=\; P(Y=y \mid C=b) + P(Y=y \mid C=c)$

D) Ninguna de las anteriores.

5 $\boxed{\text{C}}$ Dado un clasificador lineal de 2 clases $\circ$ y $\bullet$ definido por su conjunto de pesos $\mathbf{a}_\circ = (0, -1, 1)^t$ y $\mathbf{a}_\bullet = (0, 1, -1)^t$, ¿Qué conjunto de pesos de los siguientes no define un clasificador equivalente al dado?

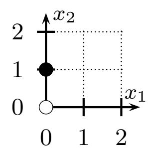
A) $\mathbf{a}_\circ = (1, -1, 1)^t$ y $\mathbf{a}_\bullet = (1, 1, -1)^t$ $\quad f(z) = az + b$ con $a = 1$ y $b = 1$

B) $\mathbf{a}_\circ = (-1, -2, 2)^t$ y $\mathbf{a}_\bullet = (-1, 2, -2)^t$ $\quad f(z) = az + b$ con $a = 2$ y $b = -1$

C) $\mathbf{a}_\circ = (0, 2, -2)^t$ y $\mathbf{a}_\bullet = (0, -2, 2)^t$ $\quad f(z) = az + b$ con $a = -2$ y $b = 0$

D) $\mathbf{a}_\circ = (0, -2, 2)^t$ y $\mathbf{a}_\bullet = (0, 2, -2)^t$ $\quad f(z) = az + b$ con $a = 2$ y $b = 0$

6 $\boxed{\text{A}}$ En la figura de la derecha se representan dos muestras de aprendizaje bidimensionales de 2 clases: (x_1,$\circ$) y (x_2,$\bullet$). Dados el conjunto de pesos $\mathbf{a}_\circ = (0, 1, -2)^t$ y $\mathbf{a}_\bullet = (0, 0, 1)^t$, si aplicamos una iteración del algoritmo Perceptrón con factor de aprendizaje $\alpha = 1.0$ y margen $b = 0.5$ a partir del conjunto de pesos y muestras de aprendizaje dadas, ¿cuántos errores de clasificación se producen sobre las muestras de aprendizaje con el nuevo conjunto de pesos?



A) 0 $\quad \mathbf{a}_\circ = (1, 1, -2)^t$ y $\mathbf{a}_\bullet = (-1, 0, 1)^t$

B) 1

C) 2

D) 3

Examen Final de Sistemas Inteligentes: Bloque 2

ETSINF, Universitat Politècnica de València, 26 de enero de 2015

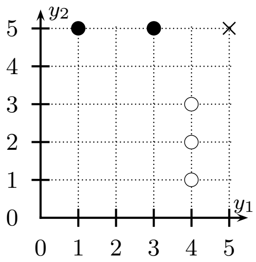
Apellidos: ______________________ **Nombre:** ______________

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

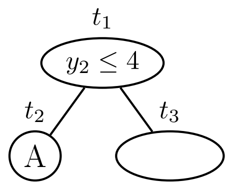
Problemas (3 puntos; tiempo estimado: 60 minutos)

1. **(1.5 puntos)**

Para aprender un árbol de clasificación se dispone de una muestra de entrenamiento formada por 6 vectores bidimensionales pertenecientes a 3 clases, A, B y C. Estos vectores se muestran en la figura a la derecha ($A = \circ$, $B = \bullet$ y $C = \times$). En las primeras invocaciones recursivas del algoritmo ADC (con $\epsilon = 0.5$ bits) se ha producido el sub-árbol con tres nodos que se muestra en la figura de abajo. Este sub-arbol corresponde a una primera división óptima de la muestra de entrenamiento en dos subconjuntos mediante el *"split"* $(2, 4.0)$ (es decir, $y_2 \leq 4$). En este proceso inicial se han obtenido los parámetros que se muestran en la tabla.



