

Machine Learning – January 30, 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 classification problem $f : X \times Y \times Z \rightarrow \{T, F\}$, with $X = \{a, b, c\}$, $Y = \{0, 1\}$, $Z = \{u, v, w\}$ and data set D 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. Draw a decision tree consistent with the data set.
2. Write a new sample for this problem (not in the dataset) and classify this new sample according to the decision tree provided above.
3. Describe the problem of overfitting in decision trees and **formally** provide a possible solution: i.e., write the pseudo-code of a suitable algorithm to deal with overfitting in decision trees, explaining input, output and all the steps.

EXERCISE 2

1. Describe the Least squares method for linear classification. Formally define the error function that is used to solve the problem and the method used to compute a solution.
2. Draw a 2D data set for binary classification and qualitatively draw a possible solution that can be obtained by Least square.
3. Add to the above data set, some examples from one class coming from a different distribution (i.e., outliers far from the current points), in such a way that the Least square solution will be significantly different from the one shown in the previous point. Draw a possible solution for Least square in this new data set.

EXERCISE 3

Assume you are given a dataset D for regression of a function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$.

1. Provide the mathematical formulation of a suitable model for the problem.
2. Describe a suitable solution method, illustrating an error function and a solution method.

EXERCISE 4

1. Give a short explanation of the *Kernel trick* (kernel substitution). What are the necessary conditions for applying the kernel trick?
2. Consider a linear model for regularized binary classification, i.e. $\mathbf{y}(\mathbf{x}; \mathbf{w}) = \mathbf{w}^T \mathbf{x}$ with regularization $\|\mathbf{w}\|^2$. Provide an example of applying the kernel trick on this problem. In detail:
 - provide the mathematical formulation of the model in its original form (before applying the kernel trick);
 - explain why it is possible to apply the kernel trick and provide the “kernelized” formulation of the model.

Note: If needed, consider the identity $(\mathbf{X}^T \mathbf{X} + \lambda I)^{-1} \mathbf{X}^T = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T + \lambda I)^{-1}$.

EXERCISE 5

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. Compute the number of *trainable* parameters of the first block (conv1,relu1,pool1) and of the block from the output of do1 to the output of do2.
3. Explain what is the role of the dropout layer in a CNN and in particular the meaning of the dropout rate.

EXERCISE 6

1. Define the unsupervised learning problem. Illustrate the concepts through a 2D data set ($|D| > 10$) and describe the concept of expected solutions for this problem.
2. Describe the difference between K-Means and Expectation-Maximization. Illustrate the differences on the above data set (or another one, if needed). In particular, show an example in which the solution provided by EM is “better” than the one provided by K-means.