WRITE FIRST NAME, LAST NAME, AND ID NUMBER (“MATRICOLA”) BELOW AND READ ALL INSTRUCTIONS BEFORE STARTING WITH THE EXAM! TIME: 1.5 hours.

FIRST NAME: ....

LAST NAME: ....

ID NUMBER: ....

INSTRUCTIONS

- solutions to exercises must be in the appropriate spaces, that is:
  - Exercise 1: pag. 1, 2
  - Exercise 2: pag. 3, 4
  - Exercise 3: pag. 5, 6,
  - Exercise 4: pag. 7, 8, 9

  Solutions written outside the appropriate spaces (including other papersheets) will not be considered.
- the use of notes, books, or any other material is forbidden and will make your exam invalid;
- electronic devices (smartphones, calculators, etc.) must be turned off; their use will make your exam invalid;
- this booklet must be returned in its entirety;
- if your solution describes material not related to the question, you will get negative points for your answer.

Exercise 1 [8 points]

Consider the problem of supervised learning.

1. Formally define when a training set is ε-representative.
2. *Provide* and *prove* the upper bound to the generalization error $L_{\mathcal{D}}(h_S)$ of the hypothesis h_S picked by empirical risk minimization when the training set is $\frac{\varepsilon}{2}$-representative (briefly motivating all the steps of the proof).

[Solution: Exercise 1]

[Solution: Exercise 1]

Exercise 2 [8 points]

Consider the *linear regression* problem with *squared loss*.

1. Within the context above, define the coefficient of determination R^2, provide an intuition for what it is capturing, and describe why it makes sense to look at R^2 instead of considering the average loss.
2. Consider the problem of feature selection in the framework described above. Provide a brief motivation for feature selection, and describe one procedure to select a model with at most k features, where k is smaller than the number d of all features, but close to it.

[Solution: Exercise 2]

[Solution: Exercise 2]

Exercise 3 [8 points]

Consider the binary classification problem with domain set $\mathcal{X} = \mathbb{R}^2$ and label set $\mathcal{Y} = \{-1, 1\}$. You decide to consider the following hypothesis class:

$$\mathcal{H} = \left\{ \mathbf{x} \to \text{sign}\left(-b + \sum_{i=1}^{3} SVM_{\mathbf{w}_i}(\mathbf{x}) \right) : \mathbf{w}_i \in \mathbb{R}^3 \ \forall i \in \{1, \ldots, 3\}, b \in \mathbb{R} \right\}$$

.

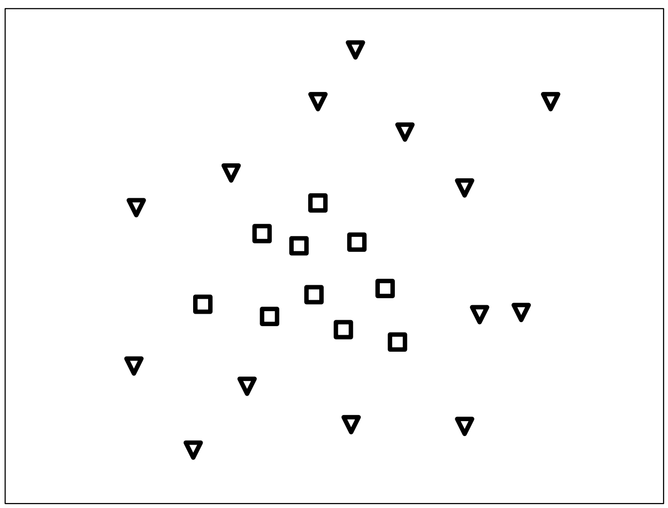
where $SVM_{\mathbf{w}_i}$ is a hard-SVM with parameters $\mathbf{w}_i \in \mathbb{R}^3$ ($\mathbf{w}_i$ includes the bias for the model $SVM_{\mathbf{w}_i}$). (Remember that the output of $SVM_{\mathbf{w}_i} \in \{-1, 1\}$.)

Assume the training data is represented by the figure below: the input $\mathbf{x} \in \mathbb{R}^2$ is given by the coordinates of the point while the label y is 1 if there is a square, and -1 if there is a triangle.

1. Is there an hypothesis $\bar{h}$ in $\mathcal{H}$ that perfectly classifies the training data? What could be a value for b in such hypothesis $\bar{h}$?
2. Plot the decision region of such a model $\bar{h}$ (i.e., where the model predicts label 1 and where the model predicts label -1) *in the figure below.*

Motivate your answers.

(***Note***: you do **not** need to find the weights $\mathbf{w}_i$ of the SVMs!)



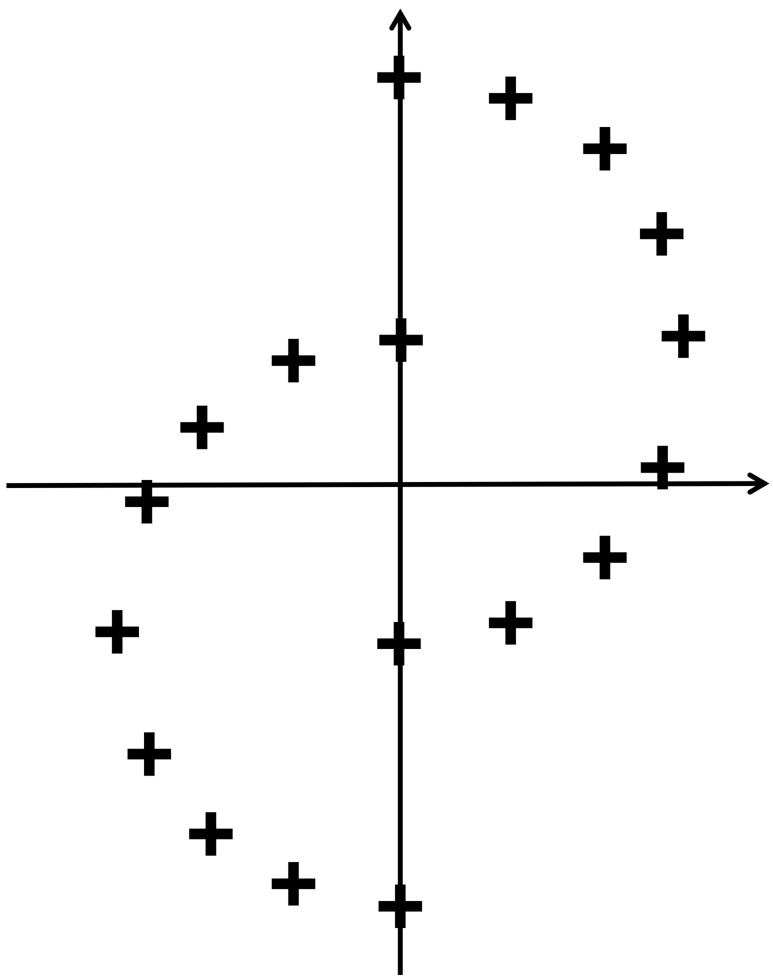
[Solution: Exercise 3]

[Solution: Exercise 3]

Consider the clustering problem.

Exercise 4 [8 points]

1. Briefly describe *single linkage-based clustering*: what is the input, what is the output, the general algorithm it employs, and a termination condition.
2. Show the output of single linkage clustering when the input is given by the points in $\mathbb{R}^2$ shown as crosses below and the termination condition is given by having the points partitioned in $k = 2$ clusters. (You can draw directly in the figure below.) Briefly describe how the algorithm reaches such output.



[Solution: Exercise 4]

[Solution: Exercise 4]