Assignment 4 : IEC2017047

GDA

QUESTION DESCRIPTIONS:

Use the microchip dataset . Use 70% of the data for training and 30% for testing.

(1)   Use raw data as given, and from there develop a GDA model (without Box-Muller transformation). Thereafter, utilize the same to predict whether a component will be accepted or rejected.

(2)   For two features, first using Box- Muller transformation create new data set having Gaussian distribution within the range of the given data set and create Gaussian Discriminant Analysis (GDA) model.Thereafter utilize the model to predict where a component will be accepted or rejected using testing data

Compare the performance of GDA in both the above cases and write comparative analysis report.

[QUESTION 1]

RESULTS:

The dataset was divided into 70% training and 30% testing.
Accuracy of the transformed testing data using raw Dataset:63.888888888888886% => The results can vary because of the random distribution of training and testing data.

SCREENSHOT OF RESULTS OBTAINED:

```
phi: 0.49382716049382713
mu0: [0.06924204 0.11264529]
mu1: [0.00431932 0.1416292 ]
sigma: [[ 0.07162791 -0.03182581]
 [-0.03182581  0.09754918]]
Correct Predictions: 23 ,Out of: 36
Accuracy:   63.888888888888886
```
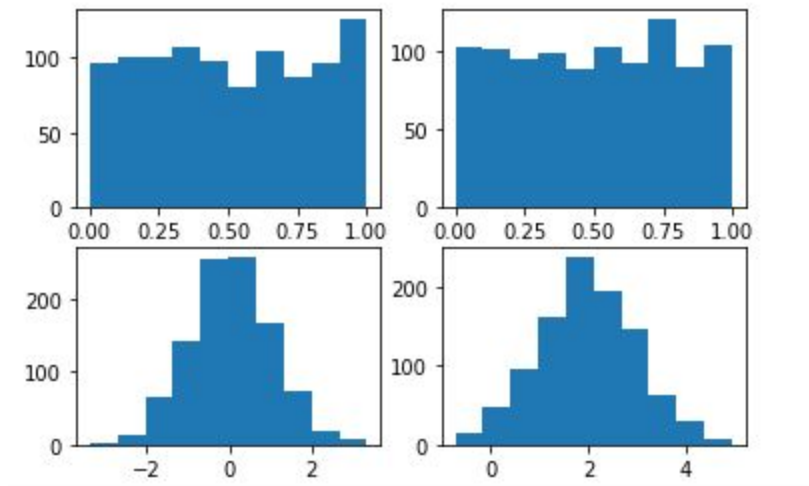
[QUESTION 2]

RESULTS:

The dataset was divided into 70% training and 30% testing.

Accuracy of the transformed testing data using raw Dataset:97.22222222222223% => The results can vary because of the random distribution of training and testing data.

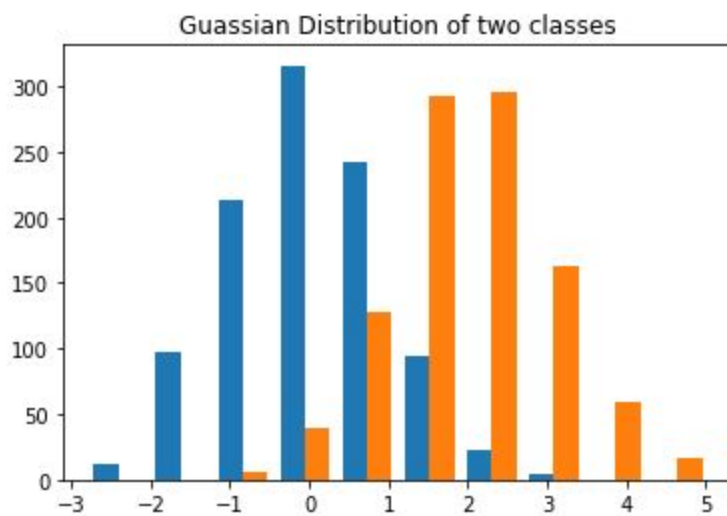## SCREENSHOT OF RESULTS OBTAINED
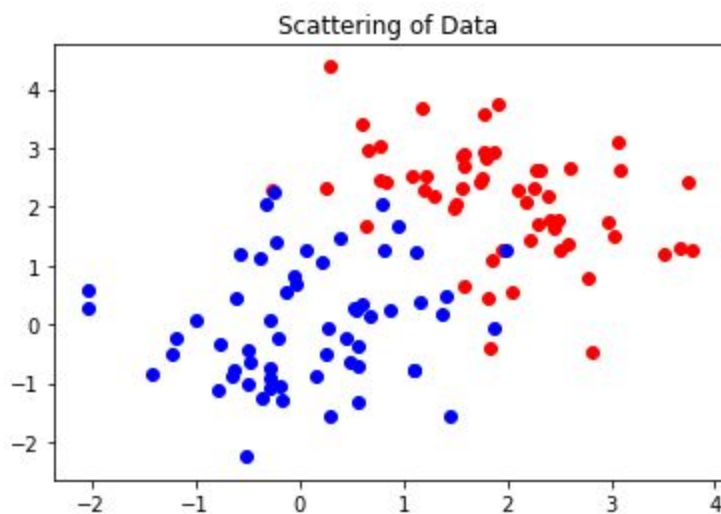
Histogram of Guassian Data

Calculated Values

```
phi: [0.48148148]
mu0: [-0.07270058  0.01431668]
mu1: [1.73793696 1.78814667]
sigma: [[0.42304878 0.16226338]
 [0.16226338 0.66253757]]
Correct Predictions: 35 ,Out of: 36
Accuracy:   97.22222222222223
```

Guassian Distribution of Two Classes

Guassian Distribution of Two Classes

CONCLUSION

It can be clearly seen when a Gaussian Dataset(created using BoxMuller Transformation) is used, predictions are very accurate but when a non-Gaussian Dataset(given raw dataset) is used, predictions are comparatively less accurate.