Sample Exercises and Exam Questions from previous years

This document contains some sample exercises in large part derived from exam questions given in previous years. Some of them are just questions about the theory while some others are exercises to be solved.

Exercise 1

1. Describe the regression task.

2. Introduce the linear regression model class under squared loss and derive the optimal solution (in terms of training error, i.e., the least squares algorithm).

3. Describe how the approach can be extended in order to avoid the problem of having too large coefficients using regularization.

Exercise 2

1. Define the clustering problem.

2. Introduce the cost function for the K-means clustering problem and describe Lloyd's iterative algorithm.

3. Mark approximately in the graph below the solution (clusters and centers) found by Lloyd algorithm for the 2 clusters (K = 2) problem, when the data ($x_i \in \mathbb{R}$) are the crosses in the figure below and the algorithm is initialised with center values indicated with the circle (cluster 1) and triangle (cluster 2) shown in the figure.

Exercise 3

1. Describe the classification task

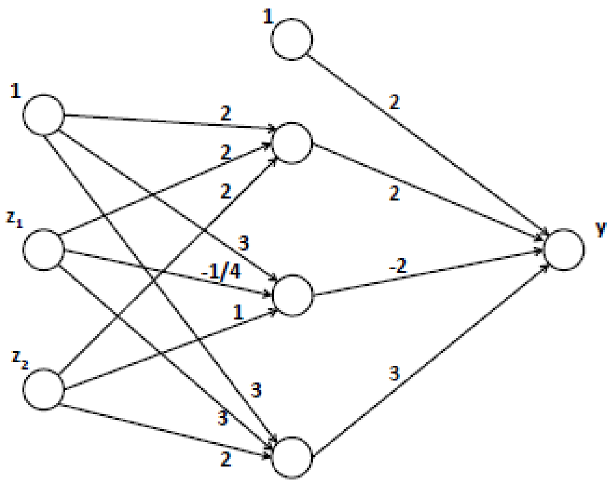
2. Describe logistic regression (model, cost function and classification rule; no need to derive the details of optimization algorithm)

Exercise 4

Consider the neural network in the figure and assume the activation function σ(x) is defined as:

$$\sigma(x) = \begin{cases} 1 & x \geq 1 \\ x & -1 \leq x < 1 \\ -1 & x < -1 \end{cases}$$

Compute the value of the output y when the input **z** is **z** = [1 3]

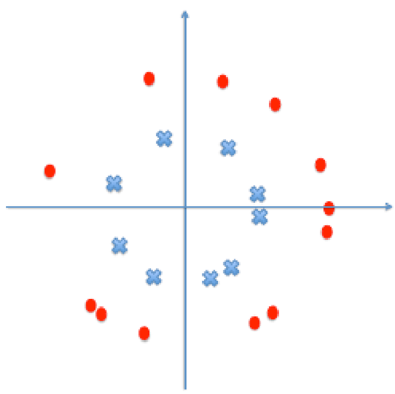


Exercise 5

1. Introduce the concept of Kernel and its use in SVM for classification.

2. Consider the configuration of training data points (crosses for class 0 and circles for class 1) in the figure below and a scalar function (feature map) $\Phi(\cdot)\colon \mathbb{R}^2 \to \mathbb{R}$ such that the data become linearly separable after the map Φ has been applied.

3. Relate the map Φ to a kernel

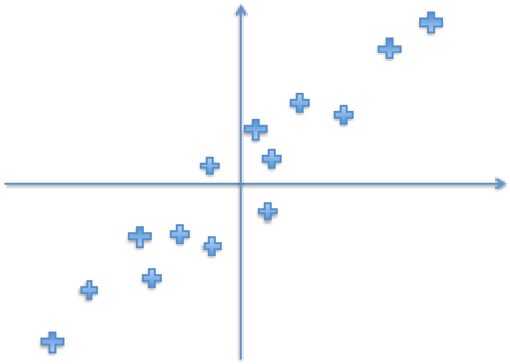


Exercise 6

1. Let $X = [\boldsymbol{x_1} \dots \boldsymbol{x_n}]\,, \boldsymbol{x_i} \in \mathbb{R}^p$ be the data matrix. Introduce the Principal Component Analysis in the context of unsupervised learning.

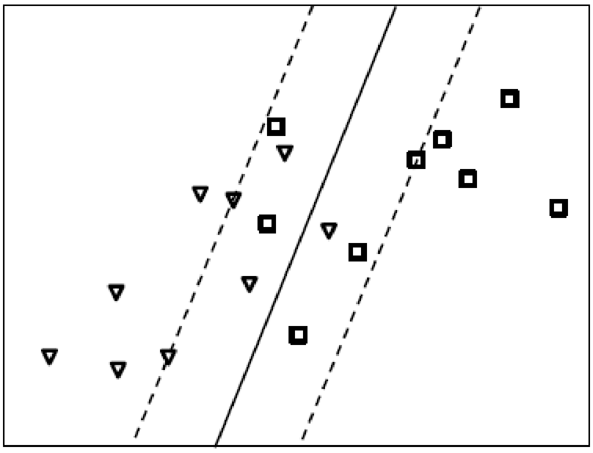
2. With reference to the figure below draw approximately the first and second right singular vectors of X.

3. Describe how PCA can be used in the context of linear regression to reduce the complexity of the model.



Exercise 7

1. Describe the linear support vector machine for classification in the case of non linearly separable data.
2. The figure shows the results (separating hyperplane and margin) of linear SVM for binary classification on the data points (in $\mathbb{R}^2$) in the figure, where the class of each point is represented by its shape (triangle or square). Mark with circles the misclassified points and draw the segments which length corresponds to the non-zero slack variables.
3. Discuss how the solution (margin width and slack variables ξ_i) changes if the value of C in the objective function $\frac{1}{2}||\beta||^2 + C\sum_i \xi_i$ increases.



Exercise 8

1. Consider a supervised learning problem, describe the concepts of training and generalization errors.
2. With reference to item 1 above, how would you state the final goal of supervised learning?
3. What role does k-fold cross validation play in estimating the errors (training and generalization) in item 1 above?

Exercise 9

1. Consider an hypothesis class $\mathcal{H}$. What do you need to show in order to demonstrate that *VCdim*($\mathcal{H}$)=*d* ?
2. Consider right triangles in the plane with the sides adjacent to the right angle both parallel to the axes and with the right angle in the lower left corner. What is the VC-dimension of this family?

Hint: Recall the axis-aligned rectangle demonstration

Exercise 10

Consider a dataset with the following six 1-dimensional points (the first element of the couple is the value *x* while the second is the label *y*):

$\{(x_i, y_i)\} = \{(-3, +1), (-2, +1), (-1, -1), (1, -1), (2, +1), (3, +1)\}$

Consider mapping these points to 2 dimensions using the mapping $\phi: x \rightarrow (x, x^2)$
Which is the maximum margin decision boundary? Which is the corresponding margin?