Examen del Bloque 2 de Sistemas Inteligentes

ETSINF, Universitat Politècnica de València, 13 de enero de 2016

Apellidos: ☐ **Nombre:** ☐

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

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

1 $\boxed{\text{D}}$ ¿Cuál de las siguientes expresiones es correcta?

A) $P(x \mid y) \;=\; \frac{1}{P(z)} \sum_x P(x, y, z)$.

B) $P(x \mid y) \;=\; \frac{1}{P(z)} \sum_z P(x, y, z)$.

C) $P(x \mid y) \;=\; \frac{1}{P(y)} \sum_x P(x, y, z)$.

D) $P(x \mid y) \;=\; \frac{1}{P(y)} \sum_z P(x, y, z)$.

2 $\boxed{\text{A}}$ Un médico sabe que:

- La enfermedad de la meningitis causa rigidez de nuca en un 70 % de los casos.
- La probabilidad a priori de que un paciente tenga meningitis es de 1 / 100 000.
- La probabilidad a priori de que un paciente tenga rigidez de nuca es del 1 %.

Con base en el conocimiento anterior, la probabilidad P de que un paciente con rigidez de nuca tenga meningitis es:

A) $0.000 \leq P < 0.001$. $\quad P = P(m \mid r) = \frac{P(m)\,P(r|m)}{P(r)} = \frac{1/100\,000 \cdot 70/100}{1/100} = 0.0007$

B) $0.001 \leq P < 0.002$.

C) $0.002 \leq P < 0.003$.

D) $0.003 \leq P$.

3 $\boxed{\text{D}}$ Considérese un problema de clasificación convencional, esto es, de C clases y objetos representados mediante vectores D-dimensionales de características reales. En términos generales, podemos decir que el problema será más difícil...

A) cuanto menor sean C y D.

B) cuanto menor sea C y mayor sea D.

C) cuanto mayor sea C y menor sea D.

D) cuanto mayor sean C y D.

4 $\boxed{\text{B}}$ Se tiene un problema de clasificación para el cual se han aprendido dos clasificadores diferentes, c_A y c_B. La probabilidad de error de c_A se ha estimado empíricamente, a partir de un cierto conjunto de 100 muestras de test, obteniéndose un valor de $\hat{p}_A = 0.10$ (10 %). La probabilidad de error de c_B se ha estimado análogamente, si bien en este caso se ha empleado un conjunto de test diferente, compuesto por 200 muestras, obteniéndose también un 10 % de error ($\hat{p}_B = 0.10$). Con base en estas estimaciones, podemos afirmar que, para un nivel de confianza del 95 %:

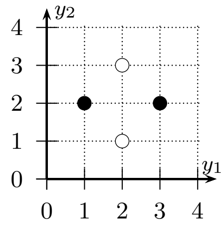
A) Los intervalos de confianza de $\hat{p}_A$ y $\hat{p}_B$ serán idénticos.

B) El intervalo de confianza de $\hat{p}_A$ será mayor que el de $\hat{p}_B$. $\quad I_A = \hat{p}_A \pm 1.96\sqrt{\frac{\hat{p}_A\,(1-\hat{p}_A)}{100}} = 0.10 \pm 0.06$

C) El intervalo de confianza de $\hat{p}_B$ será mayor que el de $\hat{p}_A$. $\quad I_B = \hat{p}_B \pm 1.96\sqrt{\frac{\hat{p}_B\,(1-\hat{p}_B)}{200}} = 0.10 \pm 0.04$

D) Los intervalos de confianza de $\hat{p}_A$ y $\hat{p}_B$ son en este caso irrelevantes ya que las tasas de error estimadas coinciden.

5 $\boxed{\text{C}}$ En la figura de la derecha se representan cuatro muestras de aprendizaje bidimensionales de 2 clases: $\circ$ y $\bullet$. A estas muestras se les aplica el algoritmo Perceptrón con pesos iniciales $\mathbf{a}_\circ = (0,1,0)^t$ y $\mathbf{a}_\bullet = (0,0,1)^t$, una constante de aprendizaje $\alpha > 0$ y un margen b. Indica cuál de las siguientes afirmaciones es correcta:



A) El algoritmo convergerá para algún $b > 0$.

B) El algoritmo solo puede converger si $b \leq 0$.

C) Si $b > 0$, no hay convergencia, pero se puede ajustar el valor de α tal que, tras un número finito de iteraciones, se obtengan buenas soluciones (con 25 % de error de resustitución).

