

Examen del bloque 2 de SIN: Test (1,75 puntos)

ETSINF, Universitat Politècnica de València, 13 de enero de 2022

Grupo, apellidos y nombre: 2,

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

- 1 ☐ Dados los siguientes 3 nodos de un árbol de clasificación con muestras pertenecientes a 3 clases:

c	n_1	n_2	n_3
1	2	3	4
2	1	5	5
3	1	1	4

donde cada fila indica el número de muestras de cada clase en el nodo. ¿Cuál de las siguientes desigualdades es cierta?

- A) $\mathcal{I}(n_1) < \mathcal{I}(n_3) < \mathcal{I}(n_2)$
B) $\mathcal{I}(n_2) < \mathcal{I}(n_3) < \mathcal{I}(n_1)$
C) $\mathcal{I}(n_2) < \mathcal{I}(n_1) < \mathcal{I}(n_3)$
D) $\mathcal{I}(n_1) < \mathcal{I}(n_2) < \mathcal{I}(n_3)$
- 2 ☐ Dado el clasificador en 2 clases definido por sus vectores de pesos $\mathbf{w}_1 = (0, -3, 1, 1)^t$, $\mathbf{w}_2 = (0, 1, 0, 2)^t$ en notación homogénea, ¿cuál de los siguientes conjuntos de vectores **no** define un clasificador equivalente al dado?
- A) $\mathbf{w}_1 = (1, -3, 1, 1)^t$, $\mathbf{w}_2 = (1, 1, 0, 2)^t$
B) $\mathbf{w}_1 = (0, -9, 3, 3)^t$, $\mathbf{w}_2 = (0, 3, 0, 6)^t$
C) $\mathbf{w}_1 = (0, 6, -2, -2)^t$, $\mathbf{w}_2 = (0, -2, 0, -4)^t$
D) $\mathbf{w}_1 = (1, -9, 3, 3)^t$, $\mathbf{w}_2 = (1, 3, 0, 6)^t$

- 3 ☐ En un problema de razonamiento probabilístico correspondiente a desplazamientos por carretera, con las variables aleatorias de interés: Climatología (C):{despejado (DES), nublado (NUB), lluvioso (LLU)}; Luminosidad (L):{día (DIA), noche (NOC)}; Seguridad (S):{seguro (SEG), accidente (ACC)}. La probabilidad conjunta de las tres variables viene dada en la tabla:

$P(s, l, c)$	DIA			NOC		
	DES	NUB	LLU	DES	NUB	LLU
SEG	0.32	0.23	0.05	0.11	0.07	0.08
ACC	0.03	0.01	0.03	0.01	0.03	0.03

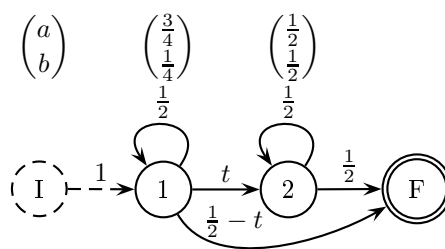
La probabilidad condicional $P(S = \text{ACC} \mid L = \text{NOC}, C = \text{NUB})$ es:

- A) 0.140
B) 0.300
C) 0.030
D) 0.100

- 4 ☐ Supóngase que estamos aplicando el algoritmo Perceptrón, con factor de aprendizaje $\alpha = 1$ y margen $b = 0.1$, a un conjunto de 3 muestras bidimensionales de aprendizaje para un problema de 2 clases. Tras procesar las primeras 2 muestras se han obtenido los vectores de pesos $\mathbf{w}_1 = (0, -1, -1)^t$, $\mathbf{w}_2 = (0, 1, 1)^t$. A continuación, se procesa la última muestra (\mathbf{x}_3, c_3) y se obtienen los vectores de pesos $\mathbf{w}_1 = (1, 0, 3)^t$, $\mathbf{w}_2 = (-1, 0, -3)^t$, ¿cuál de las siguientes es esa última muestra?

- A) $((1, 4)^t, 1)$
- B) $((5, 4)^t, 1)$
- C) $((3, 2)^t, 1)$
- D) $((4, 5)^t, 2)$

- 5 ☐ Sea M el modelo de Markov representado a la derecha, donde t , $0 < t < \frac{1}{4}$, denota la probabilidad de transición del estado 1 al 2. Dada la cadena $x = \mathbf{abb}$, la probabilidad de generar x mediante el camino $122F$, $P(\mathbf{abb}, 122F)$, depende de t . Análogamente, la probabilidad de generar x mediante el camino $111F$, $P(\mathbf{abb}, 111F)$, también depende de t (a través de la probabilidad de transición del estado 1 al F). Indica en qué caso $P(\mathbf{abb}, 111F) > P(\mathbf{abb}, 122F)$:



- A) Nunca.
- B) Si y solo si $0 < t < \frac{1}{20}$.
- C) Si y solo si $0 < t < \frac{1}{10}$.
- D) Siempre, es decir, $0 < t < \frac{1}{4}$.

