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

ETSINF, Universitat Politècnica de València, December 19th, 2024

**Group, surname(s) and name:** 1,

Tick only one choice among the given options. Score: max(0, (correct_answers−wrong_answers / 3) · 1.75 / 9).

1 $\boxed{\text{C}}$ 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 $\mathbf{w}_1 = (0, 0, -2)^t$, $\mathbf{w}_2 = (0, 0, 2)^t$ were obtained. Next, the last sample $(\mathbf{x}_3, c_3)$ is processed and the same weight vectors are obtained, which of the following samples is that last sample?

A) $((5, 5)^t, 1)$
B) $((2, 4)^t, 1)$
C) $((2, 5)^t, 2)$
D) $((4, 1)^t, 1)$

2 $\boxed{\text{A}}$ Given the following conditional probability 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 |
| P(A, B \| C) | 0.449 | 0.173 | 0.051 | 0.327 | 0.343 | 0.027 | 0.157 | 0.473 |

If $P(C = 0) = 0.81$, which is the value of $P(A = 1 \mid B = 0, C = 1)$? $P(A = 1 \mid B = 0, C = 1) = 0.135$

A) $\text{P(A=1} \mid B = 0, C = 1) \leq 0.25$
B) $0.25 < \text{P(A=1} \mid B = 0, C = 1) \leq 0.50$
C) $0.50 < \text{P(A=1} \mid B = 0, C = 1) \leq 0.75$
D) $0.75 < \text{P(A=1} \mid B = 0, C = 1) \leq 1.00$

3 $\boxed{\text{C}}$ For a two-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 $\varepsilon$ of the classifier $c(\mathbf{x})$ based on the discriminant function $g(\mathbf{x}) = 0.5 + x_1 + x_2$ defined as

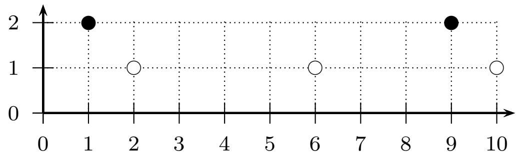
$$c(\mathbf{x}) = \begin{cases} 1 & \text{if } g(\mathbf{x}) < 0 \\ 2 & \text{otherwise} \end{cases}$$

| $\mathbf{x}$ | | $P(c \mid \mathbf{x})$ | | |
|---|---|---|---|---|
| $x_1$ | $x_2$ | $c{=}1$ | $c{=}2$ | $P(\mathbf{x})$ |
| 0 | 0 | 0.4 | 0.6 | 0 |
| 0 | 1 | 0.5 | 0.5 | 0.1 |
| 1 | 0 | 0.5 | 0.5 | 0.4 |
| 1 | 1 | 0.8 | 0.2 | 0.5 |

$\varepsilon = 0.65$

A) $\varepsilon < 0.25.$
B) $0.25 \leq \varepsilon < 0.50.$
C) $0.50 \leq \varepsilon < 0.75.$
D) $0.75 \leq \varepsilon.$

4 $\boxed{A}$ The figure below shows a partition of 5 two-dimensional points in 2 clusters, $\bullet$ and $\circ$:



If point $(9, 2)^t$ is 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:

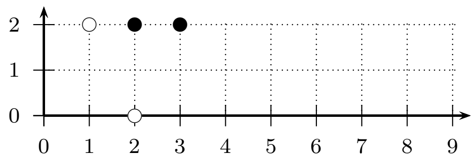
A) $\Delta J < -7$. $\Delta J = 39.5 - 64.0 = -24.5$

B) $-7 \leq \Delta J < 0$.

C) $0 \leq \Delta J < 7$.

D) $\Delta J \geq 7$.

5 $\boxed{D}$ The figure below shows a partition of 4 two-dimensional points in 2 clusters, $\bullet$ and $\circ$:



Indicate which of the following points is transferred from cluster to cluster when we apply the K-means algorithm by Duda and Hart, but not when we apply the conventional K-means algorithm:

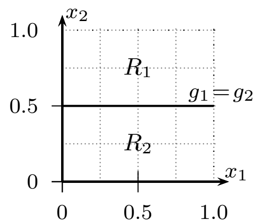
A) $(2, 0)^t$

B) $(2, 2)^t$

C) $(3, 2)^t$

D) $(1, 2)^t$

6 $\boxed{C}$ 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 classifier equivalent to the one of the figure?




