Machine Learning – January 16, 2023

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

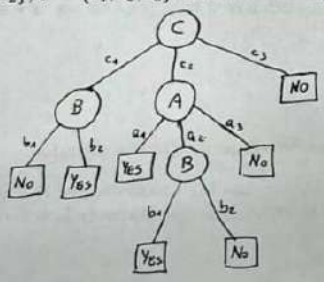
Notes

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:



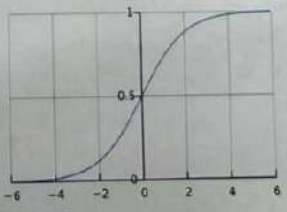
1. Provide a rule based representation of the tree T.
2. Determine if the tree T is consistent with the following set of samples
   $S = \{s_1 = \langle a_1, b_1, c_1, NO\rangle, s_2 = \langle a_2, b_1, c_2, YES\rangle, s_3 = \langle a_1, b_2, c_3, NO\rangle, s_4 = \langle a_3, b_2, c_1, NO\rangle\}$.
   Motivate your answer.
3. Compute the accuracy of T with respect to S.

EXERCISE 2

1. Define mathematically the problem solved by logistic regression
2. Consider the following dataset, two solutions $\vec{w}_1^T$ and $\vec{w}_2^T$ and the sigmoid function plotted below:

x_1	x_2	x_3	t
0	0	1	0
1	2	3	0
4	4	1	1

$\vec{w}_1^T = (2, 0, -2)$
$\vec{w}_2^T = (-2, -2, 4)$



Which one among these solutions fits the data better? Why?
(You do not need to compute explicit values of the model.)

EXERCISE 3

Consider the the k-armed bandit problem (also known as One-state MDP) with stochastic Gaussian behavior and unknown reward functions.

1. **Formally** describe the *model*, provide mathematical definitions and explanations of all the elements of the model.
2. **Formally** describe the *problem* and the *solution concept* (i.e., what is the problem to be solved with the model described in the previous point), provide mathematical definitions and explanations of all the elements of the problem.
3. **Formally** describe the Reinforcement Learning algorithm to compute the optimal policy for this problem, provide explanations of all the terms of the algorithm.

EXERCISE 4

Consider an image classification problem for RGB color images (3 channels) with resolution 128×128 and a LeNet-like CNN model, with a first convolutional layer formed by 5 kernels 7×7 with stride 1 and same padding, followed by max pooling 2×2 with stride 2 and valid padding.

1. Determine the number of trainable parameters of the first convolutional layer.
2. Determine the size of the feature maps after the first convolutional layer.
3. Provide a numerical example (invent some numbers) of the application of a max pooling operation 2×2 with stride 1 and valid padding to a feature map $8 \times 8 \times 1$.

EXERCISE 5

1. Qualitatively explain the maximum margin principle applied to Non-Linear SVM classification.
2. Draw an example of a 2-D binary non-linear classification problem showing qualitatively: i) a generic solution, and ii) a maximum margin solution computed by SVM, describing qualitatively how to obtain such a solution.

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 different use of the training data.