Machine Learning – October 22, 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

1. Explain the ID3 algorithm for Decision Tree learning: **formally** describe input, output and the main steps of the algorithm.
2. Describe the concept of pruning in Decision Tree learning, highlighting the main motivation and an operational procedure to perform it.

EXERCISE 2

Consider a classification problem $f : X \times Y \times Z \rightarrow \{T, F\}$, with $X = \{a, b, c\}, Y = \{0, 1\}, Z = \{u, v, w\}$ and the data set in the table on the right

X	Y	Z	f
a	1	v	F
b	1	w	T
c	0	u	T
b	0	w	F
a	0	w	T
c	1	u	T

1. Describe a solution of this problem based on Naive Bayes. Which terms are computed in the Naive Bayes model and how?
2. Describe how to predict the class for the new sample $(a, 0, u)$ based on Naive Bayes (just provide the formula, no need to compute the numerical result).

EXERCISE 3

1. Describe the perceptron model for classification and its training rule.
2. Draw a graphical representation of a linearly separable 2D data set for binary classification and provide a qualitative graphical example of a possible evolution of perceptron training (4 images showing a possible temporal evolution of the solution of the algorithm on the sketched data set, with the last image showing a possible final solution).

EXERCISE 4

Consider the following Convolutional Neural Network (CNN) acting on images of dimension $96 \times 96 \times 3$:

conv1	Conv2D 3×3 kernel, 8 feature maps with padding 1 and stride 1
relu1	ReLU activation function
pool1	2×2 max pooling with stride 2
conv2	Conv2D 3×3 kernel and 5 feature maps with padding 1 and stride 1
relu2	ReLU activation function
conv3	Conv2D 3×3 kernel and 3 feature maps with padding 1 and stride 1
relu3	ReLU activation function
pool3	2×2 max pooling with stride 2
flatten	flatten operation
fc1	30 units
relu4	ReLU activation function
do1	dropout with rate 0.5
fc2	10 units
relu5	ReLU activation function
do2	dropout with rate 0.5
fc3	10 units
output	softmax

1. What kind of problem and which dimension of the problem it is able to solve?
2. Explain what is the role of the dropout layer in a CNN and in particular the meaning of the dropout rate.
3. Explain and motivate what is a suitable loss function to train this network.

EXERCISE 5

Given a data set $D = \{(x_n, t_n)_{n=1}^N\}$, denoted with its design matrix $\mathbf{X}$ and its output vector $\mathbf{t}$, for a classification problem $f : \Re^d \to C$, with $|C| = K$, and a linear model $y(x; w)$ with parameters w,

1. Define a regularized squared error function to solve the problem
2. Describe a kernelized linear model obtained by solving the above problem, and provide the solution of such a problem including a formal definition of the Gram matrix.
3. Provide the dimensions of all the elements of the solution

EXERCISE 6

Given a data set D for binary classification and a set of learners $y_m(x),\ m = 1, \ldots, M$,

1. Describe the general approach of AdaBoost and provide details on its sequential training.
2. Provide the final classifier that combines the given learners $y_m(x)$.