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

B) $\mathbf{w}_1 = (1, 0, 0)^t$ and $\mathbf{w}_2 = (0, 0, 2)^t$.

C) $\mathbf{w}_1 = (0, 0, 2)^t$ and $\mathbf{w}_2 = (1, 0, 0)^t$.

D) All the above weight vectors define an equivalent classifier.

7 $\boxed{\text{C}}$ Let us suppose that we have a box with 10 oranges containing 8 oranges Washington (W) and 2 Cadenera (C) from which we draw two oranges, one after the other without replacement. Given the random variables:

- O1: variety of the first drawn orange
- O2: variety of the second drawn orange

Which of the following conditions is not true?

A) $P(O1 = W, O2 = C) = P(O1 = C, O2 = W)$
B) $P(O2 = W) < P(O2 = W \mid O1 = C)$
C) $P(O1 = C) = P(O1 = C \mid O2 = W)$
D) $P(O2 = W) > P(O2 = W \mid O1 = W)$

8 $\boxed{\text{D}}$ 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 none of the first three classifiers is of minimum error)?

A) $c(\mathbf{x}) = \underset{c=1,\dots,C}{\arg\min}\ e^{p(c|\mathbf{x})} + e^{p(\mathbf{x})}$
B) $c(\mathbf{x}) = \underset{c=1,\dots,C}{\arg\min}\ e^{p(\mathbf{x},c)}$
C) $c(\mathbf{x}) = \underset{c=1,\dots,C}{\arg\max}\ -\log p(\mathbf{x}, c)$
D) None of three classifiers is of minimum error.

9 $\boxed{\text{B|D}}$ Let $g(\mathbf{x})$ be a classifier. Which function does *not* define an equivalent classifier (or choose the last option if all three previous functions define an equivalent classifier)?

A) $f(g(\mathbf{x})) = ag(\mathbf{x}) + b \quad a > 0$
B) $f(g(\mathbf{x})) = \log g(\mathbf{x})$ $g(\mathbf{x}) > 0$
C) $f(g(\mathbf{x})) = \exp g(\mathbf{x})$
D) All three previous functions define an equivalent classifier.

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

ETSINF, Universitat Politècnica de València, December 19th, 2024

**Group, surname(s) and name:** 1,

## Problem: Logistic regression

The following table shows per row a sample with 2 dimensions that belongs to one class:

| $n$ | $x_{n1}$ | $x_{n2}$ | $c_n$ |
|---|---|---|---|
| 1 | 1 | 1 | 1 |

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

| $\mathbf{w}_1$ | $\mathbf{w}_2$ |
|---|---|
| 0.5 | -0.5 |
| 0.5 | -0.5 |
| 0.5 | -0.5 |

Answer the following questions:

1. (0.25 points) Compute the vector of logits for the training sample.
2. (0.25 points) Apply the softmax function to the vector of logits for the training sample.
3. (0.25 points) Compute the neg-log-likelihood of the training sample with respect to the initial weight matrix.
4. (0.25 points) Classify the training sample. In case of a tie, choose any class.
5. (0.5 points) Compute the gradient of the function NLL at the point of the initial weight matrix.
6. (0.5 points) Update the initial weight matrix applying gradient descent with learning rate $\eta = 1.0$.

Solution:

1. Vector of logits for the training sample:

| $n$ | $a_{n1}$ | $a_{n2}$ |
|---|---|---|
| 1 | 1.5 | -1.5 |

2. Applying the softmax function:

| $n$ | $\mu_{n1}$ | $\mu_{n2}$ |
|---|---|---|
| 1 | 0.95 | 0.05 |

3. Computation of the neg-log-likelihood:

$$\text{NLL}(\mathbf{W}) = 0.05$$

4. Classification of the training sample:

| $n$ | $\hat{c}(x_n)$ |
|---|---|
| 1 | 1 |

5. Gradient:

| $\mathbf{g}_1$ | $\mathbf{g}_2$ |
|---|---|
| -0.05 | 0.05 |
| -0.05 | 0.05 |
| -0.05 | 0.05 |

6. Updated weight matrix:

| $\mathbf{w}_1$ | $\mathbf{w}_2$ |
|---|---|
| 0.55 | -0.55 |
| 0.55 | -0.55 |
| 0.55 | -0.55 |