Tutorial 3

Machine Learning and Big Data for Economics and Finance

List of activities

I. Complete the list of exercises in this tutorial.

II. Complete **Section 3.6 Lab: Linear Regression**. Section 3.6.6 to 3.6.7.

Exercise 1. Consider the following sample of the three random variables X_1, X_2 and Y:

Obs.	X_1	X_2	Y
1	1	2	o
2	1	3	o
3	-3	1	o
4	2	2	x
5	3	2	x
6	4	1	x
7	4	3	x

Table 1.

1. In the input space, compute the distance between each point and $\boldsymbol{x}_0 = (1, 1)$.
2. Predict Y given $X_1 = 1$ and $X_2 = 1$ using K-nearest neighbor classification for $K = 1$ and $K = 3$.

Exercise 2. Load the data included in the file `MC1.csv`. The file contains a sample of size $n = 1000$ from a random variable X.

The data generating process for a new random variable Y is given by

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where $\beta_0 = -1$, $\beta_1 = 5.1$ and $\varepsilon \sim N(0, 1)$.

1. Generate a sample of size n from Y.
2. Assuming you don't know the parameters behind the data generating process, compute the least squares estimates for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}$ (the standard deviation of the error term).
3. Generate 100 different samples of size n of Y and for each compute $\hat{\beta}_{0,m}$, $\hat{\beta}_{1,m}$ and $\hat{\sigma}_m$ where $\hat{\beta}_{0,m}$ is the estimate of β_0 in sample m for $m = 1, ..., 100$.
4. Using the values $\hat{\beta}_{0,m}$, $\hat{\beta}_{1,m}$ and $\hat{\sigma}_m$, compute
    a. The sample averages of each of $\hat{\beta}_{0,m}$, $\hat{\beta}_{1,m}$ and $\hat{\sigma}_m$.
    b. The sample variance of $\hat{\beta}_{0,m}$.
    c. The sample variance of $\hat{\beta}_{1,m}$.
    d. The sample covariance of $\hat{\beta}_{0,m}$ and $\hat{\beta}_{1,m}$.
5. Plot $\hat{\beta}_{0,m}$, $\hat{\beta}_{1,m}$ and $\hat{\sigma}_m$ and discuss.
6. Compare the results of the small simulation exercise with the formula $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.