Instituto Superior Técnico

Aprendizagem 2021/22

BLUE version **Exam 1** 22 November 2021

- A duração deste exame é de 2 horas, com uma tolerância adicional de 30 minutos.
- Desistências e entregas após os primeiros 90 minutos.
- É permitida apenas a consulta de uma folha (frente e verso).
- O número total de pontos é 20.
- Marque as suas respostas NA FOLHA DE EXAME.
- Escreva o seu número e nome no topo de cada página.
- Escreva todas as fórmulas. Apresente todos os cálculos e justificações para as suas respostas.

- The exam duration is 2h with an additional 30 minutes tolerance. *Panic not.*
- Withdrawals and deliveries after the first 90 minutes.
- The consultation of a single sheet note is allowed.
- 20 points total.
- Mark your answers ON THE EXAM ITSELF.
- Write your number and name at the top of each page.
- Write all formulas. Present all the computations and justifications for your answers.

Não preencher: para o uso oficial / not fill: for official use only

1	**2**	**3**	**4**	**5**	**6**	**SUM**
4v	2v	4.5v	4v	3.5v	2v	20v

1. [3.5 pts] **Bayes and tree learning**

Consider the following discrete dataset:

	y_1	y_2	class
$\mathbf{x}_1$	A	B	A
$\mathbf{x}_2$	B	C	A
$\mathbf{x}_3$	B	C	B
$\mathbf{x}_4$	B	A	B
$\mathbf{x}_5$	C	A	B

Showing your calculus,

a) [2v] Classify $\mathbf{x} = \binom{\mathrm{B}}{\mathrm{C}}$ using naive Bayes with the maximum a posteriori (MAP) assumption

b) [1v] Consider a tree learned using ID3 (information gain). Which one of the variables is tested on the root of the tree?

c) [1v] Consider a classifier that, among the training observations, wrongly classifies $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$. Plot the model's confusion matrix and identify the training sensitivity of class B.

2. [2 pts] **Perceptron**

Consider a perceptron with weights, and which processing unit implements the following function

$$f(\mathbf{x}) = \exp\left(\left(\sum_{j=0}^{D} w_j \times x_j\right)^2 + 1\right)$$

Determine the gradient descent-training rule for squared error

$$E(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{n}(t_i - o_i)^2$$

3. [4.5 pts] **Neural networks**

Consider a network with 4 inputs, 2 hidden units and 2 outputs where the hidden units use $ReLU$ activation function and the output units $\mathrm{softmax}$ activation function.
Initializing connection weights and biases to 1.
Consider a cross-entropy loss and a stochastic gradient descent update for the training example:

$$\{\mathbf{x} = [1\ 1\ 1\ 1]^T, \mathbf{z} = [0\ 1]^T\}$$

a) [1v] Do forward propagation *(remove guiding questions?)*

b) [1v] Compute $\delta^{[2]} = \frac{\partial \mathrm{Error}}{\partial \mathrm{NET}^{[2]}}$ *(remove guiding questions?)*

c) [1v] Update $W^{[2]}$ using a learning rate of 0.1 *(remove guiding questions?)*

d) [1.5v] Compute $\delta^{[1]} = \frac{\partial \mathrm{Error}}{\partial \mathrm{NET}^{[1]}}$ *(remove guiding questions?)*

4. [4 pts] **Regression and PCA**

Consider the following observations in a Euclidean space:

	y_1	y_2	z
$\mathbf{x}_1$	3	1	0.1
$\mathbf{x}_2$	1	2	0.4
$\mathbf{x}_3$	3	3	0.3
$\mathbf{x}_4$	1	0	0.7

a) [2.5v] Estimate the quantity $z = f(\mathbf{x})$ where $\mathbf{x} = [1,1]^T$ and f is given by:

i. [1.25v] a kNN model with $k = 3$, a median estimator and uniform weights.

ii. [1.25v] a linear regression model with $\phi(\mathbf{x}) = \|\mathbf{x}\|_2{}^2$ and $\mathbf{w} = (-0.2, 0.5)^T$

b) [1.5v] The following covariance matrix and eigenvalues were produced for the given dataset:

$$C = \begin{pmatrix} 1.333 & 0.667 \\ 0.667 & 1.667 \end{pmatrix}, \quad \lambda_1 = 2.187, \quad \lambda_2 = 0.813$$

Project the input bivariate data space into a univariate data space by applying PCA with the most informative component.

5. [4.5 pts] **RBFs and clustering**

Consider the following training set described by three vectors

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 0 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 3 \\ 2 \\ 0 \\ 0 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \end{pmatrix},$$

and the corresponding binary targets:

$$t1 = 1, t2 = 1, t3 = -1$$

a) [1v] Identify the k-means clustering solution assuming $k = 2$ and cluster centers initialized as

$$\boldsymbol{c}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{c}_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Hint: *no computation of distance values for k-Means is required!*

b) [2.5 pts] Can a RBF network applied with the previous *k*-Means clustering solution and clusters with $\sigma = 1$ solve the given classification problem? Justify, showing the observations on the hidden/transformed data space.

(continuation of 5.b)

6. [2 pts] **Model complexity**

Assuming a binary classification problem with inputs of dimension $d > 3$. Consider three models:

i. a fixed feature transformation $\phi: \mathbb{R}^d \to \mathbb{R}^6$ followed by a perceptron.
ii. a learnable feature transformation $\phi: \mathbb{R}^d \to \mathbb{R}^2$ that depends on 4 parameters, followed by a perceptron
iii. a multilayer perceptron with one hidden layer of size three (i.e. architecture $[d, 3, 2]$).

Order the four models in terms of their expected risk of overfitting.

Justify each decision with a short sentence.

END