General Instructions:

1. Please set a seed so your results are reproducible.
2. Make sure to comment your code sufficiently.
3. I will not be posting sample solutions. I am always happy to talk about problems in your solutions. If you did not manage to do a problem set, ask me.

Exercise 1:

Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{with } \boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{I})$$

X_1 is a constant, $X_2 \sim \mathcal{N}(\mu = 0, \sigma^2 = 1.5)$. The error term is generated as $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 = 10)$. The true DGP uses as $\boldsymbol{\beta} = (5 - 0.5)$ and $N = 1000$.

a) Generate a training sample $(x_i, y_i)^T$ using the above specification.

b) Generate a test sample (x_0, y_0) using the same N.

c) Calculate the OLS estimate for $\hat{\beta}$.

d) Calculate the training MSE and the prediction error using the expressions given below for these two individual samples.

e) Using the training sample from above, calculate the training MSE and the avg. prediction error when sequentially increasing the degree of the polynomial for X_2 from zero (constant only) to four in the estimation equation.

Exercise 2 (Simulation Study)**:**

Using the general set-up from above

a) Repeat the simulation 1000 times, initially setting the seed at set.seed(100).

b) Calculate the average training MSE and the average prediction error using the expressions given below and store the results in a vector.

c) Plot the avg. training MSE and the avg. prediction error in two separate plots and discuss your results. Be sure to complete this simulation for the set-up described in 1 e).

d) Along which margins could you vary parameters of the initial simulation set-up and what would be your intuition based on the theoretical properties of the considered objects of interest?

Training MSE

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2 \tag{1}$$

where $\hat{f}(x_i)$ is the prediction $\hat{f}$ gives for the $i'th$ observation.

Average prediction error

$$\text{Ave}\left(\hat{f}(x_0) - y_0\right)^2. \tag{2}$$