D) El algoritmo no es aplicable a estas muestras porque no son linealmente separables.

6 $\boxed{\text{B}}$ ¿Cuál sería el número mínimo de errores de un clasificador lineal en el conjunto de muestras de la cuestión anterior?

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

7 $\boxed{\text{B}}$ Dado un clasificador lineal de 2 clases $\circ$ y $\bullet$ definido por su conjunto de pesos $\mathbf{a}_\circ = (3,1,1)^t$ y $\mathbf{a}_\bullet = (1,2,1)^t$ (en notación homogénea, cuya primera componente es el término independiente de la función lineal correspondiente). ¿Cuál de las siguientes afirmaciones es correcta?

A) Como hay dos vectores de pesos y el espacio de representación es bi-dimensional, tendremos 4 regiones de decisión.
B) Los vectores de pesos $\mathbf{a}_\circ = (2,-2,-2)^t$ y $\mathbf{a}_\bullet = (-2,0,-2)^t$ determinan la misma frontera de decisión que la del clasificador dado. La ecuación de la frontera es: $\mathbf{a}_\circ^t\mathbf{y} = \mathbf{a}_\bullet^t\mathbf{y}$. En ambos casos se obtiene: $y_1 = 2$.
C) Un clasificador equivalente al dado es el definido por $\mathbf{a}_\circ = (1,2,1)^t$ y $\mathbf{a}_\bullet = (3,1,1)^t$. Regiones de decisión opuestas.
D) Como los vectores de pesos son de tres dimensiones, la frontera viene dada por la ecuación de un plano en $\mathbb{R}^3$.

8 $\boxed{\text{D}}$ Supóngase que estamos aplicando el algoritmo de aprendizaje de árboles de clasificación para un problema de dos clases, A y B. El algoritmo ha alcanzado un nodo t que incluye dos datos: uno de la clase A y otro de la clase B. La impureza de t, $\mathcal{I}(t)$, medida como la entropía de la distribución emprírica de las probabilidades a posteriori de las clases en t, es:

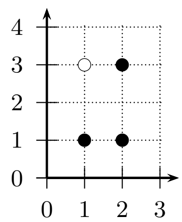
A) $\mathcal{I}(t) < 0.0$.

B) $0.0 \leq \mathcal{I}(t) < 0.5$.

C) $0.5 \leq \mathcal{I}(t) < 1.0$.

D) $1.0 \leq \mathcal{I}(t)$. $\quad \mathcal{I}(t) = -\hat{P}(A \mid t)\,\log_2 \hat{P}(A \mid t) - \hat{P}(B \mid t)\,\log_2 \hat{P}(B \mid t) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$

9 $\boxed{\text{D}}$ La figura a la derecha muestra una partición de 4 puntos bidimensionales en 2 clústers (representados mediante los símbolos $\bullet$ y $\circ$). La suma de errores cuadráticos (SEC) de esta partición es $J = \frac{30}{9}$. La transferencia del punto $(2,3)^t$ del clúster $\bullet$ al $\circ$ conduce a un incremento de la SEC, ΔJ, tal que:



A) $\Delta J > 0$.

B) $\quad 0 \geq \Delta J > -1$.

C) $-1 \geq \Delta J > -2$.

D) $-2 \geq \Delta J$. $\quad \Delta J = -\frac{21}{9} = -2.33 \quad (J = \frac{30}{9} \rightarrow J = 1)$

10 $\boxed{\text{B}}$ Dos versiones bien conocidas del algoritmo C-medias son la de *Duda y Hart* (DH) y la "popular". Suponiendo que ambas versiones se aplican a partir de un misma partición inicial, indica cuál de las siguientes afirmaciones sobre sus resultados es cierta:

A) Ambas versiones obtendrán la misma partición optimizada.
B) La versión DH obtendrá una partición final que no podrá mejorarse mediante la versión popular.
C) La versión popular obtendrá una partición final que no podrá mejorarse mediante la versión DH.
D) La partición final obtenida mediante la versión DH podrá mejorarse mediante la versión popular, y viceversa.

