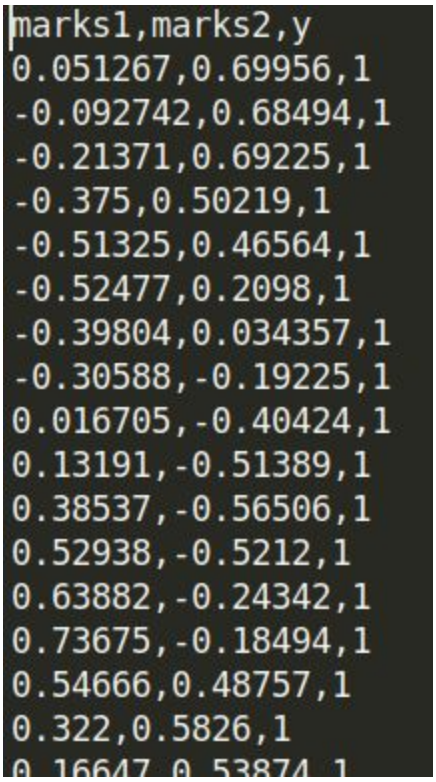
Introduction to Machine Learning

Analysis report on **Assignment 5 Q3**

Name - Prateek Mishra
Roll Number - IIT2018199

Question 3

Dataset : Assignment Marks(in CSV format)
118 rows, 3 columns



```
marks1,marks2,y
0.051267,0.69956,1
-0.092742,0.68494,1
-0.21371,0.69225,1
-0.375,0.50219,1
-0.51325,0.46564,1
-0.52477,0.2098,1
-0.39804,0.034357,1
-0.30588,-0.19225,1
0.016705,-0.40424,1
0.13191,-0.51389,1
0.38537,-0.56506,1
0.52938,-0.5212,1
0.63882,-0.24342,1
0.73675,-0.18494,1
0.54666,0.48757,1
0.322,0.5826,1
0.16647,0.53874,1
```

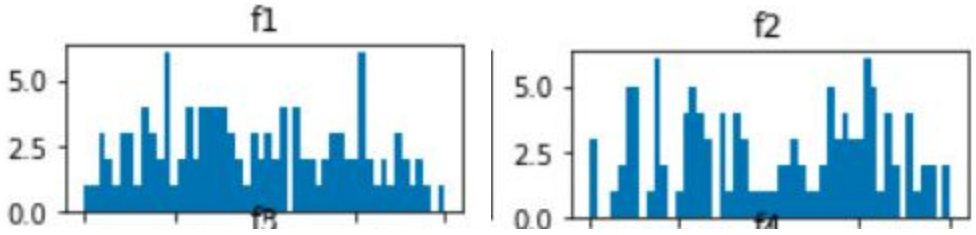
The following steps were taken to make a comparative study of Gaussian discriminant analysis with and without using box muller transformation.

Data Preprocessing(without box muller):

- Data was splitted into features and prediction(X and Y)
- The data was shuffled.
- Data was feature scaled
- Data was splitted into 70% training and 30% test set.
- I introduced higher powers of the data to improve on the accuracy of prediction.

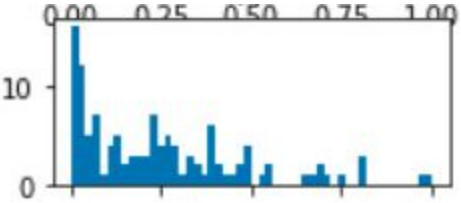
Data Preprocessing(with box muller)

- Data was splitted into features and prediction(X and Y)
- The data was shuffled.
- Data was scaled so that the values now lie between 0 and 1.
- Box muller transformation was applied afterwards to make the data follow a normal distribution.
- Data was splitted into 70% training and 30% test set.
- I introduced higher powers of the data to improve on the accuracy of prediction.
  Here is the visualization(without box muller).

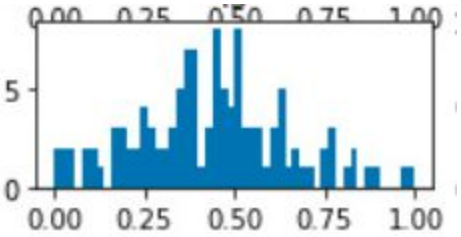


F1 F2

F1 * F1

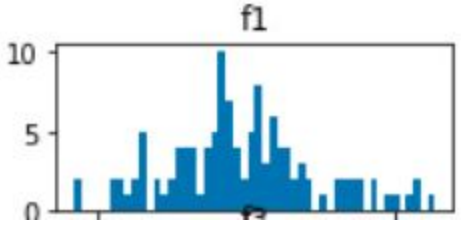


F2 * F2

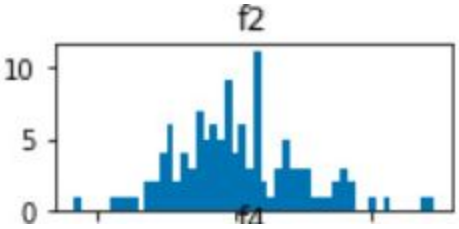


F1 * F2

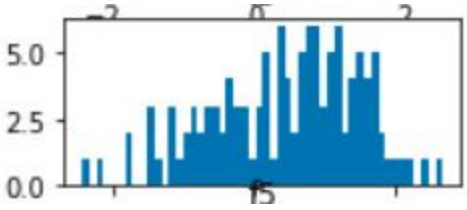
Here is the visualization(with box muller)



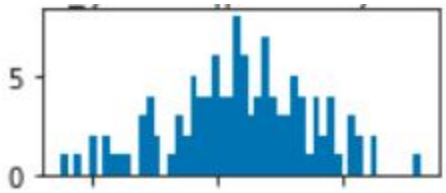
F1



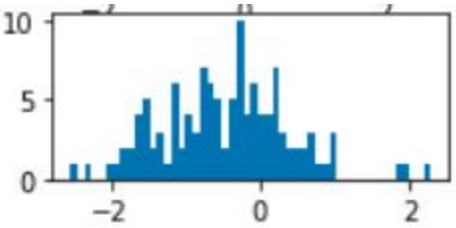
F2



F1 * F1



F2 * F2



F1 * F2

Analysis:

Both of the algorithms were run for 10 times as the data was shuffled before running GDA.

The results obtained were as follows.

1. **WIthout box muller the accuracies obtained were between 65.7% and 77% with an average of 72%.**
2. **With box muller the accuracies obtained were between 60% and 85% with an average of 80.3%.**
3. We observe the best and maximum accuracies in case of box muller to be better than without box muller by an approximate of 10%
4. The least accuracy was however obtained in case of box muller. This may be because in that specific permutation of the input data the data must have been too diverse and the normal approximation thus failed.
5. **We can safely conclude that after the application of box muller transformation on Gaussian Discriminant Analysis improves the accuracy by an approximate of 10%.**