Intelligent Systems – Final Exam (Block 2): Test (2 points)

ETSINF, Universitat Politècnica de València, January 17th, 2018

Surname(s): | **Name:**

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

Tick only one choice amongst the given options. Score: max(0, (#correct_answers - #wrong_answers/3) / 3).

1 D Among the following assertions about Artificial Intelligence (AI) and Machine Learning (ML), mark the one that **IS NOT CORRECT**:

A) One major limitation in practice of classical AI is the impossibility to check all logical conditions that should be met for an action to be applicable. For example, it turns out almost impossible to know and check every logical condition to guarantee that "we get to Manises airport on time if we leave home 90 minutes before the flight".

B) Current intelligent systems usually include *uncertainty* in the knowledge representation, which can be represented through *probabilities* associated to the events of study.

C) Most of ML-based methods build hypothesis from the data.

D) The usual methods in ML are linear classifiers and *non-linear* classifiers.

2 D Let be a classification problem among 4 equiprobable classes, $c = 1, 2, 3, 4$. For a given object, x, we know that the Bayes classifier assigns class 1 to x and that the posterior probability of belonging to that class, $p(c = 1 \mid x)$, is 1/3. Given this information, show the **CORRECT** statement:

A) The probability of error of classifying x is lower than 1/3.

B) $p(c = 1 \mid x) \; > \; p(c = 2 \mid x) + p(c = 3 \mid x) + p(c = 4 \mid x)$.

C) $p(x) > p(x \mid c = 1)$.

D) None of the above.

3 B Let be a 3-class classification problem, $c = 1, 2, 3$, for two-dimensional objects, $\mathbf{x} = (x_1, x_2)^t \in \mathbb{R}^2$. We have a linear classifier with the following weight vectors: $\mathbf{w}_1 = (w_{10}, w_{11}, w_{12})^t = (2, 0, 0)^t$, $\mathbf{w}_2 = (0, 1, 1)^t$ and $\mathbf{w}_3 = (0, 1, -1)^t$. The decision region of the class 1 defined by this classifier is:

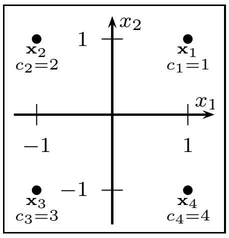
A) $\{\mathbf{x} : x_1 \geq 0 \wedge x_2 < -x_1 + 2\} \cup \{\mathbf{x} : x_1 < 0 \wedge x_2 < x_1 + 2\}$.

B) $\{\mathbf{x} : x_2 \geq 0 \wedge x_2 < -x_1 + 2\} \cup \{\mathbf{x} : x_2 < 0 \wedge x_2 > x_1 - 2\}$.

C) $\{\mathbf{x} : x_1 \geq 0 \wedge x_2 < -x_1 + 1\} \cup \{\mathbf{x} : x_1 < 0 \wedge x_2 < x_1 + 1\}$.

D) $\{\mathbf{x} : x_2 \geq 0 \wedge x_2 < -x_1 + 1\} \cup \{\mathbf{x} : x_2 < 0 \wedge x_2 > x_1 - 1\}$.

4 D The figure shows 4 samples, each belonging to one different class among 4 classes: $\mathbf{x}_1 = (1, 1)^t$ belongs to class $c_1 = 1$, $\mathbf{x}_2 = (-1, 1)^t$ belongs to class $c_2 = 2$, $\mathbf{x}_3 = (-1, -1)^t$ belongs to class $c_3 = 3$, and $\mathbf{x}_4 = (1, -1)^t$ belongs to class $c_4 = 4$. Let's assume we apply the Perceptron algorithm to these samples with learning rate $\alpha = 1$, margin $b = 0.1$ and initial null weight vectors. Once the *first 3 samples* have been processed at the first iteration of the algorithm, we get the weight vectors $\mathbf{w}_1 = (w_{10}, w_{11}, w_{12})^t = (0, 2, 0)^t$, $\mathbf{w}_2 = (-1, -1, 1)^t$, $\mathbf{w}_3 = (-1, -1, -3)^t$ and $\mathbf{w}_4 = (-3, 1, -1)^t$. Finish the first iteration of the Perceptron and mark, from the resulting weight vectors, the number of samples that are **CORRECTLY** classified:



A) 1.

