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

**FIRST NAME**: .....

**LAST NAME**: .....

**ID NUMBER**: .....

# Exercise 1 [8 points]

1. Describe the general framework of binary classification.
2. Discuss the use of the logistic regression model for binary classification and describe how it can be trained using stochastic gradient descent.

---

[Solution: Exercise 1]

---

[Solution: Exercise 1]

# Exercise 2 [8 points]

1. Introduce the concept of separating hyperplane and describe how it leads to the formulation of hard SVM.

2. In most practical situations data are not linearly separable. How can hard SVM be adapted to cope with this situation while still using a linear separation boundary? Formulate this adaptation in the form of an optimization problem, discussing the role and effect of the parameter $\lambda$ that trades the two components of the cost function.

---

[Solution: Exercise 2]

---

[Solution: Exercise 2]

---

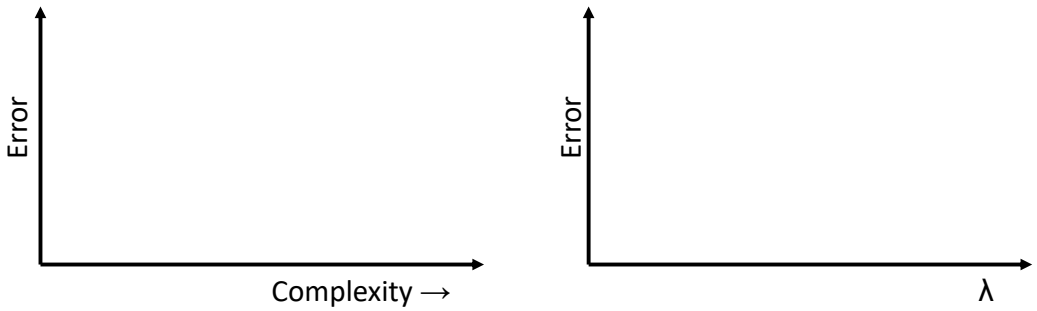
[Solution: Exercise 2]

# Exercise 3 [8 points]

With reference to the binary classification problem:

1. Describe how the true error $L_{\mathcal{D}}(h_s)$ of a predictor $h_s$ can be decomposed into approximation $\epsilon_{app}$ and estimation $\epsilon_{est}$ error. Discuss how this decomposition is related to the complexity of the selected hypothesis class and draw a plot with the expected behaviour of the approximation, estimation, and true error, as functions of the complexity of the hypothesis class.

2. Formally introduce the Tikhonov regularization model writing down the function to be optimized. In relation with the previous answer, discuss how Tikhonov regularization can be used to control the fitting-stability trade-off. In particular discuss the behaviour of the approximation, estimation, and true error, as functions of the regularization parameter $\lambda$.




---

[Solution: Exercise 3]

---

[Solution: Exercise 3]

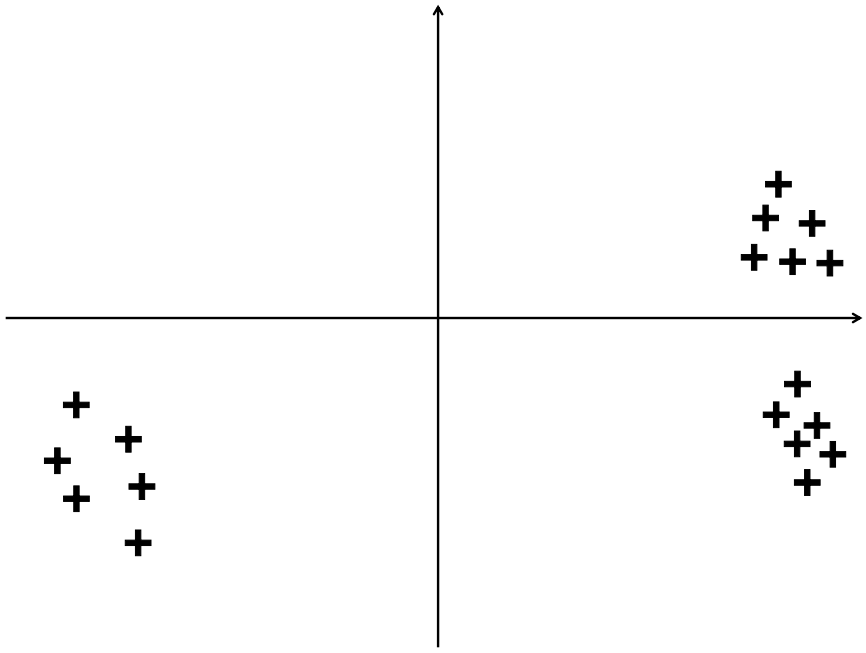
---

[Solution: Exercise 3]

# Exercise 4 [8 points]

1. Introduce the $k$-means clustering problem and present the standard (i.e., Lloyd's) algorithm for its solution.

2. Assume that you need to cluster the points shown in the figure below into $k = 3$ different clusters using $k$-means. The final solution depends on the initialization of the centroids: suggest two possible initializations, one leading to the optimal solution and one leading to a sub-optimal solution.



---

[Solution: Exercise 4]

---

[Solution: Exercise 4]

---

[Solution: Exercise 4]