**Guide to Engineering Data Science**

**SUBS6397**

**Homework 6**

**Linear Regression**

Purpose: Use linear regression techniques to find the weights (coefficients) from a set of input data and known output parameters (labels).

The data is about an iris data sets with known feature set and class labels. The key is to classify the inputs features data to a particular iris flower type.

Datasets in Blackboard: “ElectricPowerData”, "Iris\_DataSet.csv"

1. Display basic statistics for each feature in the dataset (min, max, mean, median and standard deviation) - both datasets
2. Investigate data quality (missing values): summarize missing values for each attribute/feature – both datasets
3. Impute missing values using median, if needed. Show proof?
4. Scale the variables (Standardization, normalization techniques etc.) – “Iris\_DataSet.csv” only (show a print of the X matrix)
5. Obtain Multi-Variable Linear Regression expression – “ElectricPowerData”
6. Perform a classification analysis with the Perceptron Linear Regression model.
7. Perform a classification analysis with the Adaline Linear Regression model.
8. Perform a classification analysis with the LOGISTIC Regression model.
9. Bonus - Compare the accuracies of each model.