# Intelligent Systems: Final Exam Block 2

ETSINF, Universitat Politècnica de València, January 20th, 2016

**Surname(s):** ☐ **Name:** ☐

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

**Tick only one choice among the given options.**

1 $\boxed{\text{D}}$ Which of the following assertions is **TRUE**?

A) $P(x,y) \;=\; \sum_z P(x)\, P(y)\, P(z).$

B) $P(x,y) \;=\; \sum_z P(x)\, P(y \mid z).$

C) $P(x,y) \;=\; \sum_z P(x \mid z)\, P(y \mid z)\, P(z).$

D) $P(x,y) \;=\; \sum_z P(x, y \mid z)\, P(z).$ $P(x,y) = \sum_z P(x,y,z) = \sum_z P(x,y \mid z)\, P(z)$

2 $\boxed{\text{A}}$ An entomologist discovers a rare subspecies of beetle, due to the pattern of his back. In this rare subspecies, 98 % of the specimen have this pattern. In the common subspecies, 5 % of the specimen have this pattern. The rare subspecies represents 0.1 % of the population. The probability $P$ that a beetle with the pattern of his back belongs to the rare subspecies is:

A) $0.00 \leq P < 0.05$. $P = P(r \mid p) = \frac{P(r)\,P(p|r)}{P(p)} = \frac{P(r)\,P(p|r)}{P(r)\,P(p|r)+P(c)\,P(p|c)} = \frac{1/1000 \cdot 98/100}{1/1000 \cdot 98/100 + 999/1000 \cdot 5/100} = \frac{98}{5093} = 0.0192$

B) $0.05 \leq P < 0.10$.

C) $0.10 \leq P < 0.20$.

D) $0.20 \leq P$.

3 $\boxed{\text{C}}$ Let $x$ be an object (represented with a feature vector or string of symbols) that we want to classify in one among $C$ possible classes. Indicate which of the following expressions **DOES NOT** classify $x$ by minimum classification error:

A) $c(x) \;=\; \underset{c=1,\ldots,C}{\arg\max}\ \log_2 p(c \mid x)$

B) $c(x) \;=\; \underset{c=1,\ldots,C}{\arg\max}\ \log_{10} p(c \mid x)$

C) $c(x) \;=\; \underset{c=1,\ldots,C}{\arg\max}\ a\, p(c \mid x) + b$ being $a$ and $b$ two real constants

D) $c(x) \;=\; \underset{c=1,\ldots,C}{\arg\max}\ p(c \mid x)^3$

4 $\boxed{\text{C}}$ We have three different classifiers for a two-class problem in $\Re^2$. One classifier is formed by the linear functions: $g_1(y) = 2\,y_1 + y_2 + 3$ and $g_2(y) = y_1 + 2$. The second classifier is formed by: $g_1'(y) = -2\,y_1 + y_2 - 1$ and $g_2'(y) = -y_1 + 2\,y_2$. The third classifier is formed by: $g_1''(y) = -2\,y_1 - y_2 - 3$ and $g_2''(y) = -y_1 - 2$. Which assertion is TRUE?

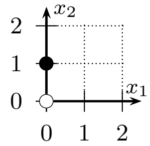
A) $(g_1, g_2)$ y $(g_1', g_2')$ are equivalent, but $(g_1, g_2)$ y $(g_1'', g_2'')$ are not.

B) $(g_1, g_2)$ y $(g_1', g_2')$ are not equivalent, but $(g_1, g_2)$ y $(g_1'', g_2'')$ are equivalent.

C) $(g_1, g_2)$ y $(g_1', g_2')$ are not equivalent, but $(g_1', g_2')$ y $(g_1'', g_2'')$ are equivalent. Common boundary $y_2 = -y_1 - 1$ but $R \neq R' = R''$

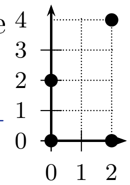
D) The three classifiers are not equivalent to each other.

5 $\boxed{C}$ The figure on the right shows two bi-dimensional samples in 2 classes: ($x_1$,$\circ$) and ($x_2$,$\bullet$). Given the weight vectors $\mathbf{a}_\circ = (0, 1, -2)^t$ and $\mathbf{a}_\bullet = (0, 0, 1)^t$, if we apply the Perceptron algorithm only to the sample $x_1$, we obtain the new weight vectors $\mathbf{a}_\circ = (1, 1, -2)^t$ and $\mathbf{a}_\bullet = (-1, 0, 1)^t$. Which is the value of the learning factor $\alpha$ and margin $b$?




