Machine Learning – June 22, 2022

Student data

Matricola	Last Name	First Name
.....	.....	.....

Notes

1. No books, slides, written notes are allowed during the exam.
2. Answers must be explicitly marked with the question they refer to (e.g., 2.1 for question 1 of exercise 2). Cumulative answers which refer to more questions at once will be considered as answering one question only (and will be evaluated as such).

Time limit: **2 hours**.

EXERCISE 1

1. Provide a formal (domain-independent and solution-independent) definition of overfitting.
2. Discuss the problem of overfitting in learning with Decision Trees and illustrate possible solutions to it.

EXERCISE 2

1. Explain when a dataset is *linearly separable*
2. Draw an example of a linearly separable dataset in a 2D setting, with two classes $C = \{+, -\}$
3. Draw an example of a non linearly separable dataset in a 2D setting, with two classes $C = \{+, -\}$
4. For each dataset shown above, draw a possible solution based on SVM and explain how it can be obtained.

EXERCISE 3

Consider a regression problem $f : \Re^d \rightarrow \Re$ with a dataset $D = \{(\mathbf{x}_n, t_n)_{n=1}^N\}$, where f is known to be non-linear in x.

1. Describe a linear model for this problem and determine the trainable parameters and the size of the model (i.e., number of trainable parameters).
2. Describe a solution of the problem in terms of least square error minimization.

EXERCISE 4

Consider the problem of estimating the function $f : \Re^3 \rightarrow \Re$, with dataset $\mathcal{D} = \{(\mathbf{x}_1^T, t_1), \ldots, (\mathbf{x}_N^T, t_N)\}$. Let us consider using a feed-forward network to estimate this function:

1. Explain how the problem is formalized by writing the parametric form of the function to be learned explaining all its terms.
2. Explain what is a suitable choice for the loss function used for training the network and write the corresponding mathematical expression.
3. Assuming that the gradients of the loss with respect to the parameters are available, describe an algorithm for training the parameters of the network. What are the hyper-parameters of the training algorithm (if any)?

EXERCISE 5

A cleaner robot is required to clean up N rooms in an environment, the robot can move to any of these rooms and clean them. The robot is equipped with perception abilities to know where it is (i.e., in which room) and all the rooms that need to be cleaned up. Assume that the environment is stationary and fully observable.

We want to model this problem as a Markov Decision Process (MDP).

1. Describe a complete model for this problem based on MDP, specifying all its elements: state space, action space, transition function, and reward function
2. Describe a complete optimal policy for the case of $N = 3$.

EXERCISE 6

Consider a problem represented with an unsupervised dataset $D = \{\mathbf{x}_n\} \subset \Re^2$ generated by an unknown Gaussian Mixture Model (GMM) with $k = 3$

1. Define the GMM model, describe the parameters of the model, and determine the size of the model (i.e., number of independent parameters) for the problem considered above.
2. Draw an example of a dataset for the above problem and show graphically the corresponding parameters of the model.