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
  - Exercise 3: pag. 5, 6
  - Exercise 4: pag. 7, 8

  Solutions written outside the appropriate spaces (including other paper-sheets) 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 [9 points]

1. Describe the classification problem in machine learning and one approach for its solution.

2. Give the definition of empirical and generalisation errors for an arbitrary hypothesis $h \in \mathcal{H}$, and give the definition of ϵ-representative sample $\mathcal{S}$.

3. Prove the following Lemma (motivating each step): If $\mathcal{S}$ is an $\epsilon/2$-representative sample (as defined above), then any empirical risk minimiser h_S satisfies the inequality

$$L_{\mathcal{D}}(h_S) \leq \min_{h \in \mathcal{H}} L_{\mathcal{D}}(h) + \epsilon.$$

[Solution: Exercise 1]

[Solution: Exercise 1]

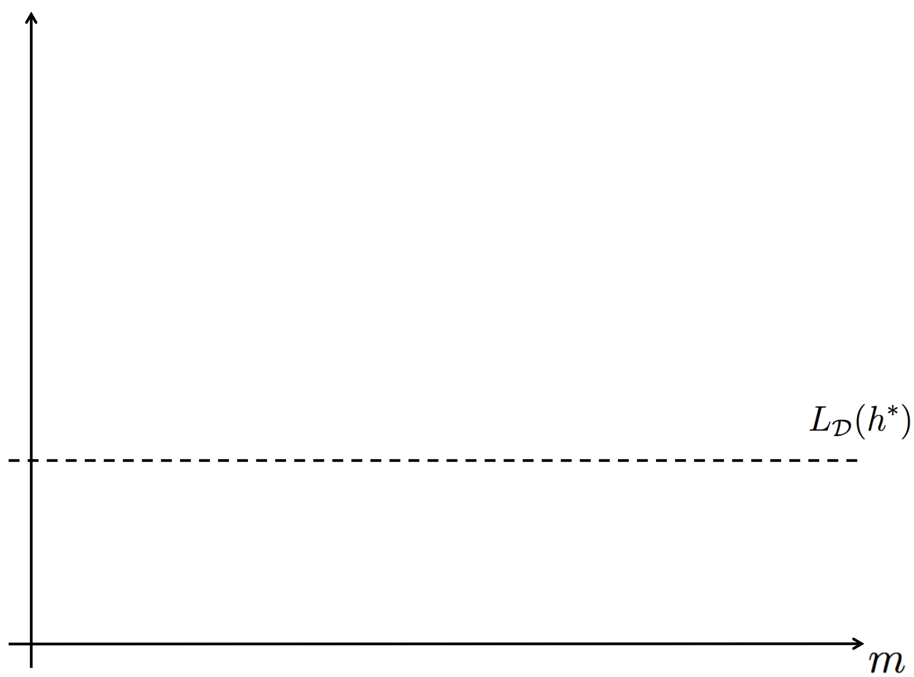
Exercise 2 [8 points]

1. Introduce and motivate the use of regularisation in regression problems.

2. Let $\mathcal{S}$ be a training set containing m i.i.d. examples, h_S be the hypothesis that minimises the empirical risk on $\mathcal{S}$, $h_{S,R}$ be the hypothesis that minimises the regularised problem defined above on $\mathcal{S}$, and

$$h^* \in \arg\min_{h \in \mathcal{H}} L_{\mathcal{D}}(h).$$

Assume the regularisation parameter (say λ) is fixed (and does not dependent on the sample size m - e.g., $\lambda = 10$). Plot below the typical behaviour of $L_{\mathcal{S}}(h_S)$, $L_{\mathcal{S}}(h_{S,R})$, $L_{\mathcal{D}}(h_S)$, $L_{\mathcal{D}}(h_{S,R})$ as a function of the training set size m.



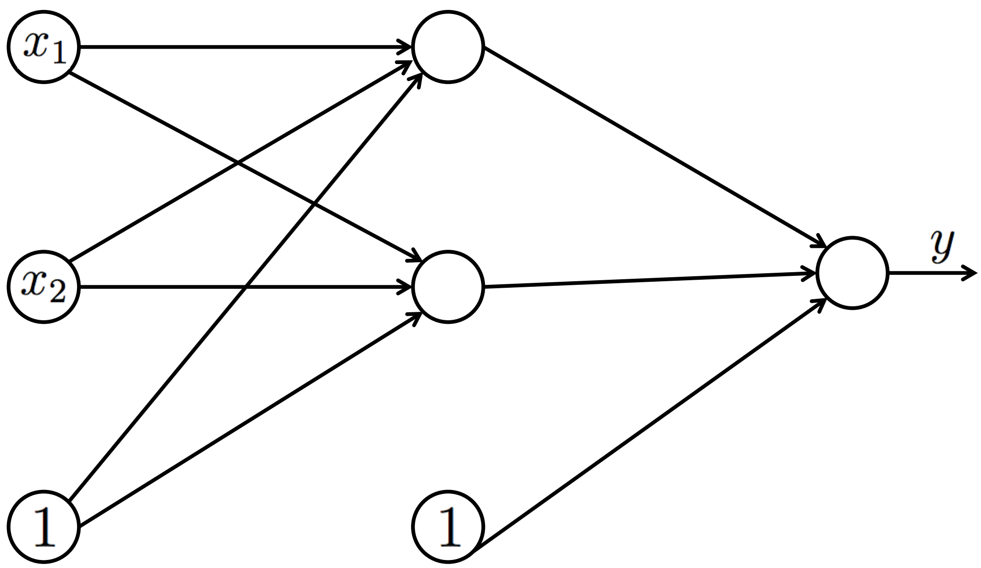
3. Describe an approach to estimate a reasonable value for the regularisation parameter λ.

[Solution: Exercise 2]

[Solution: Exercise 2]

Exercise 3 [7 points]

1. Consider the Neural Network depicted in the figure below.



Let the input variables x_1 and x_2 be binary ($x_i \in \{-1, 1\}$) and the activation function be the sign function: $\sigma(z) = sign(z)$. Assume the network weights are all constrained to take value in the discrete set $\{-1, 1\}$. Consider the training set described by the table below:

x_1	x_2	y
-1	-1	-1
-1	1	1
1	-1	1
1	1	-1

Find network weights so to that the training error is zero. (You can put the weights directly in the figure above.)

2. Using the example above motivate the fact that neural network architectures are richer than linear models.

[Solution: Exercise 3]

[Solution: Exercise 3]

Exercise 4 [8 points]

1. Introduce the clustering problem in the context of unsupervised learning.
2. Describe the k-means optimisation problem and an algorithm to solve this problem; in particular discuss whether the algorithm finds the optimal solution.
3. Introduce the Gaussian Mixture Model (GMM) for clustering points $x_i \in \mathbb{R}^d$, $i = 1, .., m$ in k classes. Argue why this can be considered a *soft* version of the *hard assignment* of points to clusters provided by k-means.

[Solution: Exercise 4]

[Solution: Exercise 4]