Nodo	Split	$P(A \mid t_i)$	$P(B \mid t_i)$	$P(C \mid t_i)$	$P_{t_i}(L)$	$P_{t_i}(R)$	$\mathcal{I}(t_i)$	$\Delta\mathcal{I}(t_1)$
t_1	(2,4)	1/2	1/3	1/6	1/2	1/2	1.459	1.000
t_2	–	1	0	0	–	–	0	–
t_3		0	2/3	1/3				
t_4								
t_5								



a) Explicar cómo se obtienen los siguientes valores de la tabla: $P(A \mid t_1)$, $P(B \mid t_1)$, $P(C \mid t_1)$, $P_{t_1}(R)$ y $\mathcal{I}(t_1)$.

El nodo raiz (t_1) representa a los 6 los datos disponibles. De ellos hay 3 de la clase A, 2 de la clase B y 1 de la clase C. Por tanto: $P(A \mid t_1) = 3/6 = 1/2$, $P(B \mid t_1) = 2/6 = 1/3$, $P(C \mid t_1) = 1/6$

El *"split"* $(2, 4.0)$ $(t_2 \leq 4)$ divide el árbol raiz en dos subárboles: uno enraizado en t_2, que representa 3 datos tales que $y_2 \leq 4$, y otro en t_3, que representa otros 3 datos tales que $y_2 > 4$. Así pues: $P_{t_1}(R) = 3/6 = 1/2$

Finalmente: $\mathcal{I}(t_1) \;=\; -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{3}\log_2\frac{1}{3} - \frac{1}{6}\log_2\frac{1}{6} \;\approx\; 1.459$ bits

b) Calcular la impureza del nodo t_3.

$\mathcal{I}(t_3) \;=\; 0 - \frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \;\approx\; 0.918$ bits

c) Encontrar el *"split"* óptimo para el nodo t_3, completar la ejecución del algoritmo ADC y completar las celdas de tabla que están en blanco.

El nodo t_3 representa los vectores $((1,5)^t, B), ((3,5)^t, B), ((5,5)^t, C)$ para los que solo hay dos particiones posibles, correspondientes a los *"splits"*: $y_1 \leq 2$ y $y_1 \leq 4$. Los decrementos de impureza correspondientes son:

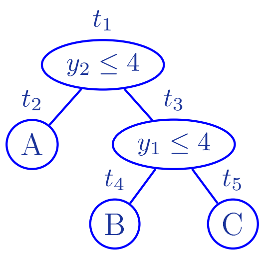
$$\Delta\mathcal{I}(1,2,t_3) = \mathcal{I}(t_3) - \tfrac{1}{3}\mathcal{I}(t_4) - \tfrac{2}{3}\mathcal{I}(t_5) \;\approx\; 0.918 - 0 - \tfrac{2}{3}\cdot 1 \;=\; 0.251 \text{ bits}$$

$$\Delta\mathcal{I}(1,4,t_3) = \mathcal{I}(t_3) - \tfrac{2}{3}\mathcal{I}(t_4) - \tfrac{1}{3}\mathcal{I}(t_5) \;\approx\; 0.918 - 0 - 0 \;=\; 0.918 \text{ bits}$$

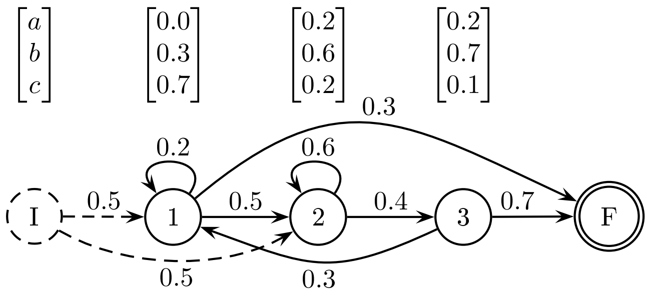
De estos, el mayor decremento es para el split $(1,4)$ (o sea, $y_1 \leq 4$).

El árbol resultante y los parámetros correspondientes se muestran abajo en la figura y tabla, respectivamente.

