Intelligent Systems - Final Exam (Block 2): Test (1.75 points)

ETSINF, Universitat Politècnica de València, January 27th, 2022

Group, surname(s) and name: 1,

Tick only one choice among the given options. Score: $\max(0, (\text{correct\_answers} - \text{wrong\_answers} / 3) \cdot 1.75 / 6)$.

1 $\boxed{A}$ Given the following 3 nodes of a classification tree with samples belonging to 3 classes:

c	1	2	3
n_1	2/12	5/12	5/12
n_2	3/11	4/11	4/11
n_3	5/11	3/11	3/11

where each row is the posterior probability of each class in the node. Which of the following inequalities is true?

A) $\mathcal{I}(n_1) < \mathcal{I}(n_3) < \mathcal{I}(n_2)$

B) $\mathcal{I}(n_3) < \mathcal{I}(n_2) < \mathcal{I}(n_1)$

C) $\mathcal{I}(n_1) < \mathcal{I}(n_2) < \mathcal{I}(n_3)$

D) $\mathcal{I}(n_2) < \mathcal{I}(n_3) < \mathcal{I}(n_1)$

2 $\boxed{D}$ Let M be a Markov model with set of states $Q = \{1, 2, F\}$ and alphabet $\Sigma = \{a, b\}$. Given the string $x = \texttt{bbb}$, the Viterbi aproximation to $P_M(x)$, $\tilde{P}_M(x)$, has been found using the Viterbi algorithm:

$$
\begin{aligned}
V_{11} &= \pi_1 B_{1b} = 0.3000\\
V_{21} &= \pi_2 B_{2b} = 0.3333\\
V_{12} &= \max(V_{11}A_{11}B_{1b}, V_{21}A_{21}B_{1b}) = \max(0.0450, 0.1000) = 0.1000\\
V_{22} &= \max(V_{11}A_{12}B_{2b}, V_{21}A_{22}B_{2b}) = \max(0.0500, 0.0556) = 0.0556\\
V_{13} &= \max(V_{12}A_{11}B_{1b}, V_{22}A_{21}B_{1b}) = \max(0.0150, 0.0167) = 0.0167\\
V_{23} &= \max(V_{12}A_{12}B_{2b}, V_{22}A_{22}B_{2b}) = \max(0.0167, 0.0093) = 0.0167\\
\tilde{P}(\texttt{bbb}) &= \max(V_{13}A_{1F}, V_{23}A_{2F}) = \max(0.0083, 0.0042) = 0.0083
\end{aligned}
$$

The most probable path (one of the most probable paths, if there are more than one) through which M generates x is:

A) 1 1 2 F

B) 2 1 1 F

C) 1 2 2 F

D) 2 2 1 F

3 $\boxed{C}$ For a three-class classification problem of objects of type $\mathbf{x} = (x_1, x_2)^t \in \{0, 1\}^2$, we have the probability distributions shown in the table. Show the interval of the probability of error of the classifier $c(\mathbf{x})$ provided in the table, ε:

A) $\varepsilon < 0.25$.

B) $0.25 \leq \varepsilon < 0.50$.

C) $0.50 \leq \varepsilon < 0.75$.

D) $0.75 \leq \varepsilon$.

$\mathbf{x}$		$P(c \mid \mathbf{x})$				
x_1	x_2	$c{=}1$	$c{=}2$	$c{=}3$	$P(\mathbf{x})$	$c(\mathbf{x})$
0	0	0.2	0.1	0.7	0.2	2
0	1	0.4	0.3	0.3	0	1
1	0	0.3	0.4	0.3	0.4	3
1	1	0.4	0.4	0.2	0.4	1

$\varepsilon = 0.70$

4 $\boxed{\text{C}}$ Given the following joint frecuency distribution for the 3 random variables

A	0	0	0	0	1	1	1	1
B	0	0	1	1	0	0	1	1
C	0	1	0	1	0	1	0	1
N(A,B,C)	124	28	227	175	126	222	23	75

Which is the value of $P(A = 1 \mid B = 1, C = 0)$?

A) 0.023
B) 0.250
C) 0.092
D) 0.446

5 $\boxed{\text{C}}$ Let M be a Markov model with set of states $Q = \{1, 2, F\}$ and alphabet $\Sigma = \{a, b\}$. After the application of an iteration of the Viterbi's reestimation algorithm, the table of state transition probabilities shown on the right has been obtained. From which table of state transition frequencies has been obtained?

A	1	2	F
1	$\frac{4}{9}$	$\frac{1}{9}$	$\frac{4}{9}$
2	$\frac{4}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

A)

A	1	2	F
1	4	1	15
2	4	1	15

B)

A	1	2	F
1	4	1	9
2	4	1	6

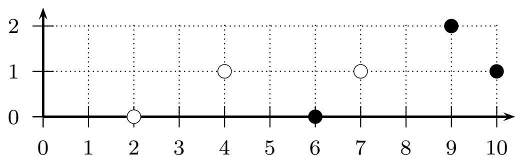
C)

A	1	2	F
1	8	2	8
2	12	3	3

D)

A	1	2	F
1	8	2	4
2	12	3	1

6 $\boxed{\text{D}}$ The figure below shows a partition of 6 two-dimensional points in 2 clusters, $\bullet$ and $\circ$:



If points $(10, 1)^t$ and $(7, 1)^t$ are transferred from one cluster to the other, a variation of the Sum of Square Errors (SSE) is produced, $\Delta J = J - J'$ (SSE after the transfer minus SSE before the transfer), such that:

A) $\Delta J < -7$.
B) $-7 \leq \Delta J < 0$.
C) $0 \leq \Delta J < 7$.
D) $\Delta J \geq 7$.

$\Delta J = 42.0 - 24.0 = 18.0$

Intelligent Systems - Final Exam (Block 2): Problem (2 points)

ETSINF, Universitat Politècnica de València, January 27th, 2022

Group, surname(s) and name: 1,

Problem: Perceptron algorithm

The table on the left side shows a training set of 3 two-dimension samples that belong to 3 classes, while the table on the right side shows the initial weight vector for each class.

n	x_{n1}	x_{n2}	c_n
1	-2	-2	1
2	0	0	2
3	2	2	3

	$\mathbf{w}_1$	$\mathbf{w}_2$	$\mathbf{w}_3$
w_{c0}	0	-1	-1
w_{c1}	-2	0	4
w_{c2}	-2	0	4

Answer the following questions:

1. (1.5 points) Show a trace of the Perceptron algorithm running for one iteration with learning rate $\alpha = 1$, margin $\beta = 0.1$ using the initial weight vectors provided.
2. (0.5 points) Plot the decision regions of the resulting classifier, as well as the decision boundaries involved.

Solution:

1. One iteration of the Perceptron algorithm with 1 misclassified sample provides the following final weight vectors:

	$\mathbf{w}_1$	$\mathbf{w}_2$	$\mathbf{w}_3$
w_{c0}	-1	0	-2
w_{c1}	-2	0	4
w_{c2}	-2	0	4

2. The following plot shows the three decision regions along with the two decision boundaries involved: