Examen del Bloc 2 de Sistemes Intel·ligents

ETSINF, Universitat Politècnica de València, 18 de gener de 2017

Cognoms: | | **Nom:** | |

Grup: ☐ **3A** ☐ **3B** ☐ **3C** ☐ **3D** ☐ **3E** ☐ **3F** ☐ **3FLIP**

Marca cada requadre amb una única opció d'entre les donades.

1 $\boxed{\text{C}}$ Quina de les següents expressions és *incorrecta*?

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

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

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

D) $P(x \mid y) \;=\; \dfrac{P(y \mid x)\; P(x)}{P(y)}$

2 $\boxed{\text{B}}$ Es tenen dues borses. La primera conté 3 pomes de color roig i 5 de color verd; la segona, 2 roges, 2 verdes i 1 groga. S'escull una borsa a l'atzar i, seguidament, una poma a l'atzar de la mateixa. Suposeu que les borses tenen la mateixa probabilitat de ser escollides i que, donada una borsa qualsevol, les seues pomes també tenen idèntica probabilitat de ser escollides. Si la poma escollida és roja, quina és la probabilitat P que siga de la primera borsa?

A) $0.00 \leq P < 0.25$

B) $0.25 \leq P < 0.50$

C) $0.50 \leq P < 0.75$

D) $0.75 \leq P$

$$P = P(B=1 \mid C=r) = \frac{P(B=1)P(C=r|B=1)}{P(C=r)}$$
$$= \frac{P(B=1)P(C=r|B=1)}{P(B=1)P(C=r|B=1)+P(B=2)P(C=r|B=2)}$$
$$= \frac{\frac{1}{2}\cdot\frac{3}{8}}{\frac{1}{2}\cdot\frac{3}{8}+\frac{1}{2}\cdot\frac{2}{5}} = \frac{15}{31} = 0.4839$$

3 $\boxed{\text{A}}$ Siga x un objecte (vector de característiques o cadena de símbols) a classificar en una classe de C possibles. Indica quin dels següents classificadors *no* és d'error mínim:

A) $c(x) = \underset{c=1,\ldots,C}{\arg\max}\; p(x \mid c)$

B) $c(x) = \underset{c=1,\ldots,C}{\arg\max}\; p(x, c)$

C) $c(x) = \underset{c=1,\ldots,C}{\arg\max}\; \log p(x, c)$

D) $c(x) = \underset{c=1,\ldots,C}{\arg\max}\; P(c \mid x)$

4 $\boxed{\text{A}}$ Per a un problema de classificació de dues classes en $\mathbb{R}^2$ tenim un classificador compost per dues funcions discriminants lineals de vectors de pesos $\mathbf{a}_\circ = (-1,1,2)^t$ i $\mathbf{a}_\bullet = (1,1,1)^t$, en notació homogènia. Indica quines són les regions de decisió definides pel classificador anterior.

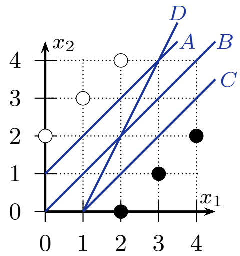
A) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_2 > 2\}$ i $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_2 < 2\}$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \rightarrow x_2 = 2 \;\wedge\; g_\circ((0,0)^t) < g_\bullet((0,0)^t)$

B) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_1 > 2\}$ i $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_1 < 2\}$

C) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_1 < 2\}$ i $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_1 > 2\}$

D) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_2 < 2\}$ i $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_2 > 2\}$

5 $\boxed{\text{B}}$ En la figura de la dreta es representen 6 mostres d'aprenentatge bidimensionals de 2 classes. Després d'aplicar l'algorisme Perceptró amb diferents valors del paràmetre b, s'obtenen els classificadors caracteritzats pels vectors de pesos donats a continuació. Quin d'ells proporciona una frontera de decisió de major folgança (més "centrada"), i per tant de menor error esperat?



