

Practical Machine Learning

Assignment I

Part 1

knn.py contains all functions and program to compute the accuracy of the knn algorithm implemented. The program is executed using Pycharm on windows PC.

The calculateDistances() function calculates the Euclidean distance between the query point and training data instances. This calculateDistances() function is implemented in knn-Algorithm. The knn.py code provides accuracy of the algorithm implemented.

The classify() function manipulates the nearest neighbour and classify the predicted class either benign or malignant. The sixth column of the data is the class. Slicing method is used to take first five column data to manipulate the neighbours. Used k nearest neighbours and the classification to classify the data point.

The function calculateDistances() used argsort sorting method to sort the calculated distance for the array.

After executing the code for k=3, The accuracy of the algorithm is 70.31

Part 2

The performance of the k-NN algorithm varies based on the k value.

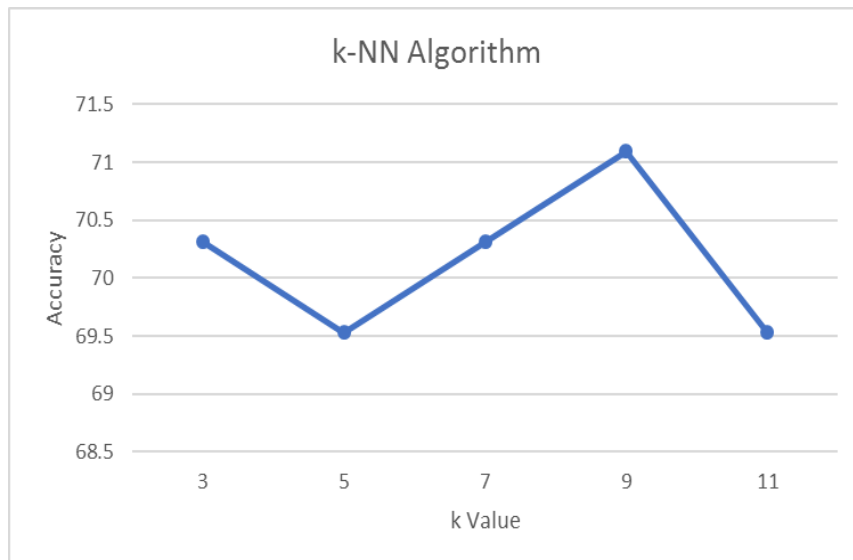
k=3, the accuracy of the algorithm 70.31

k=5, the accuracy of the algorithm 69.53

k=9, the accuracy of the algorithm 71.09

k=11, the accuracy of the algorithm 69.53

By varying the value of k, the accuracy varies. It seems for k=9, the accuracy seems high and the algorithm performs better. If alternate distance measurements are used. The accuracy of the algorithm differs.



Part 3

IBL.py contains all the functions and program to execute the accuracy of the IBL algorithm using R Square metrics.

- The R2 metrics provides the statistics of how best the model fits.
- The value of R2 ranges between 0 and 1.
- If new variables are introduced inside the model to value of R2 increases.