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

ETSINF, Universitat Politècnica de València, January 24th, 2025

Group, surname(s) and name: 2,

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

1 $\boxed{\text{A}}$ Given the following probability distributions for the random variables:

	$P(A=0 \mid B, C)$				$P(B, C)$			
B	0	0	1	1	0	0	1	1
C	0	1	0	1	0	1	0	1
	0.049	0.431	0.022	0.842	0.038	0.292	0.462	0.208

Which is the value of $P(A=1, B=1 \mid C=1)$? $P(A=1, B=1 \mid C=1) = 0.066$

A) $P(A=1, B=1 \mid C=1) \leq 0.25$

B) $0.25 < P(A=1, B=1 \mid C=1) \leq 0.50$

C) $0.50 < P(A=1, B=1 \mid C=1) \leq 0.75$

D) $0.75 < P(A=1, B=1 \mid C=1) \leq 1.00$

2 $\boxed{\text{D}}$ For a four-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 Bayes probability of error, ε^*:

A) $\varepsilon^* < 0.40$.

B) $0.40 \leq \varepsilon^* < 0.45$.

C) $0.45 \leq \varepsilon^* < 0.50$.

D) $0.50 \leq \varepsilon^*$.

$\mathbf{x}$		$P(c \mid \mathbf{x})$			
x_1	x_2	$c=1$	$c=2$	$c=3$	$P(\mathbf{x})$
0	0	0.3	0.3	0.1	0.2
0	1	0.1	0.2	0.2	0.2
1	0	0.3	0.1	0.3	0.1
1	1	0.1	0.2	0.2	0.5

$\varepsilon^* = 0.56$

3 $\boxed{\text{A|B}}$ Let $\mathbf{x}$ be a object that we want to classify in one among C classes. Which expression is a minimum error classifier (or choose the last option if all three classifiers are of minimum error)?

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

B) $c(\mathbf{x}) = \underset{c=1,\dots,C}{\arg\max} \ \log p(c \mid \mathbf{x}) - \log p(\mathbf{x})$

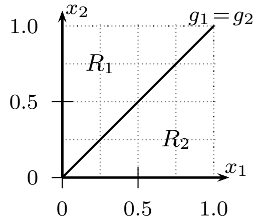
C) $c(\mathbf{x}) = \underset{c=1,\dots,C}{\arg\max} \ \frac{\log p(c \mid \mathbf{x})}{\log p(\mathbf{x})}$

D) All three classifiers are of minimum error.

4 $\boxed{\text{D}}$ Let's suppose that we are applying the Perceptron algorithm, to a set of 3 bidimensional learning samples for a problem of 2 classes. After processing the first 2 samples, the weight vectors $\mathbf{w}_1 = (0, -4, 1)^t$, $\mathbf{w}_2 = (0, 4, -1)^t$ were obtained. Next, the sample $(\mathbf{x}_3 = (1, 5),\ c_3 = 1)$ is processed, which of the following values of margin b is the minimum needed to update the weights with this sample?

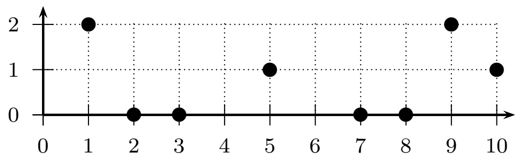
A) 0.0
B) 0.1
C) 1.0
D) 10.0

5 $\boxed{\text{C}}$ The figure on the right represents the decision boundary and the two regions of a binary classifier. Which of the following weight vectors (in homogeneous notation) defines a **non-equivalent** classifier to the one of the figure?



A) $\mathbf{w}_1 = (0, -1, 0)^t$ and $\mathbf{w}_2 = (0, 0, -1)^t$.
B) $\mathbf{w}_1 = (0, 0, 1)^t$ and $\mathbf{w}_2 = (0, 1, 0)^t$.
C) $\mathbf{w}_1 = (0, 1, 0)^t$ and $\mathbf{w}_2 = (0, 0, 1)^t$.
D) All the above weight vectors define non-equivalent classifiers to the one of the figure.

6 $\boxed{\text{D}}$ The figure below shows a dataset of 8 two-dimensional points:



What is the number of clusters that minimizes the sum of squared errors (SEC) of this dataset?

A) 1
B) 4
C) 5
D) 8

Intelligent Systems - Re-take Exam (Block 2): Problem (2 points)

ETSINF, Universitat Politècnica de València, January 24th, 2025

Group, surname(s) and name: 2,

Problem: Logistic regression

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

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

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.
-0.25	0.25
-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 $\eta = 1.0$.

Solution:

1. Vector of logits for each training sample:

n	a_{n1}	a_{n2}
1	0.	0.
2	-0.5	0.5

2. Applying the softmax function:

n	μ_{n1}	μ_{n2}
1	0.5	0.5
2	0.27	0.73

3. Classification of every sample:

n	$\hat{c}(x_n)$
1	1
2	2

4. Gradient:

$\mathbf{g}_1$	$\mathbf{g}_2$
-0.12	0.12
0.13	-0.13
0.13	-0.13

5. Updated weight matrix:

$\mathbf{w}_1$	$\mathbf{w}_2$
0.12	-0.12
-0.38	0.38
-0.38	0.38