Linear Regression Analysis

Dataset

The dataset and it's description is found here `https://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant`. Your task will be to predict the EP(power consumption) values and the loss over a test set. The test set in this case will be the last 2000 examples in this dataset, with the rest for training. You can use *pandas* to read data from the *.xlsx* file.

Part A: Normal Equations Method

You will be implementing the solution detailed in section 3.1.1 of Bishop. Refer to equations 3.15-3.17.

Part B: Gradient Descent Method

You will define the loss function to be half the sum of squared errors over the training set and minimize the loss function using Gradient Descent. Choose an appropriate learning rate and stopping criteria.

Part C: Gradient Descent with Regularization

As an extension to Part B, you will experiment with L1 and L2 regularization. Using a part of the training data for validation, determine the best value for the regularization coefficient.