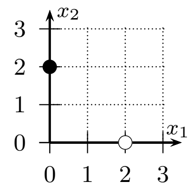
A) $\mathbf{a}_\circ = (-1, 1, 2)^t$ i $\mathbf{a}_\bullet = (\ \ 0, 2, 1)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \to x_2 = x_1 + 1$

B) $\mathbf{a}_\circ = (\ \ 1, 1, 2)^t$ i $\mathbf{a}_\bullet = (\ \ 1, 2, 1)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \to x_2 = x_1$

C) $\mathbf{a}_\circ = (\ \ 1, 1, 2)^t$ i $\mathbf{a}_\bullet = (\ \ 0, 2, 1)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \to x_2 = x_1 - 1$

D) $\mathbf{a}_\circ = (\ \ 1, 1, 1)^t$ i $\mathbf{a}_\bullet = (-1, 3, 0)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \to x_2 = 2x_1 - 2$

6 $\boxed{\text{C}}$ En la figura de la dreta es representen dues mostres d'aprenentatge bidimensionals de 2 classes: $(\mathbf{x}_1,\circ)$ i $(\mathbf{x}_2,\bullet)$. Donats els pesos $\mathbf{a}_\circ = (0, 1, 0)^t$ i $\mathbf{a}_\bullet = (1, 0, 0)^t$, si apliquem una iteració de l'algorisme Perceptró processant únicament la mostra $\mathbf{x}_1$, quin és el valor mínim del marge b amb el qual s'actualitzen els vectors de pesos?



A) $b = 0.5$

B) $b = 1.0$

C) $b = 1.5$ $\quad g_\circ(\mathbf{x}_1) = 2 \quad g_\bullet(\mathbf{x}_1) = 1 \quad \text{if } (g_\bullet(\mathbf{x}_1) + b > g_\circ(\mathbf{x}_1))$

D) Cap dels anteriors

7 $\boxed{\text{D}}$ Es té un problema de classificació per al qual s'ha après un classificador. L'interval de confiança al 95% per a la probabilitat d'error d'aquest classificador s'ha estimat empíricament, a partir d'un cert conjunt de mostres de test. Indica quina de les següents opcions permetria reduir la grandària de l'interval estimat:

A) Reduir significativament el conjunt de test.

B) Mantenir el conjunt de test i re-entrenar el classificador amb l'algorisme C-mitjanes de Duda i Hart.

C) Mantenir el conjunt de test i re-entrenar el classificador amb l'algorisme C-mitjanes convencional ("popular").

D) Augmentar significativament el conjunt de test.

8 $\boxed{\text{A}}$ Siga un problema de classificació en 3 classes (A, B i C) per al qual es disposa de 6 dades d'aprenentatge representades mitjançant vectors de característiques tridimensionals (veure taula a la dreta). Si desitgem aplicar l'algorisme d'aprenentatge d'arbres de classificació amb aquestes dades, quin és el nombre N de particions diferents que es poden generar en el node arrel de l'arbre? Nota: no heu de tenir en compte les particions que donen lloc a nodes buits.

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	2	0	3	2	3
c_n	A	A	B	B	C	C

A) $0 \leq N \leq 5$ $\quad \{(1,0),(2,1),(2,2),(3,0),(3,2)\}$

B) $5 < N \leq 10$

C) $10 < N \leq 20$

D) Es poden generar infinites particions.

9 $\boxed{\text{D}}$ Suposeu que estem aplicant l'algorisme d'aprenentatge d'arbres de classificació per a un problema de quatre classes; açò és, $\mathcal{C} = \{1, 2, 3, 4\}$. L'algorisme ha arribat a un node t que inclou vuit dades: 4 de la classe 1, 2 de la 2, 1 de la 3 i 1 de la 4. La impuresa de t, $\mathcal{I}(t)$, mesurada com l'entropia de la distribució empírica de les probabilitats a posteriori de les classes en t, és:

A) $0.00 \leq \mathcal{I}(t) < 0.25$

B) $0.25 \leq \mathcal{I}(t) < 0.50$

C) $0.50 \leq \mathcal{I}(t) < 0.75$

D) $0.75 \leq \mathcal{I}(t)$ $\quad \mathcal{I}(t) = -\sum_{c=1}^{4} \hat{P}(c \mid t) \log_2 \hat{P}(c \mid t) = -\frac{4}{8}\log_2\frac{4}{8} - \frac{2}{8}\log_2\frac{2}{8} - 2\,\frac{1}{8}\log_2\frac{1}{8} = \frac{7}{4} = 1.75$

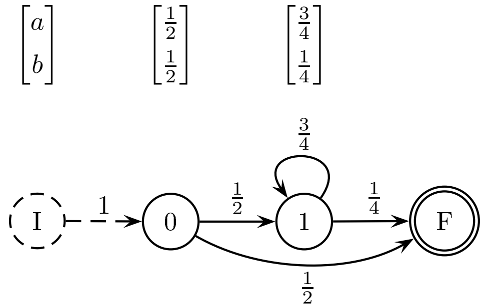
10 $\boxed{\text{B}}$ Indica quina de les següents afirmacions sobre aprenentatge supervisat (AS) i no-supervisat (ANS) és correcta:

A) Tant en ANS com en AS es requereixen dades d'entrenament sense etiqueta de classe.
B) En ANS es requereixen dades d'entrenament sense etiqueta de classe; en AS, amb etiqueta.
C) En ANS es requereixen dades d'entrenament amb etiqueta de classe; en AS, sense etiqueta.
D) Tant en ANS com en AS es requereixen dades d'entrenament amb etiqueta de classe.