11 $\boxed{A}$ Dado el modelo de Markov M_A de la pregunta 12, la aproximación de Viterbi a la probabilidad exacta que este modelo asigna a la cadena "`bba`" es:

A) 0.003200. $\tilde{P}(bba, q_1q_2q_3 = 111 \mid M_A) = 0.2 \cdot 0.8 \cdot 0.5 \cdot 0.8 \cdot 0.5 \cdot 0.2 \cdot 0.5 = 0.0032$

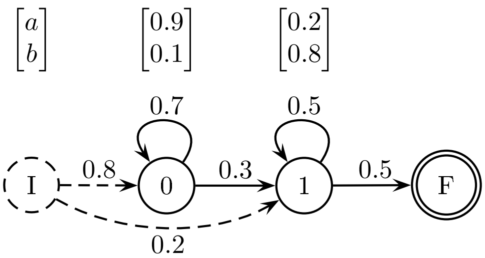
B) 0.004328.

C) 0.006400.

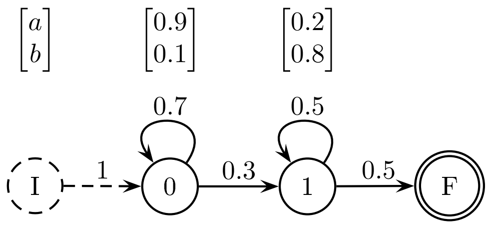
D) Ninguno de los resultados anteriores es correcto.

12 $\boxed{B}$ Se tiene un problema de clasificación en dos clases equiprobables (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 de Markov:

Modelo M_A: $P(x \mid A) = P(x \mid M_A)$



Modelo M_B: $P(x \mid B) = P(x \mid M_B)$



Por mínima probabilidad de error, la cadena "`bba`" quedaría clasificada en la clase:

A) Indistintamente en A ó B ya que las clases son equiprobables.

B) En la clase A. $\hat{c} = \arg\max_c P(c \mid \text{"}bba\text{"}) = \arg\max_c P(c)P(\text{"}bba\text{"} \mid c) = \arg\max_c P(\text{"}bba\text{"} \mid c)$

C) En la clase B. $P(\text{"}bba\text{"} \mid A) \approx \tilde{P}(\text{"}bba\text{"}|A) = 0.0032 \gg P(\text{"}bba\text{"} \mid B) \approx \tilde{P}(\text{"}bba\text{"}|B) = 0.0012 \rightarrow \hat{c} = A$

D) No se puede determinar ya que M_B no cumple las condiciones de normalización.

13 $\boxed{C}$ Dado el modelo de Markov M_A de la pregunta 12, si aplicamos el algoritmo *forward* con la cadena "`bba`", se cumple que:

A) $\alpha(q = 1, t = 3) = \alpha(q = 0, t = 2) \cdot A_{01} \cdot B_{1a}$.

B) $\alpha(q = 1, t = 3) = \alpha(q = 1, t = 2) \cdot A_{11} \cdot B_{1a}$.

C) $\alpha(q = 1, t = 3) = \alpha(q = 0, t = 2) \cdot A_{01} \cdot B_{1a} + \alpha(q = 1, t = 2) \cdot A_{11} \cdot B_{1a}$.

D) $\alpha(q = 1, t = 3) = \alpha(q = 0, t = 2) \cdot A_{01} \cdot B_{1a} \cdot \alpha(q = 1, t = 2) \cdot A_{11} \cdot B_{1a}$.

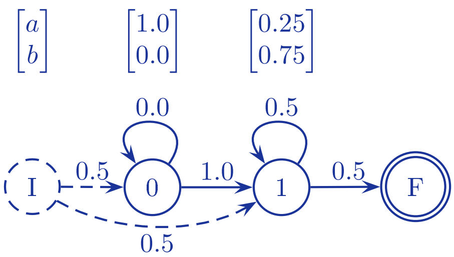
14 $\boxed{D}$ Dado el modelo de Markov M_A de la pregunta 12, tras *una* iteración de re-estimación por Viterbi a partir de las cadenas de entrenamiento "`bba`" y "`ab`" se cumple que:

A) $\pi_0 = 1$.

B) No se produce ningún cambio en el modelo.

C) Todas las probabilidades de transición modifican su valor.

D) El estado 0 tiene algunas probabilidades de emisión y/o transición nulas.



15 $\boxed{B}$ El modelo de Markov de conjunto de estados $Q = \{0, 1, F\}$ y alfabeto $\Sigma = \{a, b\}$ estimado mediante una inicialización con una segmentación lineal a partir de las cadenas de entrenamiento "`bbaa`" y "`ab`":

A) Tiene algunas probabilidades de emisión nulas.

B) Cumple que $A_{00} = A_{11}$ y $A_{01} = A_{1F}$.

C) Cumple que $\pi_0 = \pi_1$.

D) Cumple que $B_{0a} = B_{1a}$.