California State University Long Beach
BME 455 Spring 2021
Lab 5
Instructor: Elsa Harris

In this lab you will work on Classification using Logistic Regression

Part 1: Work on the python code for chapter 7 from “Python Machine Learning” by Wei-Meng Lee.
Part 2: In this part you will use LR as a classifier for diabetes prediction. Use part 1 as a guide.
Part 3: In this part you will use the diabetes dataset and create a KNN classifier. Tune the parameter K to find the optimal K for the classification task. Then perform k-fold cross validation, find the best k that minimizes the misclassification rate. Use part Lab 4 part 1 as a guide for that
Part 4: For the KNN algorithm with the optimal K in part 4, calculate the accuracy, the precision and the recall. Also graph the ROC and estimate AUC. Use part 1 as a guide. Compare the KNN and the LR algorithms. Which one performs better as a classifier for diabetes prediction. Include a table that compares the accuracy metrics of the two algorithms

When done please submit for each part of the lab:
a. A pdf printout of your complete code. Please run the code and make sure the outputs (numerical values, graphs, etc) are printed in the file you submit
b. A Python notebook file

Name each submission file as Lab5_PartX_FirstName_LasstName