General Instructions:

1. Set a seed so your results are reproducible.
2. Make sure to comment your code sufficiently.
3. I will **not** be posting sample solutions. If you did not manage to do a problem set, you can always ask me during the office hours.

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 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ using the above specification.

b) Generate a test sample $\{(\mathbf{x}'_i, y'_i)\}_{i=1}^N$ using the same N.

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

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

e) Using the training and the test samples 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 five in the estimation equation (i.e. include X_2^2, X_2^3, X_2^4 X_2^5 as regressors).

Exercise 2 (Simulation Study)**:**

Using the general set-up from above

a) Repeat the simulation 1000 times, each time setting the seed at $100+$ the number of the simulation run.

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 one plot 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

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\rho}\left(\mathbf{x}_i\right)\right)^2 \tag{1}$$

where $\hat{\rho}\left(x_i\right)$ is the prediction $\hat{\rho}$ gives for the ith observation.

Average prediction error

$$\text{Ave}\left(\hat{\rho}\left(\mathbf{x}'_i\right) - y'_i\right)^2 . \tag{2}$$