11 $\boxed{\text{D}}$ Considereu l'algorisme C-mitjanes en la seua versió convencional o "popular" (CM) i en la versió de Duda i Hart (DH). Encara que ambdues optimitzen la *suma d'errors quadràtics* (SEQ), els seus resultats poden diferir ja que:

A) DH mimimiza la SEQ i CM la maximitza.
B) DH maximitza la SEQ i CM la minimitza.
C) Ambdues maximitzen la SEQ, si bé DH pot aconseguir millors solucions que CM.
D) Cap de les anteriors.

12 $\boxed{\text{C}}$ Considereu el model de Markov ocult M que es mostra en la figura de la dreta, en el qual $P_M(a) = P_M(b) = \frac{1}{4}$. Quin és el valor $S = \sum_x P_M(x)$ on x és qualsevol possible cadena formada per dos o més símbols?

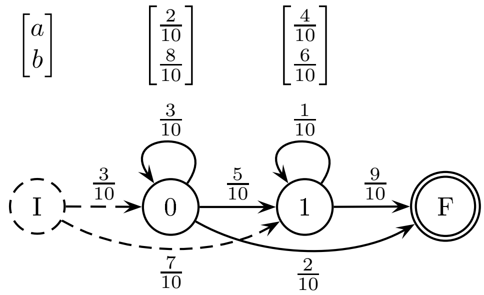


A) $0 \leq S < \frac{1}{4}$.
B) $\frac{1}{4} \leq S < \frac{2}{4}$.
C) $\frac{2}{4} \leq S < \frac{3}{4}$.
D) $\frac{3}{4} \leq S \leq 1$.

13 $\boxed{\text{B}}$ Siga M un model de Markov ocult i x una cadena tal que $P_M(x) > 0$. Sempre es compleix que:

A) La seqüència d'estats de M que genera la cadena x amb màxima probabilitat és única.
B) L'aproximació de Viterbi a $P_M(x)$ és única.
C) La seqüència d'estats de M que genera la cadena x amb màxima probabilitat no és única.
D) L'aproximació de Viterbi a $P_M(x)$ no és única.

14 $\boxed{\text{D}}$ Es té un problema de classificació en dues classes (A i B) d'objectes representats mitjançant cadenes de símbols en l'alfabet $\Sigma = \{a, b\}$. Les funcions de probabilitat condicional de les classes vénen caracteritzades pels models de Markov ocults M_A i M_B. Suposeu que $P(A) = 0.45$, $P(ba \mid A) = P_{M_A}(ba) = 0.0612$ i $P(ba \mid B) = P_{M_B}(ba)$, sent M_B el model representat en la figura de la dreta. A quina classe s'assignaria la cadena "`ba`" per mínima probabilitat d'error?:



A) Amb les dades aportades no es pot determinar.
B) Indistintament en A o B ja que $P_{M_A}(ba) = P_{M_B}(ba)$.
C) En la classe A.
D) En la classe B.

$\hat{c} = \arg\max_c P(c)P(ba \mid c)$
$P(A)P(ba \mid A) = 0.45 \cdot 0.0612$
$P(B)P(ba \mid B) = 0.55 \cdot 0.0612$
$\hat{c} = B$

15 $\boxed{\text{A}}$ Donat el model de Markov ocult M_B de la pregunta anterior, després *d'una* iteració de re-estimació per Viterbi a partir de les cadenes d'entrenament "`ba`", "`b`" i "`aa`", indica quin dels següents resultats és cert:

A) $A_{01} = A_{1F} = 1$
B) $B_{0a} = B_{1a} = \frac{1}{2}$
C) $\pi_0 = \frac{1}{3}$
D) $\pi_1 = \frac{2}{3}$

A	0	1	F
0	0	1	0
1	0	0	1

B	a	b
0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{2}{3}$	$\frac{1}{3}$

$\pi_0 = \frac{2}{3}$ $\pi_1 = \frac{1}{3}$