Faculty of Sciences
Department of Mathematics, Statistics, and Computer Science

First Practical Assignment

Submission Deadline: April 6

Problem 1. In this section, you will work with two one-dimensional datasets and perform data fitting using the least squares method. For this purpose, two datasets named data1.csv and data2.csv have been provided. After removing noise from the data to smooth the function, fit a polynomial of appropriate degree to each dataset.

Problem 2. Various regularization terms are used in the regularized least squares framework. One of the most well-known is the L2-norm of the parameter vector, referred to as ridge regression.

The basic least squares problem is:

$$\min_x ||Ax - b||^2 \tag{1}$$

In ridge regression, the problem becomes (where D is the identity matrix):

$$\min_x ||Ax - b||^2 + \lambda||Dx||^2 \quad (\lambda > 0) \tag{2}$$

Implement ridge regression using the diabetes.csv dataset (measuring diabetes progression in patients) and answer:

- How does the parameter λ affect the solution x? Test with different λ values.
- How does regularization influence the final solution? When is it preferred over the base case?
- Research lasso regression and elastic net. Compare their advantages over ridge regression.

Problem 3. Weighted least squares is used when data points should have unequal influence on the solution. The objective function is:

$$\min_x \frac{1}{n} \sum_{i=1}^{n} w_i (A_i^T x - b_i)^2 \quad (w, b \in \mathbb{R}^n) \tag{3}$$

For diagonal weight matrix W, the matrix form is:

$$\min_x \frac{1}{n} (Ax - b)^T W (Ax - b) \tag{4}$$

Implement this method with california_housing.csv for three weight vectors w:

(a) Sampled from a uniform distribution (range [0.5, 3]).

(b) Sampled from a multinomial distribution ($p_i = \frac{1}{n}$).

(c) Sampled from a Dirichlet distribution (unit parameters).

Run each case for 100 iterations (use np.random.uniform, np.random.multinomial, and np.random.dirichlet) and average the results. Compare the performance.