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]

Consider a regression problem in which data y_i, $i = 1, .., m$, are noisy measurements of an unknown target function $f(x_i)$, i.e.

$$y_i = f(x_i) + n_i.$$

Consider the model class of linear functions $\mathcal{H} := \{h(x) := w^\top x + b\}$.

1. Denoting with $L_S(h)$ the empirical risk, introduce the Empirical Risk Minimization principle, discuss the overfitting problem and how this can be cured.

2. Introduce the concept of regularization, discuss which "tuning knobs" one typically has, how they can be tuned and what would be epected to happen when $m \to \infty$.

[Solution: Question 1]

[Solution: Question 1]

Question 2 [6 points]

Let $x_i \in \mathbb{R}^d$, $i = 1, ..., m$, be measured (i.i.d.) data from an unknown distribution $\mathcal{D}_x$ (with probability density function $p(x)$).

Assume you would like to build an estimate of the probability density function $p(x)$.

1. Under which conditions would you expect that (linear) dimensionality reduction can be helpful to the purpose? (you can also use a graphical illustration to explain your answer).

2. Under these conditions, and assuming you have perfomed linear dimensionality reduction, how can you now exploit Gaussian Mixture Models to obtain an estimate of the pdf $p(x)$ of x?

[Solution: Question 2]

[Solution: Question 2]

Question 3 [6 points]

In the context of binary classification:

1. Introduce the perceptron algorithm and discuss its properties;
2. Introduce Support Vector Machines (hard and soft) and, using your answer to the question above, motivate their use discussing their advantages (both for hard and soft).

[Solution: Question 3]

[Solution: Question 3]

Question 4 [6 points]

Consider the GMM model for clustering data $\{x_i\}_{i=1,..,m}$ in K (known and fixed) clusters.

1. Formulate the problem of estimating the parameters of the GMM using the ML principle in combination with the EM algorithm (you don't need to report all the specific equations of EM, only describe the two main steps)

2. How would you modify the formulation if, for a subset of indexes $i \in \mathcal{I} \subset \{1,..,m\}$, labelled data were available (i.e. we would also know the cluster indexes $y_i \in \{1,..,K\}$, $\forall i \in \mathcal{I}$)?

[Solution: Question 4]

[Solution: Question 4]