ID5030- Machine learning for Engineering and science applications

Homework 1 – Linear Regression

Assignment Given on : Jan 20, 2023

Due Date : Jan 26, 2023 (Online submission)

Context : The purpose of this assignment is to get familiar with some of the basic packages used in Machine Learning and get used to the data utilization pipeline. *The expectation of this assignment is that you will be able to use the Linear Regression module within sci-kit learn as a black box as demonstrated within the Tutorial class.* You don't need to understand linear regression in order to do the assignment. The understanding of linear regression itself will be provided in detail within the next several classes.

Tools you will might get familiar with with this assignment:

1. The Python programming language
2. Packages to be used: numpy, matplotlib, seaborn, pandas and sci-kit learn

==========

Consider the following Concrete Compressive Strength **dataset ((**For the downloading of the dataset please refer to this Link.)

1. The first 8 columns indicate features such as cement quantity, age, etc. These are to be used to predict the final column 'Concrete compressive strength(MPa, megapascals)**'** which is the Target variable (that is, the predicted variable). Divide the dataset into Training, Test datasets using a 70:30 ratio.
2. Obtain a linear regression fit for predicting the compressive strength from fitting the linear regression model on just the Training dataset. Use R-squared value (coefficient of determination) and mean Square error as evaluation metrics.
3. Using the trained model, predict the value of 'Concrete compressive strength' for the given Test data in the submission.csv file.

4. Obtain the following values.
    a. Mean Squared error and R-squared value for training data and test data respectively.
    b. Plot the scatter plot for y_{true} against $y_{predicted}$ for training and test data respectively.

5. **Upload** the sumission.csv file and submit the code as a single .ipynb file. We will provide details of the submission process later in class.