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

ETSINF, Universitat Politècnica de València, February 1st, 2024

Group, surname(s) and name: 2,

Tick only one choice among the given options. Score: max(0, (correct_answers−wrong_answers / 3) · 1.75 / 6).

1 C Given the following probability distributions:

B	0	0	1	1
C	0	1	0	1
$P(A=0 \mid B,C)$	0.222	0.298	0.234	0.118
$P(B,C)$	0.025	0.467	0.219	0.290

Which is the value of $P(A=1, B=1 \mid C=0)$? $P(A=1, B=1 \mid C=0) = 0.689$

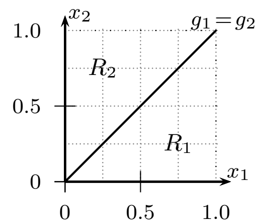
A) $\text{P(A=1, B=1} \mid C=0) \leq 0.25$

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

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

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

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

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

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

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

3 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 4 bidimensional learning samples for a problem of 4 classes, $c = 1, 2, 3, 4$. At a given moment in the execution of the algorithm, we have obtained the weight vectors $\mathbf{w}_1 = (-2, -8, -5)^t$, $\mathbf{w}_2 = (-2, -8, -9)^t$, $\mathbf{w}_3 = (-2, 0, -3)^t$, $\mathbf{w}_4 = (-2, -4, -9)^t$. Assuming that the sample $(\mathbf{x}, c) = ((5, 4)^t, 1)$ is then going to be processed, how many weight vectors will be modified?

A) 0

B) 2

C) 3

D) 4

4 $\boxed{\text{B}}$ The estimated probability of error of a classifier is 5%. Which is the minimum number of testing samples, M, so that the 95% confidence interval of this estimated probability of error is not higher than $\pm 1\%$; that is, $I = [4\%, 6\%]$: $M = 1825$

A) $M < 1000$.

B) $1000 \leq M < 2000$.

C) $2000 \leq M < 3000$.

D) $M \geq 3000$.

5 $\boxed{\text{B}}$ Given the following dataset to train a classification tree with 5 bidimensional samples that belong to 2 classes:

n	1	2	3	4	5
x_{n1}	4	1	2	1	3
x_{n2}	4	4	1	1	1
c_n	1	1	1	1	2

How many different partitions can be generated at the root node? Do not consider those partitions in which all data samples are assigned to the same child node.

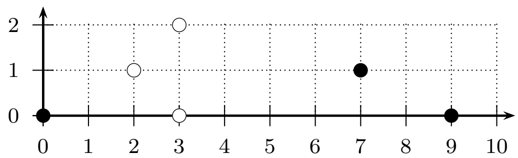
A) 6

B) 4

C) 3

D) 2

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



What point when transferred minimises the variation of the Sum of Square Errors (SSE), $\Delta J = J - J'$ (SSE after the transfer minus SSE before the transfer)? $\Delta J = 11.2 - 48.0 = -36.8$

A) $(0, 0)^t$

B) $(9, 0)^t$

C) $(2, 1)^t$

D) $(3, 0)^t$

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

ETSINF, Universitat Politècnica de València, February 1st, 2024

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	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.
0.25	-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	0.73	0.27

3. Classification of every sample:

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

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