- 6 ☐ La probabilidad de error de un clasificador se estima que es del 3%. Determina cuál es el número mínimo de muestras de test necesario, M , para conseguir que el intervalo de confianza al 95% de dicho error no supere el $\pm 1\%$; esto es, $I = [2\%, 4\%]$:

- A) $M < 2000$.
- B) $2000 \leq M < 3500$.
- C) $3500 \leq M < 5000$.
- D) $M \geq 5000$.

- 7 ☐ Sea M un modelo de Markov de conjunto de estados $Q = \{1, 2, F\}$ y alfabeto $\Sigma = \{a, b\}$. Durante la aplicación de una iteración del algoritmo de reestimación por Viterbi, se ha obtenido un par “(cadena, camino más probable)” por cada cadena de entrenamiento. Seguidamente, a partir de todos los pares obtenidos, se han obtenido las cuentas (frecuencias absolutas) de transición entre estados mostradas en la tabla a la derecha. La normalización *correcta* de estas cuentas resultará en la tabla de probabilidades de transición entre estados:

A	1	2	F
1	2	1	2
2	2	1	4

A)

A	1	2	F
1	$\frac{2}{4}$	$\frac{1}{2}$	$\frac{2}{6}$
2	$\frac{2}{4}$	$\frac{1}{2}$	$\frac{4}{6}$

B)

A	1	2	F
1	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{2}{12}$
2	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{4}{12}$

C)

A	1	2	F
1	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
2	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{4}{7}$

D)

A	1	2	F
1	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{2}{4}$
2	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{4}{4}$

- 8 ☐ Se tiene una partición de un conjunto de datos 3-dimensionales en un número de clústers dado, $C \geq 2$. Considérese la transferencia del dato $\mathbf{x} = (4, 5, 2)^t$ de un clúster i a otro j , $j \neq i$. Se sabe que el clúster i contiene 3 datos (contando \mathbf{x}) y el j 2. Asimismo, se sabe que la media del clúster i es $\mathbf{m}_i = (10, 8, 4)^t$ y la del j $\mathbf{m}_j = (7, 7, 1)^t$. Si se realiza dicha transferencia, se producirá un incremento de la suma de errores cuadráticos, ΔJ , tal que:

- A) $\Delta J < -70$
 B) $-70 \leq \Delta J < -30$
 C) $-30 \leq \Delta J < 0$
 D) $\Delta J \geq 0$

- 9 ☐ Dado el siguiente conjunto de datos utilizado para entrenar un árbol de clasificación con 5 muestras bidimensionales que pertenecen a 2 clases:

n	1	2	3	4	5
x_{n1}	5	4	5	4	3
x_{n2}	1	2	2	3	1
c_n	1	2	1	1	2

¿Cuántas particiones diferentes se podrían generar en el nodo raíz? No consideres aquellas particiones en que todos los datos se asignan al mismo nodo hijo.

- A) 4
 B) 2
 C) 3
 D) 5

Examen del bloque 2 de SIN: Problemas (2 puntos)

ETSINF, Universitat Politècnica de València, 13 de enero de 2022

Grupo, apellidos y nombre: 2,

Problema sobre Forward y Viterbi

Sea M un modelo de Markov de conjunto de estados $Q = \{1, 2, F\}$; alfabeto $\Sigma = \{a, b\}$; probabilidades iniciales $\pi_1 = \frac{1}{2}, \pi_2 = \frac{1}{2}$; y probabilidades de transición entre estados y de emisión de símbolos:

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

B	a	b
1	$\frac{3}{6}$	$\frac{3}{6}$
2	$\frac{1}{4}$	$\frac{3}{4}$

Sea $x=ab$. Se pide:

- (0,75 puntos) Realiza una traza del algoritmo *Forward* para obtener la probabilidad con la que M genera la cadena x , $P_M(x)$.
- (0,75 puntos) Realiza una traza del algoritmo de *Viterbi* para obtener la aproximación de Viterbi a la probabilidad con la que M genera la cadena x , $\tilde{P}_M(x)$.
- (0,25 puntos) A partir de la traza realizada en el apartado anterior, determina un camino más probable con el que M genera x .
- (0,25 puntos) Determina la probabilidad con la que M genera x siguiendo un camino distinto al más probable determinado en el apartado anterior.