A) $\alpha = 1.0$ y $b = 0.0$
B) $\alpha = -1.0$ y $b = 0.5$
C) $\alpha = 1.0$ y $b = 0.5$
D) It is not possible to determine the value of $\alpha$ and $b$

6 $\boxed{A}$ Consider the partition $\Pi = \{X_1 = \{(0, 0)^t, (0, 2)^t\}, X_2 = \{(2, 0)^t, (2, 4)^t\}\}$ for the points in the figure. The mean points of the clusters are $\mathbf{m}_1 = (0, 1)^t$ and $\mathbf{m}_2 = (2, 2)^t$. The Sum of Square Errors (SSE) of the partition is 10. If the point $(0, 2)^t$ is transferred to cluster $X_2$, then:




A) The new SSE value will be >10. $\|(0, 2)^t - (4/3, 2)^t\|^2 + \|(2, 0)^t - (4/3, 2)^t\|^2 + \|(2, 4)^t - (4/3, 2)^t\|^2 = 32/3$
B) The new SSE value will be >8 and <10
C) The new SSE value will be >6 and <8
D) The new SSE value will be <6.

# Intelligent Systems: Final Exam Block 2

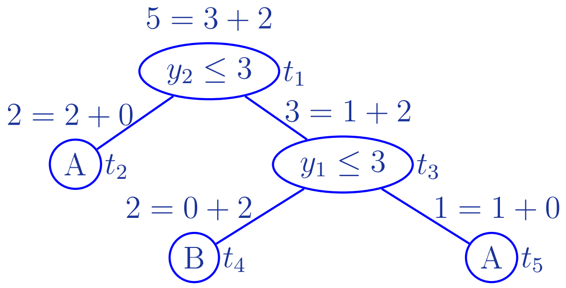
ETSINF, Universitat Politècnica de València, January 20th, 2016

## Problems (3 points; estimated time: 45 minutes)

1. **(1 point)** We have the 5 two-dimensional samples shown in the table to learn a classification tree. For each sample, we show its feature vector and the class it belongs to. The first *split* is $(2, 3)$; that is, $y_2 \leq 3$; and the second and last split is $(1, 3)$; that is, $y_1 \leq 3$.

| | | | | | |
|---|---|---|---|---|---|
| $y_1$ | 2 | 2 | 2 | 4 | 6 |
| $y_2$ | 2 | 4 | 6 | 6 | 2 |
| $c$ | $A$ | $B$ | $B$ | $A$ | $A$ |

*a*) Represent graphically the classification tree and classify the sample $(4, 4)^t$




The sample $(4, 4)^t$ goes through the tree until it reaches $t_5$. Therefore, the classification hypothesis is class $A$.

*b*) For each non-terminal node, $t$, calculate:

- Probability of the classes, $P(c \mid t),\ c \in \{A, B\}$
  $P(A \mid t_1) = 3/5, \quad P(B \mid t_1) = 2/5; \qquad P(A \mid t_3) = 1/3, \quad P(B \mid t_3) = 2/3$

- Probability of choosing the left node and the right node, $P_t(L),\ P_t(R)$
  $P_{t_1}(L) = 2/5, \quad P_{t_1}(R) = 3/5 \qquad P_{t_3}(L) = 2/3, \quad P_{t_3}(R) = 1/3$

*c*) Calculate the number of bits of the impurity, $\mathcal{I}(t_1)$, of the root node, $t_1$

$$\begin{aligned}\mathcal{I}(t_1) &= -P(A \mid t_1)\, log_2 P(A \mid t_1) - P(B \mid t_1)\, log_2 P(B \mid t_1) \\ &\approx -0.6(-0.737)\ -0.4(-1.322) \ = \ 0.971 \text{ bits.}\end{aligned}$$

*d*) For each terminal node, $t$, calculate:

- Probability of the terminal node, $P(t)$
  $P(t_2) = 2/5, \quad P(t_4) = 2/5, \quad P(t_5) = 1/5$

- Impurity in bits, $\mathcal{I}(t)$
  $\mathcal{I}(t_2) = \mathcal{I}(t_4) = \mathcal{I}(t_5) = 0$ bits.

*e*) Estimated resubstitution error (misclassification error) of the tree.
Since the three terminal nodes are pure nodes, the estimated resubstitution error is 0.

