EXAM 21/02/2020

Exercise 1 [8 points]

1. Provide the formulation of PAC learning (including the definition of loss, risk, training algorithms, etc.).
2. In the context of PAC learning define the concept of realizability and discuss how the formulation above changes when the realizability assumption cannot be made.
3. ~~Provide a (probabilistic) bound on the generalization error for the ERM when working~~ with finite hypothesis classes (proof not needed).

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. Describe how the output layer of a neural network for binary classification can be designed.

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 is this number related to the number of neurons in each layer?

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

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 $(\mathbf{x}_i, y_i)$, $i \in [m]$ data points with m "small", describe a procedure to perform the selection of M (deciding the "most suitable" degree of the polynomial boundary), $M \in \{2, 3, 4, .., 10\}$.

Exercise 4 [8 points]

1. Introduce the problem of dimensionality reduction, describing what is the input, what is the output, and what is its goal.

2. Consider a linear regression problem with squared loss, where the input feature vectors are $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$. Assume that the (design) matrix $\mathbf{X}$, whose i-th row is $\mathbf{x}_i^\top$ is such that $\mathbf{X}^T\mathbf{X}$ is almost singular. (Remember that a square matrix is singular if and only if it is not invertible.) Explain how dimensionality reduction can be used to reduce the number of parameters to be estimated and to learn a model in this situation.

3. Consider the two datasets with points $\mathbf{x} \in \mathbb{R}^2$ shown in Figure (a) and (b) below. For both datasets, describe i) whether it is possible to meaningfully reduce the dimensionality of the data and ii) what is the most appropriate dimension of the data after dimensionality reduction.