WRITE FIRST NAME, LAST NAME, AND ID NUMBER ("MATRICOLA") BELOW AND READ ALL INSTRUCTIONS BEFORE STARTING WITH THE EXAM! TIME: 2.5 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, 5
  - Exercise 3: pag. 6, 7
  - Exercise 4: pag. 8, 9, 10

  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.

Exercise 1 [8 points]

1. Discuss which are the main ingredients of a learning problem, how learning can be formulated as on optimisation problem, and how the objective of learning can be encoded.

2. Define the concept of model class and a way to measure its complexity.

3. Discuss the role of model class complexity on the learning problem. In the context of PAC learning, provide a bound on sample complexity for finite model classes with loss function $\ell : \mathcal{H} \times Z \to [0, 1]$.

[Solution: Exercise 1]

[Solution: Exercise 1]

[Solution: Exercise 1]

Exercise 2 [8 points]

1. Describe and motivate the regression problem in Machine Learning.

2. Provide an example of linear regression problem where the hypothesis class is

$$Y = X\theta \qquad Y \in \mathbb{R}^n \quad \theta \in \mathbb{R}^d$$

   in which it is of interest to perform variable selection and discuss how this can be solved using regularisation, defining explicitly the cost function to be minimised as a function of the usual regularisation parameter λ.

3. Let λ be the regularization parameter in the sparse regression problem discussed above. Draw a typical plot (regularisation path) of how the estimated coefficients $\hat{\theta}_i$ (entries of the parameter vector θ) vary as a function the regularisation parameter λ (one line for each $\hat{\theta}_i$, $i = 1, ., d$, assuming $d = 4$).

[Solution: Exercise 2]

[Solution: Exercise 2]

[Solution: Exercise 2]

Exercise 3 [8 points]

Consider a neural network with two hidden layers, inputs x, and output y, where the first hidden layer has 5 nodes (say ξ_i, $i = 1, .., 5$) and the second hidden layer 1 node (say z_1) where

$$\xi_i = \mathbf{1}(w_{1,i}^\top x + b_i) \quad i = 1, .., 5$$

and

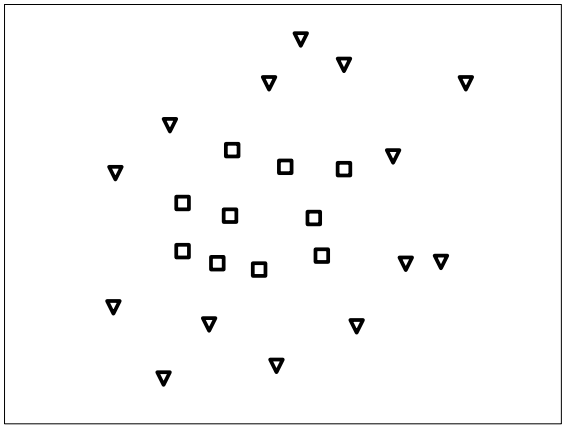
$$z_1 = \mathbf{1}(w_{2,1}^\top \xi - 4.5)$$

where $w_{2,1}^\top = [1\ 1\ 1\ 1\ 1]$, $\mathbf{1}(a)$ is the indicator function

$$\mathbf{1}(a) = \begin{cases} 1 & a \geq 0 \\ 0 & a < 0 \end{cases}$$

and $y = z_1$.

1. Draw a schematic picture of the neural network
2. Assuming the network is trained for the binary classification problem with the data depicted in figure below (the input $x \in \mathbb{R}^2$ are the coordinates of the points while the output y are the labels), say whether there is a combination of weights for which the training error is exactly equal to zero (i.e. the network perfectly classifies the training data). *Note*: you do not need to find the exact weights.
3. Interpret, and illustrate in the picure below, the two hidden layers in the context of linear classification on training data.



[Solution: Exercise 3]

[Solution: Exercise 3]

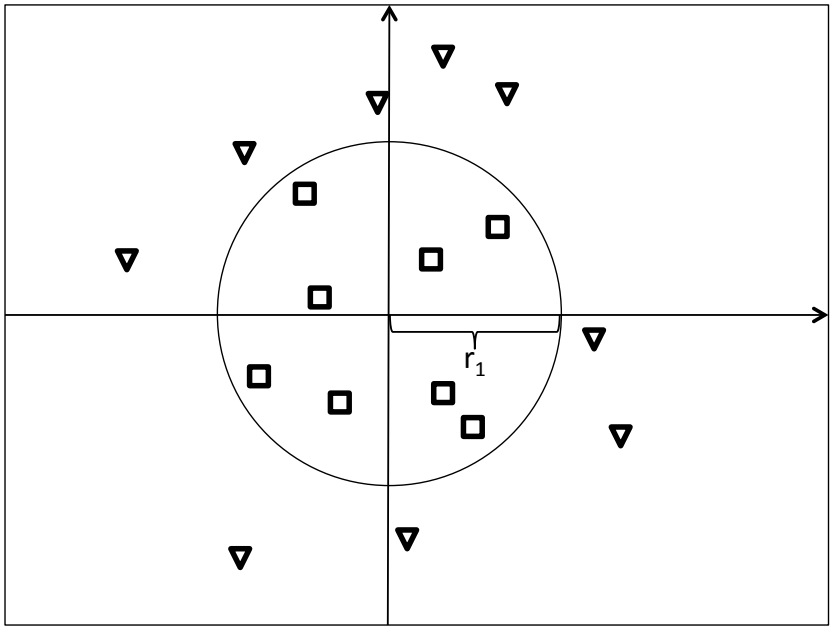
Exercise 4 [8 points]

You want to cluster the points in the figure below using k-means. with $k = 2$.

1. Is there a way to cluster the points using k-means so that in the solution the two clusters corresponds to the two sets with different marks (triangles, squares)? Given a short explanation for your answer.

2. Before applying clustering, you can apply a transformation to the dataset. Describe a transformation such that the application of k-means with $k = 2$ to the transformed datasets results in two clusters corresponding to the two sets with different marks, and plot the transformed dataset.

3. Briefly describe the execution of k-means on the transformed dataset (*note*: choose the first centers so that only few iterations are required and that the final clustering corresponds to the two sets with different marks).



[Solution: Exercise 4]

[Solution: Exercise 4]

[Solution: Exercise 4]