CEEL 82B, Data Science, 2022

Lab 1: Study of Linear Regression

**Objective:** To carry out linear regression (including multiple regression) and build a regression model using Python Platform

Case studies to consider:

1. Estimating horse fatalities from colic- use logistic regression to try to predict if a horse with colic will live or die
2. Credit Score prediction – a Model to predict the probabilities of default. Use Linear Regression to predict the probabilities of default and assign credit to potential borrowers

**Outcomes:**

1. To learn how to define, fit, and use a model in Python
2. To interpret the results

**System Requirements:** Linux/MaC/Windows OS with Anconda platform with *Pandas, numpy, scipy, matplotlib, seaborn* and *scikit-learn* ML library.

**Part-A: S**imple linear regression and Multiclass linear regression with data preprocessing (Handling NA values)

Use the case study relevant csv and files to build the models and evaluate the models.

**General Steps:**

1. Load the dataset (Use pandas )
2. Data Preprocessing (Handling NA values)
3. Exploratory Data Analysis (understanding the relationships between the variables with help of plot, scatter-plot, enery-plot etc) Use matplotlib
4. Data Partition (80% for training and 20% for testing) (Use scikit-learn)
5. Build the model (use scikit learn)
6. Summarize the model.
7. Prediction
8. Evaluate the model
9. Tuning the model

**[1 Hour]**

**Part-B:** Logistic Regression

Follow the general steps to carry out logistic regression as mentioned in Part-A.

Calculate the performance metrics-Accuracy, Miss-classification rate, Receiver operating characteristics.

**[1 Hour]**

**Conclusion:** (Write in own words)

**Note:** Complete your write-up with conclusion and upload your outputs on your github account