# Machine Learning – B – February 11, 2020

Time limit: **2 hours**.

Last Name ..... First Name ..... Matricola .....

**Note:** if you are not doing the regular exam for ML 2019/20, write below name of exam, CFU, and academic year (when you were supposed to attend the course). Please specify also if you are an Erasmus student.

.....

---

## EXERCISE A1

Assume you are given the following dataset, representing the samples of a function $f$:

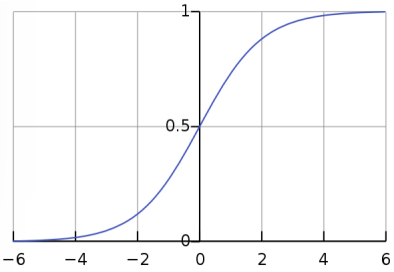
| $x_1$ | $x_2$ | $x_3$ | f |
|---|---|---|---|
| 0.6 | 3 | 1 | 4.6 |
| 1 | 2 | 3 | 2.1 |
| 4 | 4 | 1 | 10 |

1. Which technique would you use to estimate $f$?
2. Provide a mathematical formulation of the problem solved by the chosen technique.
3. Provide an example of solution using a simple dataset of your choice. (Show the solution only, you don't have to illustrate the steps followed to obtain it).

## EXERCISE A2

1. Define mathematically the problem solved by logistic regression
2. Consider the following dataset and the sigmoid function:

| $x_1$ | $x_2$ | $x_3$ | t |
|---|---|---|---|
| 0 | 0 | 1 | 0 |
| 1 | 2 | 3 | 0 |
| 4 | 4 | 1 | 1 |




Which one among the following solutions fits the data better? Why?

$$\vec{w}_1^T = (2, 0, -2) \qquad \vec{w}_2^T = (-2, -2, 4)$$

A plot of the sigmoid function is reported above. You do not need to compute explicit values of the model.

**EXERCISE B1**

1. Explain what properties a kernel function should typically satisfy.
2. Indicate which of the following kernel functions are not valid explaining why:
   (a) $k(\mathbf{x}, \mathbf{x}') = 1$;
   (b) $k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T\mathbf{x}' + \gamma)^4$;
   (c) $k(\mathbf{x}, \mathbf{x}') = \sum_i [\sin(\mathbf{x}_i) - \sin(\mathbf{x}'_i)]$;
   (d) $k(\mathbf{x}, \mathbf{x}') = \sum_i -\log(\mathbf{x}_i) \log\left(\frac{\mathbf{x}'_i}{\mathbf{x}_i}\right)$, with $\mathbf{x}_i, \mathbf{x}'_i > 0$ for all $i$;
   (e) $k(\mathbf{x}, \mathbf{x}') = 1 - \frac{|\mathbf{x}^T\mathbf{x}'|}{\|\mathbf{x}\|\|\mathbf{x}'\|}$;

**EXERCISE B2**

Consider the problem of finding a function which describes how the salary of a person (in hundreds of euros) depends on his/her age (in years), the months in higher education and average grades in higher education. A dataset in the form $\mathcal{D} = \{(\mathbf{x}_1^T, t_1), \ldots, (\mathbf{x}_N^T, t_N)\}$ is provided, with $\mathbf{x} \in \mathbb{R}^3$ denoting the input values and $t \in \mathbb{R}$ the target values (salary). Assuming that one tries to estimate this function with a deep feed-forward network:

1. Explain how the problem is formalized by writing the parametric form of the function to be learned highlighting the parameters $\boldsymbol{\theta}$.
2. Explain what are suitable choices for the activation functions of the hidden and output units of the network.
3. Explain what is a suitable choice for the loss function used for training the network and write the corresponding mathematical expression.
4. Assuming that the gradients of the loss with respect to the parameters are available, describe an algorithm for training the parameters of the network. What are the hyper-parameters of the training algorithm (if any)?

**EXERCISE C1**

1. Describe the Markov Decision Process (MDP) model used in reinforcement learning, provide its mathematical formulation, and explain the elements of the model.
2. Describe the Q-learning algorithm, referring to the mathematical formulation of the MDP given above.

**EXERCISE C2**

1. Describe the K-means algorithm in a formal way (i.e., with precise mathematical formulas and equations), including: input and output of the algorithm, its main steps, and the termination condition.
2. Draw a suitable 2-D data set for K-means.
3. Simulate the execution of K-means in such 2-D data, showing at least three steps of the algorithm and the final output.