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 notion of stable algorithms (OAROS) in machine learning

2. Show that if an algorithm is (OAROS) stable it does not overfit

[Solution: Exercise 1]

[Solution: Exercise 1]

Exercise 2 [6 points]

Assume you are given i.i.d. data $(x_1, ..., x_m)$, $x_i \in \mathbb{R}$ from an unknown distributon $p(x)$. Your task is to estimate the distribution $p(x)$ and, to do so, you use the Maximum Likelihood criterion and choose as a model class a mixture of Gaussians ($K \in \mathbf{N}$ is fixed):

$$p_\theta(x) = \sum_{\ell=1}^{K} \alpha_\ell p_{m_\ell, \sigma_\ell}(x)$$

where the parameter vector θ collects of all the parameteres $\alpha_\ell, m_\ell, \sigma_\ell$, $\ell = 1, .., K$

$$\sum_{\ell=1}^{K} \alpha_\ell = 1 \qquad \alpha_\ell \geq 0$$

and $p_{m_\ell, \sigma_\ell}(x)$ is the density of a Gaussian random variable with mean m_ℓ and variance σ_ℓ^2.

1. Discuss the relation of this problem with the clustering problem and

2. Describe an algorithm to estimate the unknown parameter θ.

[Solution: Exercise 2]

[Solution: Exercise 2]

Exercise 3 [6 points]

Assume you are given noisy measurements y_i of an (unknown) nonlinear function

$$y_i = f(x_i) + n_i \quad i = 1, .., m$$

where x_i are (known) input locations and n_i are i.i.d. noises.

Describe a framework to estimate the (unknown) function $f(\cdot)$ from measurements (y_i, x_i), $i = 1, .., m$

Which are the "user choices" and how are they made?
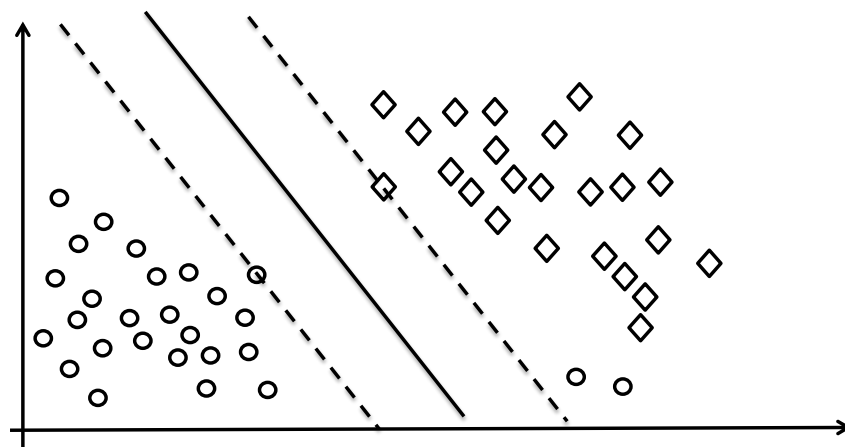

[Solution: Exercise 3]

[Solution: Exercise 3]

Exercise 4 [6 points]

Given training data (x_i, y_i), $i = 1, .., m$, $y_i \in \{-1, 1\}$, $x_i \in \mathbb{R}^d$, the (binary) soft-SVM classifier aims at minimizing the following function:

$$\lambda ||\mathbf{w}||^2 + \frac{1}{m} \sum_{i=1}^{m} \xi_i.$$

1. Briefly explain how the Soft-SVM classification method works and which are the constraints under which the function has to be minimized.

2. The figure shows the results of a binary classification performed using a Soft-SVM model with parameter $\lambda = 1$. The training samples are the circles and diamonds and the two shapes correspond to the two classes $(+1, -1)$ to which the samples belong. The solid line is the computed separating hyperplane, while the dotted lines represent the margins. For which points ξ_i is different from 0? How would the separating hyperplane look like if $\lambda \simeq 0$?

[Solution: Exercise 4]

[Solution: Exercise 4]