B) 2. $\mathbf{w}_1 = (-1, 1, 1)^t$, $\mathbf{w}_2 = (-1, -1, 1)^t$, $\mathbf{w}_3 = (-2, -2, -2)^t$, $\mathbf{w}_4 = (-2, 2, -2)^t$

C) 3. $\mathbf{x}_1$: $g_1 = 1$, $g_2 = -1$, $g_3 = -6$, $g_4 = -2$ $\quad$ $\mathbf{x}_2$: $g_1 = -1$, $g_2 = 1$, $g_3 = -2$, $g_4 = -6$

D) 4. $\mathbf{x}_3$: $g_1 = -3$, $g_2 = -1$, $g_3 = 2$, $g_4 = -2$ $\quad$ $\mathbf{x}_4$: $g_1 = -1$, $g_2 = -3$, $g_3 = -2$, $g_4 = 2$

5 B We want to apply the classification tree learning algorithm to a 4-class problem, $c = 1, 2, 3, 4$. The algorithm reaches a node t that comprises 8 data: 2 data belong to class 1, 4 to class 2, 1 to class 3 and 1 to class 4. The entropy impurity, $\mathcal{I}(t)$, measured as the impurity of the posterior probability of the classes in t is:

A) $0.00 \leq \mathcal{I}(t) < 1.00$

B) $1.00 \leq \mathcal{I}(t) < 2.00$

C) $2.00 \leq \mathcal{I}(t) < 3.00$

D) $3.00 \leq \mathcal{I}(t)$

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

6 D The table on the right shows 6 three-dimensional training samples. We believe that a natural partition of this set in 2 clusters would group together the first 4 samples in one cluster and the last 2 samples in the second cluster. The Sum of Square Errors (SSE) of this partition, J, is:

A) $J < 3$

B) $3 \leq J < 6$

C) $6 \leq J < 12$

D) $12 \leq J$ $\quad (1 + 2 + 1 + 2) + (3 + 3) = 6 + 6 = 12$

$\mathbf{x}_n = (x_{n1}, x_{n2}, x_{n3})^t$

n	x_{n1}	x_{n2}	x_{n3}
1	0	1	1
2	2	1	0
3	1	2	1
4	1	0	2
5	4	6	4
6	6	4	6

Intelligent Systems – Final Exam (Block 2): Problem (3 points)

ETSINF, Universitat Politècnica de València, January 17th 2018

Surname(s): ______________________ **Name:** ____________

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

Let M be a Markov model with set of states $Q = \{1, 2, F\}$; alphabet $\Sigma = \{a, b\}$; initial probabilities $\pi_1 = \frac{1}{2}, \pi_2 = \frac{1}{2}$; and transition probability matrix and emission probability matrix:

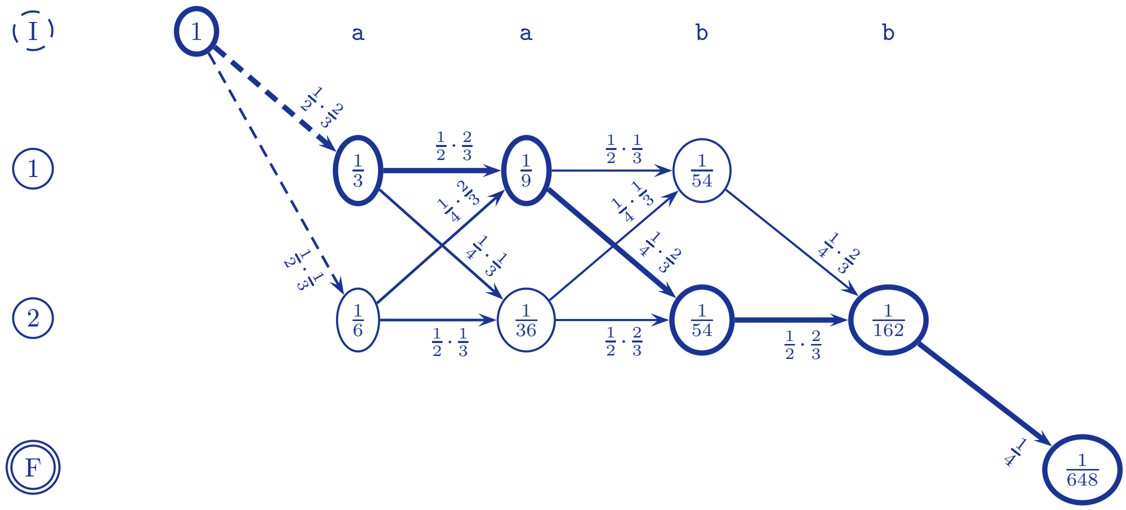
A	1	2	F
1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

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

1. (1.5 points) Apply the *Viterbi* algorithm to obtain the most probable sequence of states with which M generates the string "`aabb`".

2. (1.5 points) From the training samples "`aabb`" and "`a`", re-estimate the parameters of M by applying the Viterbi re-estimation algorithm (up to convergence).

1.



$$\tilde{q} = 1122F$$

2. En la primera iteración, a partir de los pares ("`aabb`",$1122F$) y ("`a`",$1F$), obtenemos:

$$\pi_1 = \frac{2}{2}, \pi_2 = 0$$

A	1	2	F
1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
2	0	$\frac{1}{2}$	$\frac{1}{2}$

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

At the second iteration, we start from a MM in which the symbol "`a`" is only emitted by state 1, and the symbol "`b`" is only emitted by state 2. Thus, "`aabb`" can only be generated with the sequence of states $1122F$, and"`a`" with $1F$. This means that the pairs (training sample, most probable sequence of states) are the same as in the first iteration. Consequently, the second iteration ends with the same MM as in the first iteration and the re-estimation algorithm stops (convergence).