CSE-4155 Introduction to Machine Learning Lab
4th Year 1st Semester, 2024 (Held in March, 2025)
Department of Computer Science and Engineering, University of Dhaka
Course Instruction: Dr. Muhammad Ibrahim, Associate Professor

Lab 2 (27-4-2025)

In this lab, you shall implement:
1) Linear regression with multiple variables.
2) k-fold cross validation.
3) Polynomial regression with a single variable.

IMPORTANT: You will be evaluated on the cleanliness and modularity of your code. This is the code that you'll use throughout the entire course, so make it as standard as possible.

(A) Linear regression with multiple variable

Load the supplied data file `data_02a.csv` that contains 4 features and a target variable and 9568 examples.

Plot the individual features against the target variable. There will be 4 plots in total.

Then, apply linear regression on this dataset.
    Show training and validation error curves.
    Report the best validation error, training error, and the values of the learnt parameters.
    Compare performance before and after data normalization/scaling.

(B) Cross Validation
Implement 5-fold cross validation on the above problem. Compare the best parameters between this and the previous settings.

(C) Polynomial Regression

Load the supplied data file `data_02b.csv` that contains one feature and a target variable. Plot the feature.

Tune the parameter *d* (the degree of the polynomial) with *d* = 1, 2 and 3. Find the best hypothesis based on the validation set error. You can use 1-fold cross validation here for the sake of simplicity.
    Plot all the three fitted curves onto a single graph and a bar plot with the three validation errors.
    Report the learnt parameters with the best *d* (found after tuning above).
    Draw, as before, the training and validation error curves against iterations for all three *d* values.