Machine Learning – June 28, 2024

Matricola Last Name First Name

....

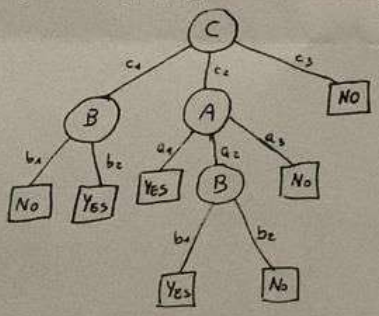
Notes

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 will be evaluated as answering one question only.

Time limit: **2 hours**.

EXERCISE 1

Given a classification problem for the function $f : A \times B \times C \rightarrow \{YES, NO\}$, with $A = \{a_1, a_2, a_3\}, B = \{b_1, b_2\}, C = \{c_1, c_2, c_3\}$ and the following decision tree T that is the result of training on some data set:



1. Provide a rule based representation of the tree T.
2. Determine if the tree T is consistent with the following set of samples

   $S \equiv \{(s_1 = \langle a_2, b_1, c_2 \rangle, YES), (s_2 = \langle a_1, b_1, c_1 \rangle, NO), (s_3 = \langle a_1, b_2, c_3 \rangle, YES), (s_4 = \langle a_3, b_2, c_1 \rangle, YES), (s_5 = \langle a_3, b_1, c_2 \rangle, NO)\}$.

   Motivate your answer.
3. Compute the accuracy of T with respect to S.

EXERCISE 2

1. Describe the K-nearest neighbors (K-NN) algorithm for classification and discuss its limitations.
2. Draw a 2D classification dataset for three classes $\{circle, star, plus\}$, choose one query point close to points in the dataset belonging to different classes, and determine the answer of K-NN for such a query point for K=1, K=3, and K=5, illustrating a situation in which the answer of K-NN depends on K (i.e., you get different solutions for K=1, K=3, K=5). Motivate your answer, showing (with a graphical drawing) which instances contribute to the solution.

EXERCISE 3

1. Describe the difference between bagging and boosting when combining multiple learners.
2. Consider applying both bagging and boosting to a classification problem and the following models $y_1(x), \ldots, y_M(x)$ for bagging and $z_1(x), \ldots, z_M(x)$ for boosting as results of the training phase. Write down the formal equations to combine the predictions of the different models for each of the two approaches.

EXERCISE 4

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 5

Consider a dataset D for the classification problem $f : \Re^5 \mapsto \{C_1, C_2, C_3\}$.

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 6

Consider a two-layers ANN which receives in input real-valued vectors $\mathbf{x}$ of dimension 100 and produces in output real-valued vectors $\mathbf{y}$ of dimension 10. The hidden layer of the ANN is composed of 50 units which use the ReLU activation function. The output units use a linear activation function.

1. Compute the number of trainable parameters including the bias terms, motivating the answer (i.e., show how to compute this value)
2. Provide the formula explicitly stating how the values of $\mathbf{y}$ are computed given an input vector $\mathbf{x}$ in terms of the weight matrices and the activation functions (including the bias terms).
3. Provide a suitable loss function for training the network. Motivate your answer, answers with no motivation will not be considered.