WRITE FIRST NAME, LAST NAME, AND ID NUMBER ("MATRICOLA") ON YOUR ASSIGNMENT. TIME: 1.5 hours.

FIRST NAME: .....

LAST NAME: .....

ID NUMBER: .....

Question 1 [6 points]

1. Introduce the regression problem in machine learning.
2. Describe the Gaussian regression framework and show that, under suitabkle assumtions, the estimated function is linear.

[Solution: Question 1]

[Solution: Question 1]

Question 2 [6 points]

Consider a linear regression model of the form

$$h(x) = w^\top x \quad x \in \mathbb{R}^d$$

and assume we are given data (x_i, y_i), $i = 1, .., m$

1. When is the problem of estimating the model from data ill-conditioned (and what does it mean)?
2. How can you cure ill-conditioning? Show mathematically that the approach you suggest avoids the ill-conditioning issue.

[Solution: Question 2]

[Solution: Question 2]

Question 3 [6 points]

1. Describe the PAC learning framework.

2. Define the concept of sample complexity; with reference to the notation used in class, for binary classification and assuming the model class is finite and that realizability holds true, draw two qualitative plots of how sample complexity changes as a function $1/\epsilon$ (with δ fixed) and as a function of $1/\delta$ (with ϵ fixed).

[Solution: Question 3]

[Solution: Question 3]

Question 4 [6 points]

With referemce to the binary classification problem:

1. Describe (linear) soft SVMs and formulate mathematically the optimization problem to be solved;
2. Let λ be the regularization parameter associated with the slack variables; discuss what happens to the "margin" as λ varies. Is it possible to recover (possibly under suitable assumptions) the "hard" SVM solution? If so, how?

[Solution: Question 4]

[Solution: Question 4]