2. **(2 points)** Let $M$ be a Markov model with states $Q = \{1, 2, F\}$; alphabet $\Sigma = \{a, b, c\}$; initial probabilities $\pi_1 = \frac{6}{10}, \pi_2 = \frac{4}{10}$; and transition and emission probabilities:

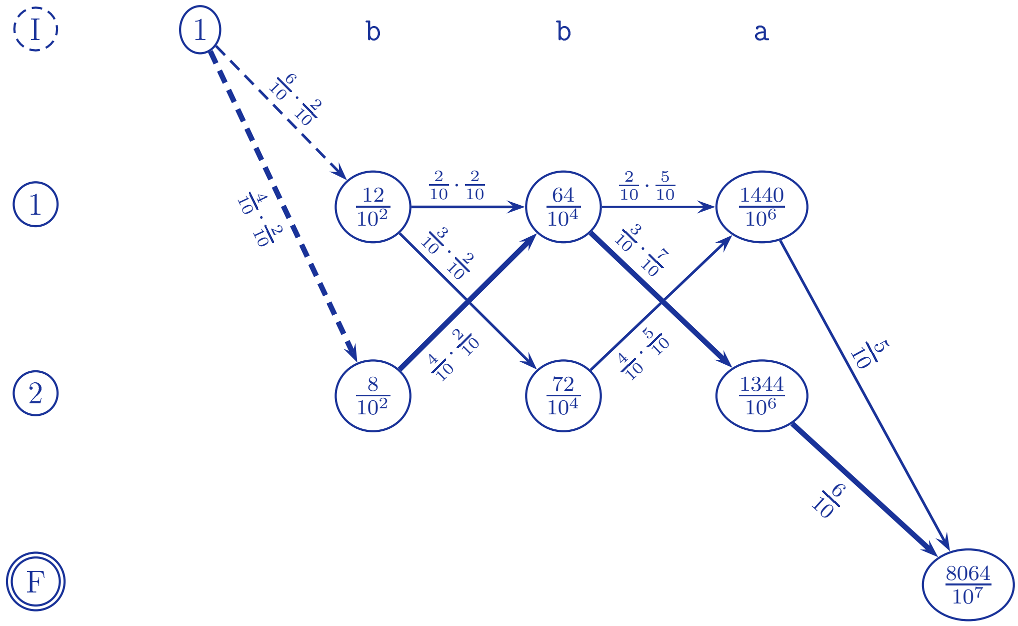
| $A$ | 1 | 2 | $F$ |
|---|---|---|---|
| 1 | $\frac{2}{10}$ | $\frac{3}{10}$ | $\frac{5}{10}$ |
| 2 | $\frac{4}{10}$ | 0 | $\frac{6}{10}$ |

| $B$ | $a$ | $b$ | $c$ |
|---|---|---|---|
| 1 | $\frac{5}{10}$ | $\frac{2}{10}$ | $\frac{3}{10}$ |
| 2 | $\frac{7}{10}$ | $\frac{2}{10}$ | $\frac{1}{10}$ |

*a*) Apply the *Viterbi* algorithm in model $M$ to obtain the most probable state sequence for the string "bba".

*b*) Calculate the model $M'$ after ONE iteration of Viterbi re-estimation algorithm using the string in the previous question ("bba") and the strings "ac","cacb" and "a". For this calculation, use the following data: $\tilde{P}(ac \mid M) = P(ac, q_1 q_2 = 21 \mid M)$ (i.e.; the optimal state sequence for "ac" is 21); $\tilde{P}(cacb \mid M) = P(cacb, q_1 q_2 q_3 q_4 = 1212 \mid M)$ and $\tilde{P}(a \mid M) = P(ac, q_1 = 2 \mid M)$.

*a*)




$$\tilde{Q} = (2, 1, 2, F)$$

*b*)

$$\pi_1 = \frac{1}{4}, \pi_2 = \frac{3}{4}$$

| $A$ | 1 | 2 | $F$ |
|---|---|---|---|
| 1 | 0 | $\frac{3}{4}$ | $\frac{1}{4}$ |
| 2 | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ |

| $B$ | $a$ | $b$ | $c$ |
|---|---|---|---|
| 1 | 0 | $\frac{1}{4}$ | $\frac{3}{4}$ |
| 2 | $\frac{2}{3}$ | $\frac{1}{3}$ | 0 |