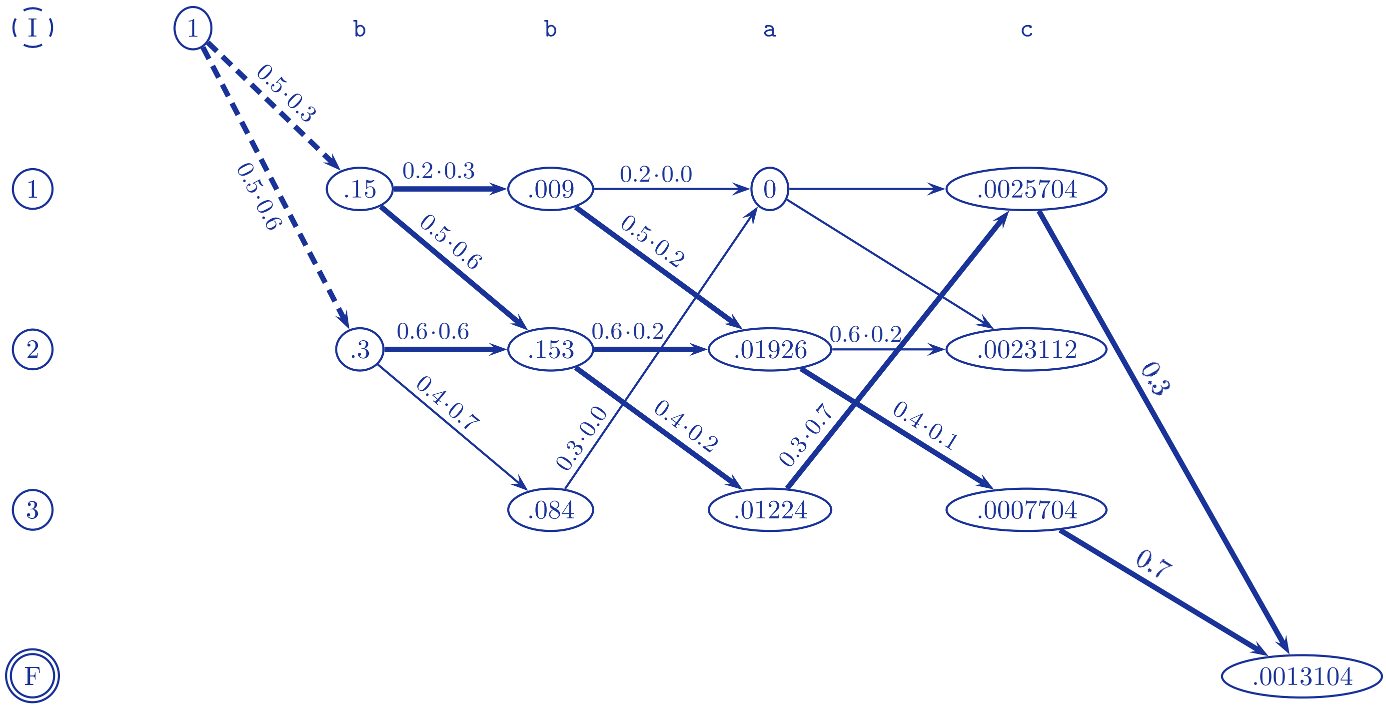
Nodo	Split	$P(A \mid t_i)$	$P(B \mid t_i)$	$P(C \mid t_i)$	$P_{t_i}(L)$	$P_{t_i}(R)$	$\mathcal{I}(t_i)$	$\Delta\mathcal{I}(t_1)$
t_1	(2,4)	1/2	1/3	1/6	1/2	1/2	1.459	1.000
t_2	–	1	0	0	–	–	0	–
t_3	(1,4)	0	2/3	1/3	2/3	1/3	0.918	0.918
t_4	–	0	1	0	–	–	0	–
t_5	–	0	0	1	–	–	0	–



2. **(1.5 puntos)** Sea M el modelo de Markov:



Calcula la probabilidad exacta de que M genere la cadena *bbac*, $P_M(bbac)$, mediante el algoritmo *Forward*.



$$P_M(bbac) \;=\; 0.0013104$$