ANALYSIS REPORT

Name: Kumar Raju Bandi

Roll Number: IWM2017502

The K-nearest neighbors model was implemented from scratch for both Regression and classification for different values of K using the dataset of two exam results.

Colab Link :
https://colab.research.google.com/drive/1SLoeL79BXPIE7EIWLby3o7JmYttJ5hMy?usp=sharing

KNN Classification

Firstly, KNN classification was implemented and the results were analysed.
It was observed that K = 5 was giving better results than K = 3.
The accuracy observed for K = 5 was 100% while the accuracy observed for K = 3 was 80%.

The accuracies for different K from 1 to 10 was also analysed and the result of the same has been shown at the end of this report.

The KNN Classification was performing well for the given dataset with the test and train ratio being 3:1. That is the train data was 75% of the whole dataset and test data was 25% of the whole dataset. I have also generated examples which can be found in the implementation file.

KNN Regression

Firstly, KNN regression was implemented and the results were analysed which is as follows:
It was observed that K = 5 was giving better results than K = 3.

To convert regression to 1 or 0 label, I have used a threshold of 0.5. So, if the regression returned a value greater than 0.5 than, the predicted value is 1, else 0.
Using this the accuracy was calculated.
Coincidentally, the accuracies for classification and regression were exactly the same because the threshold value made sure that the majority label of the k nearest neighbors is given as output.

The accuracy observed for K = 5 was 100% while the accuracy observed for K = 3 was 80%.
The KNN regression was performing well similar to classification for the given dataset.

I have also generated examples which can be found in the implementation file.

I also was motivated to analyse the K vs Accuracy graph and the result was as follows:
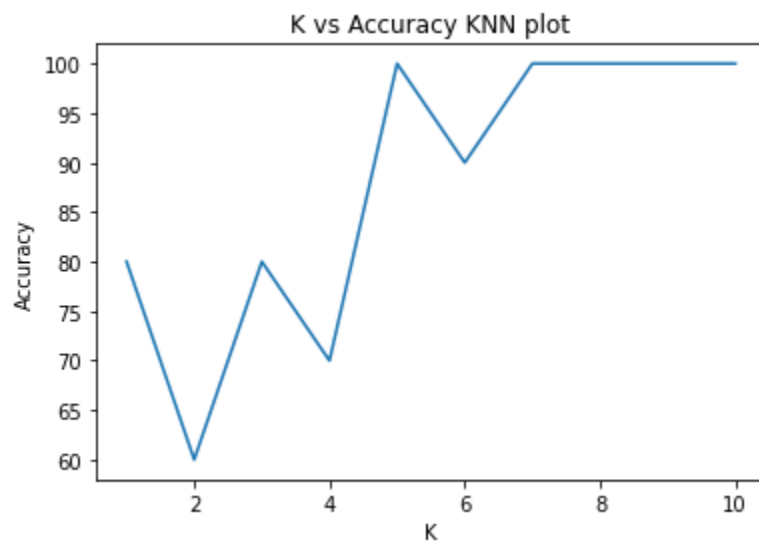
From the above graph it is evident that the values, K = 5,7,8,9 and 10 are giving the highest accuracy.