Report :: TIPR Assignment - 1

Prasanna Patil
11/2/19

Python version 2 is used to implement the code.

## Task 1

Generated a random matrix of N x D with Gaussian distribution. The data matrix is then multiplied with random matrix and each entry is divided by $\sqrt{D}$.
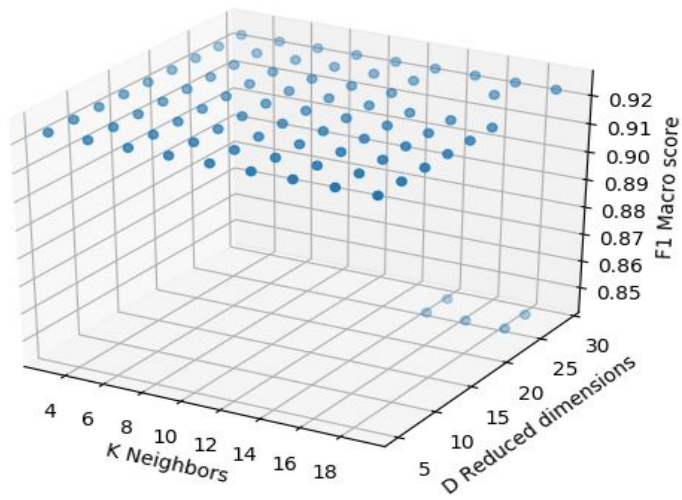
Task 2

Implemented Bayes classifier for 3 different distributions: Multinomial, Multivariate Bernoulli, Gaussian using their respective maximum likelihood parameter estimation formulas, as covered in lectures.
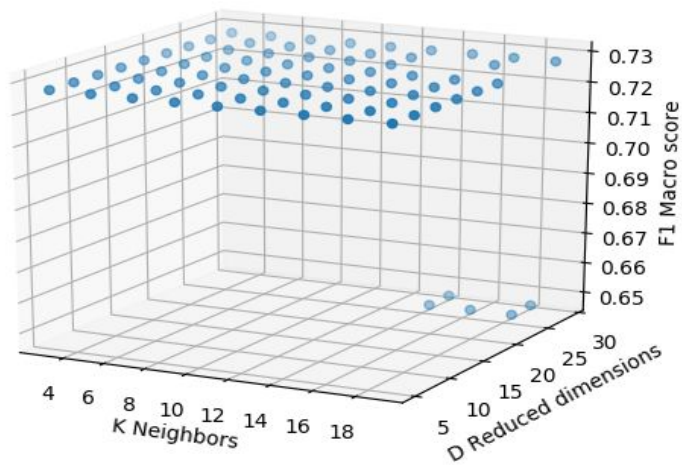
Task 3

