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

ETSINF, Universitat Politècnica de València, January 16th, 2024

Group, surname(s) and name: 2,

Tick only one choice among the given options. Score: $\max(0, (\text{correct\_answers} - \text{wrong\_answers} / 3) \cdot 1.75 / 9)$.

1 ☐ In a problem of probabilistic reasoning corresponding to flu diagnosis, the random variables of interest are: Flu (F):{positive (POS), negative (NEG)}; Ventilation (V):{high (HIG), low (LOW)}; Activity (A):{silence (SIL), talking (TAL), exercise (EXE)}. The joint probability of the three random variables is provided by the following table:

	HIG			LOW		
$P(f, v, a)$	SIL	TAL	EXE	SIL	TAL	EXE
POS	0.01	0.02	0.02	0.01	0.03	0.05
NEG	0.29	0.19	0.10	0.14	0.10	0.04

The conditional probability $P(G = \text{POS} \mid V = \text{LOW}, A = \text{EXE})$ is:

A) $P \leq 0.25$

B) $0.25 < P \leq 0.50$

C) $0.50 < P \leq 0.75$

D) $0.75 < P \leq 1.0$

2 ☐ Let $\mathbf{x}$ be a object that we want to classify in one among C classes. Which expression is *not* a minimum error (error risk) classifier (or choose the last option if the first three are minimum error classifiers)?

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

B) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ e^{p(c|\mathbf{x})} + e^{p(\mathbf{x})}$

C) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ e^{p(\mathbf{x},c)} - e^{p(\mathbf{x})}$

D) All three are minimum error classifiers.

3 ☐ 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	0.5	0.1	0.4	0.3	1
0	1	0.6	0.4	0	0.3	2
1	0	0.1	0.4	0.5	0.1	2
1	1	0	0.5	0.5	0.3	1

4 ☐ Let's suppose that we are applying the Perceptron algorithm, with learning rate $\alpha = 1$ and margin $b = 0.1$, to a set of 3 bidimensional learning samples for a problem of 2 classes. After processing the first 2 samples, the weight vectors $\mathbf{w}_1 = (0, 2, 1)^t$, $\mathbf{w}_2 = (0, -2, -1)^t$ were obtained. Next, the last sample $(\mathbf{x}_3, c_3)$ is processed and the following weight vectors $\mathbf{w}_1 = (-1, 1, -3)^t$, $\mathbf{w}_2 = (1, -1, 3)^t$ are obtained, which of the following samples is that last sample?

A) $((2, 1)^t, 1)$

B) $((3, 1)^t, 1)$

C) $((1, 4)^t, 2)$

D) $((2, 4)^t, 1)$

5 ☐ Given a classifier for 3 classes defined by their weight vectors $\mathbf{w}_1 = (2, 1, 1)^t$, $\mathbf{w}_2 = (1, -3, -3)^t$, $\mathbf{w}_3 = (2, 0, -1)^t$ in homogeneous notation, which of the following weight vectors do **not** define a classifier equivalent to the one given?

A) $\mathbf{w}_1 = (-2, -1, -1)^t$, $\mathbf{w}_2 = (-1, 3, 3)^t$, $\mathbf{w}_3 = (-2, 0, 1)^t$

B) $\mathbf{w}_1 = (4, 2, 2)^t$, $\mathbf{w}_2 = (2, -6, -6)^t$, $\mathbf{w}_3 = (4, 0, -2)^t$

C) $\mathbf{w}_1 = (4, 1, 1)^t$, $\mathbf{w}_2 = (3, -3, -3)^t$, $\mathbf{w}_3 = (4, 0, -1)^t$

D) $\mathbf{w}_1 = (6, 2, 2)^t$, $\mathbf{w}_2 = (4, -6, -6)^t$, $\mathbf{w}_3 = (6, 0, -2)^t$

6 ☐ Which of the following statements about logistic regression is *incorrect* (or choose the last option if the first three are correct)?:

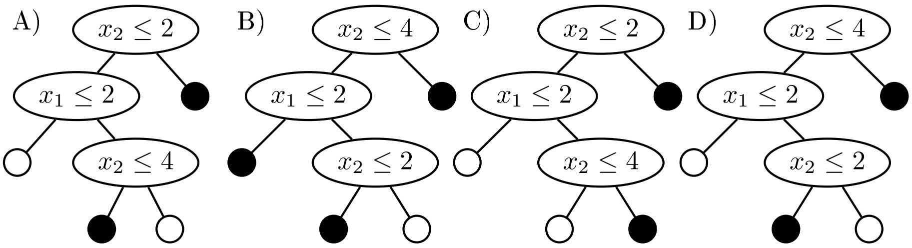
A) Logistic regression is a classification probabilistic model based on logit predictive function linearly dependent on the input

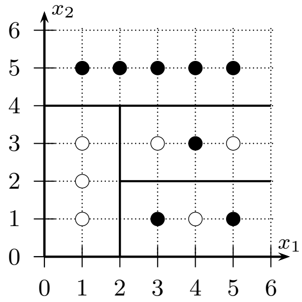
B) Being a a classification probabilistic model, logistic regression allows the application of decision rules more general than the MAP (maximum a posteriori) rule

C) Logistic regression is a classification probabilistic model based on the categorical distribution

D) All three previous statements are correct

7 Given the two-class (○ and ●) samples of the figure on the right, which of the following classification trees is coherent with the partition of the figure?





8 Suppose we apply the classification tree algorithm for a 3-class problem $c = 1, 2, 3$. The algorithm reaches a node t that is split into a left node with 3 samples of class 1, 1 sample of class 2 and 2 samples of class 3; and a right node with 1 sample of class 1, 0 samples of class 2 and 0 samples of class 3. Which impurity reduction is achieved with this split?

A) $0.00 \le \Delta\mathcal{I} < 0.25$.
B) $0.25 \le \Delta\mathcal{I} < 0.50$.
C) $0.50 \le \Delta\mathcal{I} < 0.75$.
D) $0.75 \le \Delta\mathcal{I}$.

9 We have a partition of a set of 3-dimensional data points into a given number of clusters, $C \ge 2$. Consider the transfer of the data point $\mathbf{x} = (4, 10, 4)^t$ from a cluster i to another one j, $j \ne i$. We know that cluster i contains 4 data points (including $\mathbf{x}$) and cluster j 2. We also know that the centroid (mean) of cluster i is $\mathbf{m}_i = (1, 8, 2)^t$, while that of cluster j is $\mathbf{m}_j = (10, 2, 10)^t$. If the transfer is carried out, an increase of the sum of square errors, ΔJ, will be produced such that:

A) $\Delta J < -70$
B) $-70 \le \Delta J < -30$
C) $-30 \le \Delta J < 0$
D) $\Delta J \ge 0$

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

ETSINF, Universitat Politècnica de València, January 16th, 2024

Group, surname(s) and name: 2,

Problem: Logistic regression

The following table shows a training set of 2 samples with 2 dimensions that belong to 2 classes:

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

In addition, the following table represents an initial weight matrix with the weights of each class per columns:

$\mathbf{w}_1$	$\mathbf{w}_2$
0.	0.
0.	0.
0.25	-0.25

Answer the following questions:

1. (0.5 points) Compute the vector of logits for each training sample.
2. (0.25 points) Apply the softmax function to the vector of logits for each training sample.
3. (0.25 points) Classify every training sample. In case of a tie, choose any class.
4. (0.5 points) Compute the gradient of the function NLL at the point of the initial weight matrix.
5. (0.5 points) Update the initial weight matrix applying gradient descent with learning rate $\eta = 1.0$.