WRITE FIRST NAME, LAST NAME, AND ID NUMBER ("MATRICOLA") BELOW AND READ ALL INSTRUCTIONS BEFORE STARTING WITH THE EXAM! TIME: 2.5 hours.

FIRST NAME: .....

LAST NAME: .....

ID NUMBER: .....

INSTRUCTIONS

- solutions to exercises must be in the appropriate spaces, that is:
  - Exercise 1: pag. 1, 2, 3
  - Exercise 2: pag. 4, 5, 6
  - Exercise 3: pag. 7, 8, 9
  - Exercise 4: pag. 10, 11, 12

  Solutions written outside the appropriate spaces (including other papersheets) will not be considered.
- the use of notes, books, or any other material is forbidden and will make your exam invalid;
- electronic devices (smartphones, calculators, etc.) must be turned off; their use will make your exam invalid;
- this booklet must be returned in its entirety.

Exercise 1 [8 points]

1. Provide the formulation of PAC learning (defining losses, risks, training algorithms and probabilistic requirements)

2. In the context if PAC learning define the concept of realizability and discuss how item 1 above changes when the realizability assumption cannot be made.

3. Provide a (probabilistic) bound on the generalization error for the ERM when aorking with finite hypothesis classes (proof not needed)

[Solution: Exercise 1]

[Solution: Exercise 1]

Exercise 2 [8 points]

1. Introduce the neural network model highlighting its main components and pointing out which are the parameters to be learned in the training process. Explain which are the main design parameters AND/OR describe how the output layer of a neural network for binary classification is made.

2. Consider a fully connected neural network N with $L = 4$ layers, with $n_1 = 5$ neurons in the input layer, $n_2 = n_3 = 4$ neurons in the two inner layers and a single neuron ($n_4 = 1$) in the last (i.e., output) layer. How many trainable parameters are there in the network? How this number is related to the number of neurons in each layer ? E.g., is it a linear, quadratic or exponential relationship?

3. Which provisions are used in the Convolutional Neural Network (CNN) model to reduce the number of trainable parameters? Highight in your answer the differences with respect to the fully connected model.

[Solution: Exercise 2]

[Solution: Exercise 2]

[Solution: Exercise 2]

Exercise 3 [8 points]

With reference to the binary classification problem:

1. Describe a framework under which the decision boundary can be an arbitrary polynomial function of degree M.
2. Discuss how this can be related to kernel SVM, possibly highlighting which is the advantage of the "kernel" interpretation.
3. Assuming one has (x_i, y_i), $i \in [m]$ data points with m "small", describe a procedure to perform the selection of M (deciding the "most suitable" degree of the polynomial boundary), $M \in \{2, 3, 4, .., 10\}$.

[Solution: Exercise 3]

[Solution: Exercise 3]

[Solution: Exercise 3]

Exercise 4 [8 points]

- Introduce the problem of dimensionality reduction - - given a regression problem with X almost singular; how can be dimensionality reduction used to reduce the number of parameters to be estimated.

- given the two cases in the figure, tel whetehe it is possible to reduce dimension and, if so , which is the most approriate dimension of the reduced problem.

1.

[Solution: Exercise 4]

[Solution: Exercise 4]

[Solution: Exercise 4]