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

FIRST NAME: ....

LAST NAME: ....

ID NUMBER: ....

Exercise 1 [6 points]

1. Define the framework of PAC learning with particular reference to the case the model class $\mathcal{H}$ is finite.

2. Discuss the relation between the approximation error (ϵ), the violation probability δ and the training set size m.

Exercise 2 [6 points]

Consider a regression problem with model class

$$\mathcal{H} := \{h_w(x) = \sum_{i=1}^{d} w_i \phi_i(x),\ w_i \in \mathbb{R}\}$$

where the functions $\phi_i(x) : \mathbb{R} \to \mathbb{R}$ are given and linearly independent.

1. Given a training set $z_i = (x_i, y_i),\ i = 1, .., m,\ x_i \in \mathbb{R},\ y_i \in \mathbb{R}$, write the explicit expression of the empirical risk minimizer $\hat{w}_s \in \mathbb{R}^d$ under the square loss

$$\ell(z, w) = (y - h_w(x))^2$$

2. Under the following assumptions:

   - d is large
   - it is expected that the "true" regression function $h(x)$ can be written exactly as a linear combination of a small number of the basis function $\phi_i(x)$, yet it is NOT known which basis functions are needed.

   Which properties would you like to favour in $\hat{w}$? Describe a procedure to estimate the parameter w.

Motivate your answers.

Exercise 3 [6 points]

1. Describe how support vector machines can be extended to perform non-linear (binary) classification (i.e. solve classification problems with non-linear boundaries)
2. Formulate the associated optimization problem and discuss a procedure to select the regularization parameters.

Motivate your answers.

[Solution: Exercise 3]

[Solution: Exercise 3]

[Solution: Exercise 3]

Exercise 4 [6 points]

Let $x_1, ..., x_m$, $x_i \in \mathbb{R}^d$, $m >> d$ be i.i.d. samples from a distribution $\mathcal{D}_x$ that has zero mean, and let

$$USV^\top = X \tag{1}$$

be the singular value decomposition of the matrix $X = [x_1, .., x_m]^\top$. Assume that

$$S := \text{diag}\{s_1, .., s_d\}$$

is such that

$$s_1 \geq s_2 \geq \cdots \geq s_k > 0$$

while $s_{k+1} = \cdots = s_d = 0$.

1. Which is the rank of the covariance matrix $\Sigma := \mathbf{E}xx^\top$?

2. You would like to find a vector β, $\|\beta\| = 1$, such that $\beta^\top x$ has as large variance as possible.

   - How would you express β as a function of the singular value decomposition (1)?
   - Can you provide and estimate of the variance of $\beta^\top x$?

Motivate your answers.