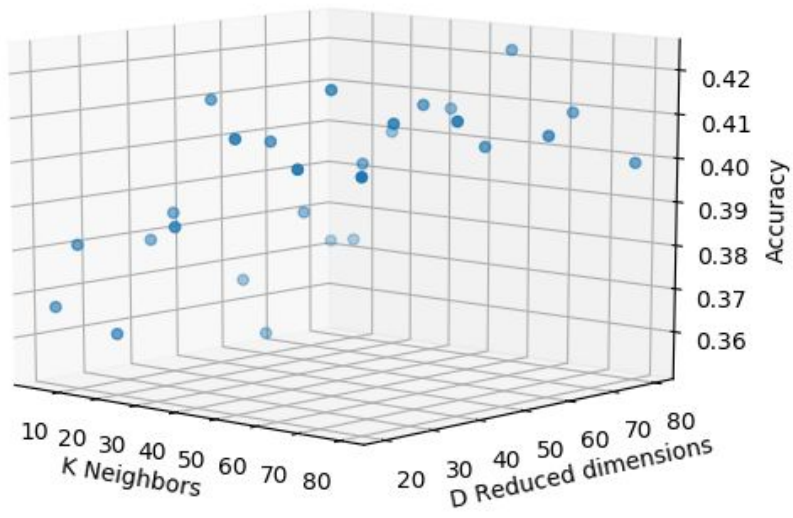Dolphins data set - Accuracy plot.
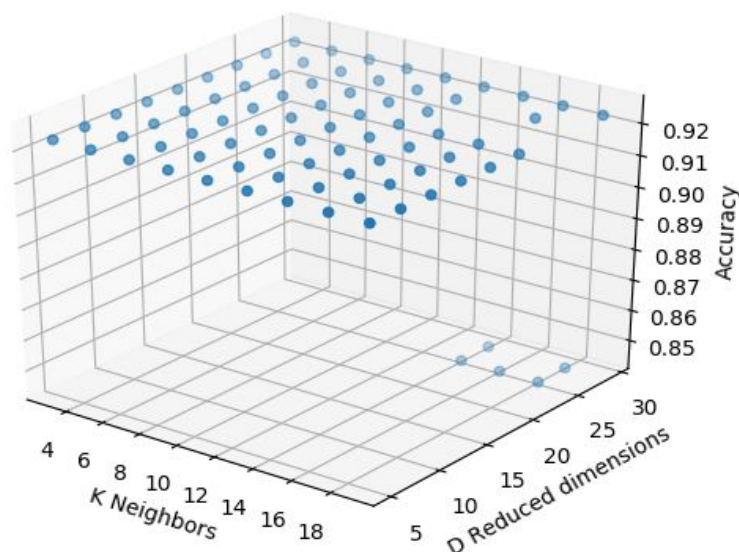
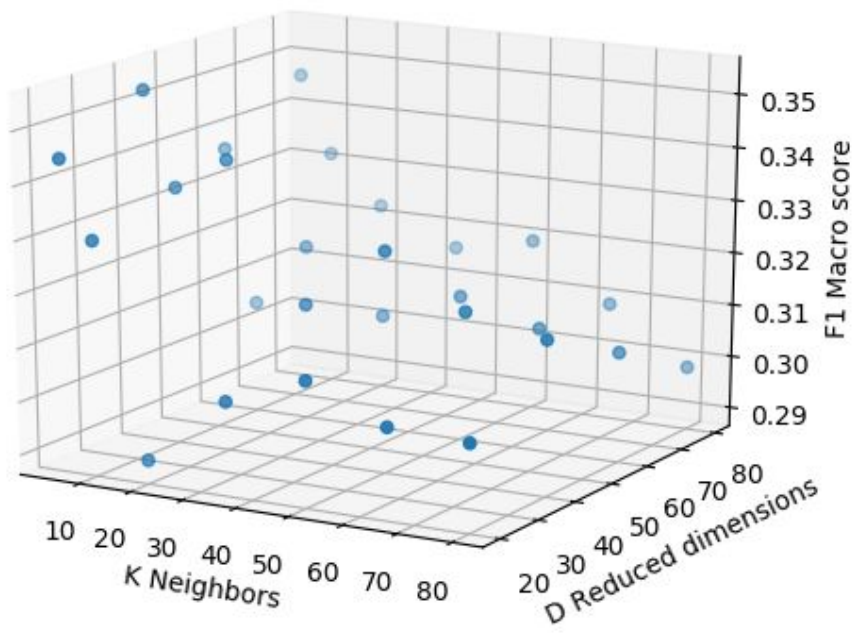Dolphins data set - F1 Micro score
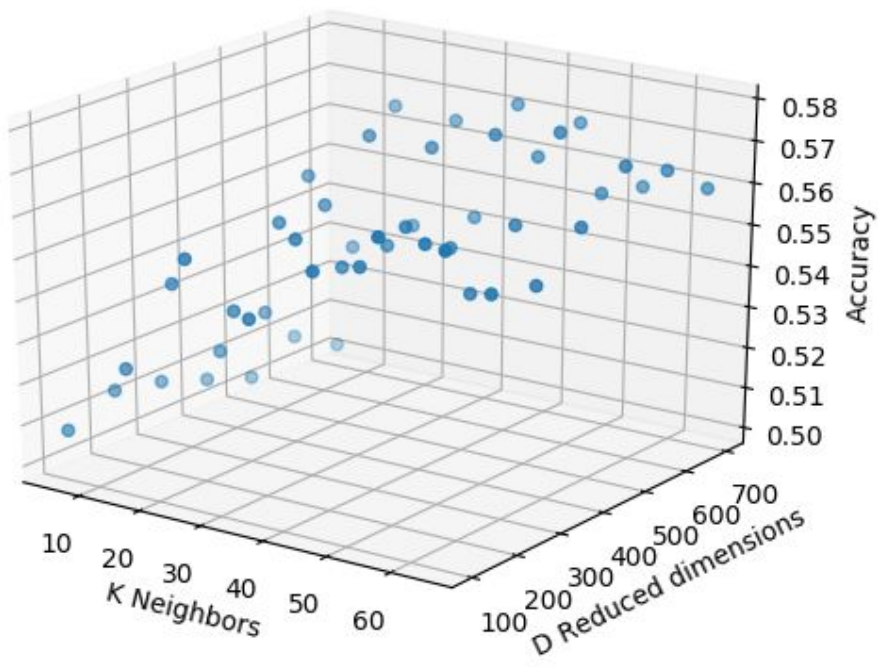
Dolphins dataset - F1 Macro score
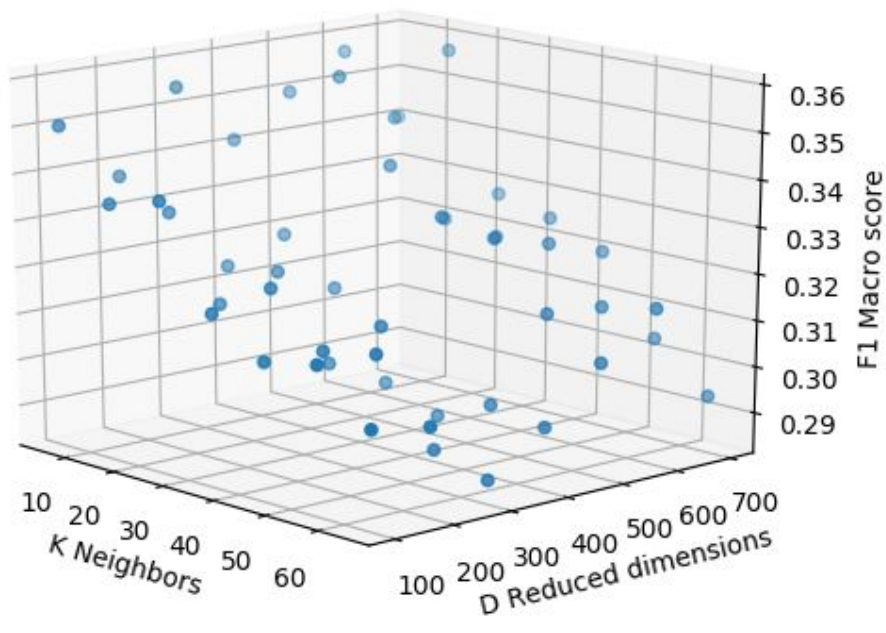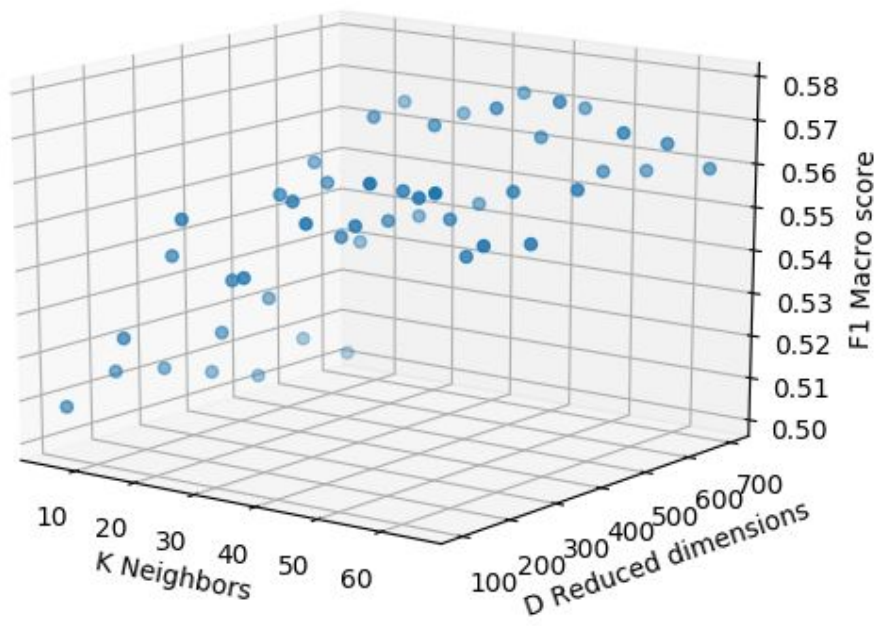
Pubmed dataset - accuracy

Pubmed dataset - F1 Macro

Twitter dataset - Accuracy

Twitter dataset - F1 Macro

Twitter dataset - F1 Micro

Task 7

Ac

Task 6

Implemented 2 variants of LSH: (1) Hamming distance, (2) E2LSH for L2 distance. Note that in E2LSH, sometimes points fail to classify in any of the hash bins, hence such points are ignored during calculation of accuracy and F1 Score.
curacy with PCA (n = 500) is 50% with F1 macro score 0.34.
Accuracy with LSH is 57%.