Machine Learning – December 17, 2025

Matricola Last Name First Name

....

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 will be evaluated as answering one question only.

Time limit: **2 hours**.

EXERCISE 1

Consider a binary classification problem $X \rightarrow \{T, F\}$, with $X = \{T, F\}^3$, i.e. $(x_1, x_2, x_3) \in X$ and $x_i \in \{T, F\}$, and the dataset $D = \{\langle (F, F, F), F\rangle, \langle (F, T, T), T\rangle, \langle (T, T, F), T\rangle, \langle (T, F, T), T\rangle\}$. Consider the two hypothesis $h_1 = (x_1 \wedge \neg x_2 \wedge x_3) \vee x_2$ and $h_2 = (\neg x_1 \wedge x_2 \wedge x_3) \vee x_1$.

1. Determine whether h_1 and h_2 are consistent with D, showing all the passages needed to answer.
2. Assuming the likelihood probabilities $P(D|h_1) = 0.6$ and $P(D|h_2) = 0.8$ and the prior probabilities $P(h_1) = 0.2$ and $P(h_2) = 0.1$, determine the higher a poteriori hypothesis between h_1 and h_2.

EXERCISE 2

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

1. Explain what is a suitable choice for the loss function used for training the network and write the corresponding mathematical expression.
2. 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 3

1. Describe the principle of maximum margin used by SVM classifiers through its formal mathematical definition.
2. Draw a linearly separable dataset for 2D binary classification. Draw a possible solution obtained by SVM and highights the margin and the support vectors.
3. Discuss why the maximum margin solution is preferred for the classification problem.

EXERCISE 4

Consider a dataset D for the binary classification problem $f : \Re^3 \mapsto \{Y, N\}$.

1. Describe a probabilistic generative model for such a classification problem, assuming Gaussian distributions.
2. Identify the parameters of the model and determine the size of the model (i.e., the number of independent parameters).

EXERCISE 5

Consider the following dataset, containing the samples of a function f:

x_1	x_2	f
0.0	0.0	1.0
0.6	0.3	2.2
1.2	0.4	3.0
1.5	0.5	3.5

1. Based on the available data, select a reasonable model for learning f, explicitly indicating its parameters.
2. Show an optimal and a non-optimal solution, explicitly indicating, for each of them, the corresponding value of the loss function.

EXERCISE 6

1. Describe the concept of bagging in the definition of an ensemble model. Describe precisely the training procedure for such a model and the final formula used for prediction.
2. Discuss the difference between bagging and voting, highlighting in particular the use of different types of models.