Assignment 9-B

Due date: 27th February 2020

Guidelines for submitting assignments

1. Assignment document to be submitted as a .pdf file with final results properly written.
2. In case of a hand-written solution, ensure that it is legible, properly scanned and converted to pdf. Other formats such as .jpeg are not allowed.
3. In case of a typed report, please ensure a minimum font size of 14.
4. If you have multiple files to be submitted, please zip it and upload.
5. The files should be named with the following convention: 'firstName_assignment#.pdf/.zip/. py'. In case of multiple files for each problem, please follow 'firstName_assignment#_question#.py'
6. Script files should be submitted in. py format only. Other formats such as .pdf or .txt are strictly not allowed.
7. Standard programming practices should be followed. The variable names should either be camelCased or underscore_separated and should signify their intent or purpose. Variable names such as a, b, a1 etc. are not entertained.
8. If you are relying on additional libraries for your script to run, provide information of the same at the start of the script in the comment block.
9. Marks will be deducted if the above guidelines are ignored.

Logistic Regression & KNN

Automotive Service Study: Problem Statement

- An automotive service chain is launching its new grand service station this weekend and they offer to service a wide variety of cars
- The current capacity of the station is to check 315 cars thoroughly per day
- As an inaugural offer, they claim to freely check all cars that arrive on their launch day, and report whether they need servicing or not!
- Unexpectedly, they get 450 cars. The service men won't work longer than the working hours, but the data analysts have to! Can you save the day for the new service station?

How can a data scientist save a day for them?

- He has been a data set which contains some attributes of car that can be easily measured and won't require much time and a conclusion that if service is needed for that or not. - "serviceTrainData.csv"
- Now for the cars they cannot check in detail, they measure those attributes- "serviceTestData.csv"
- Use knn and logistic regression classification techniques to classify the cars they cannot test manually and say whether service is needed or not