

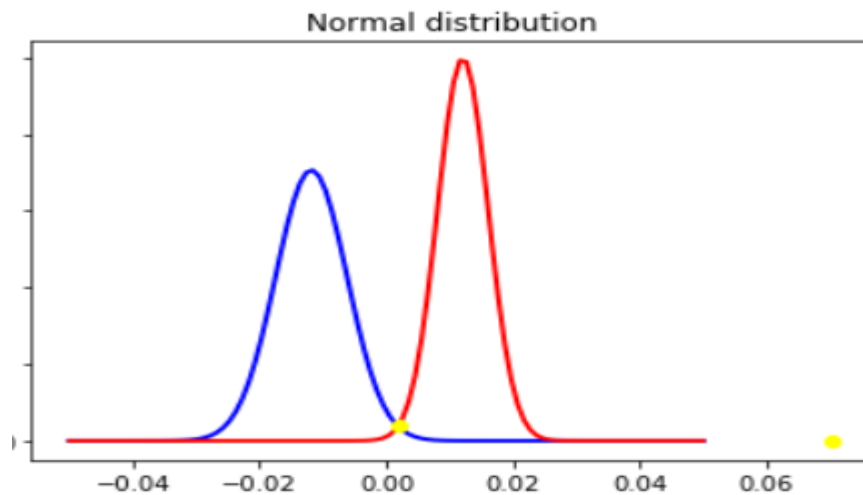
BITS F464 – Machine Learning (2019-2020 Sem II)

Report

Fisher's Linear Discriminant:

2D Plot:

Plot a normal distribution for each class and use the intersection point of both curves as the classification threshold. Report your implementation's accuracy and F-score.



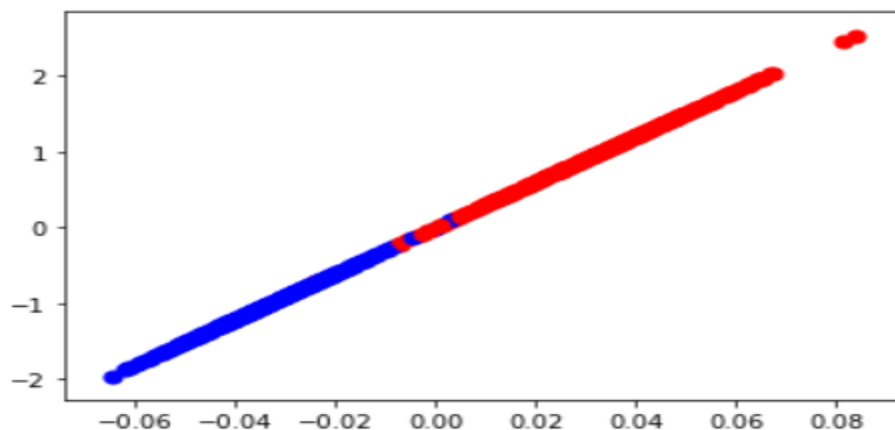
The intersection of 2 gaussian curves gives us classification threshold.

Blue=Positive (Class 1); Red= Negative (Class 0)

Accuracy: 99.3%

F-Score: 0.9929929929929929

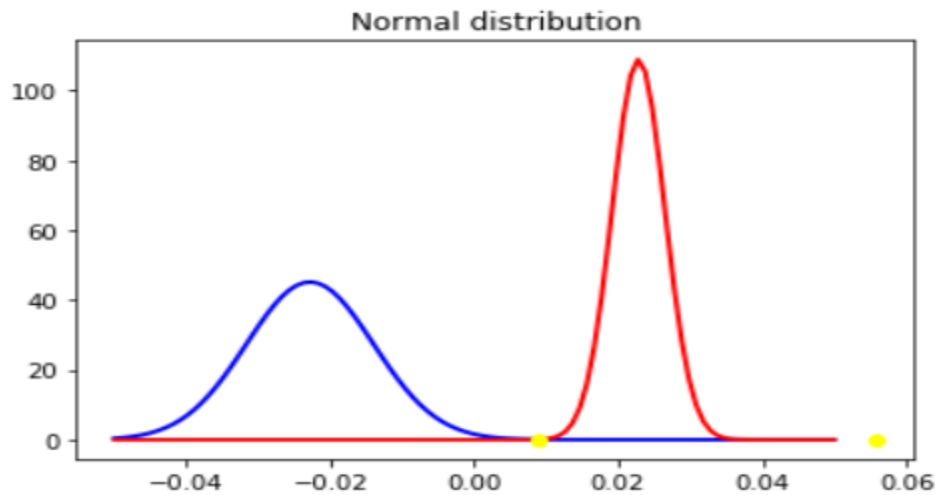
Once the points are transformed to one dimension, visualize them as points on a line and colour points of both classes separately.



Blue= Negative Red =Positive

3D Plot:

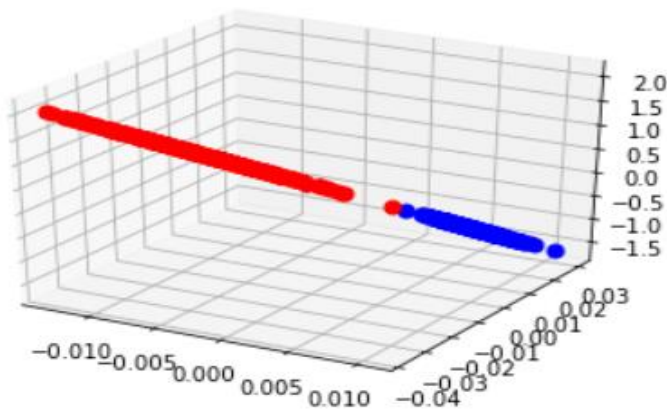
Once the points are transformed to one dimension, plot a normal distribution for each class and use the intersection point of both curves as the classification threshold. Report your implementation's accuracy and F-score.



Accuracy :100.0 %

F-score=1.0

Once the points are transformed to one dimension, visualize them as points on a line and colour points of both classes separately.



Naïve Bayes:

Data Preprocessing:

- Special characters were removed using regular expression.
- All the words were converted to lower case so as to not differentiate between words like The,THE,the.
- Stopwords were removed.

- Laplacian smoothening was used to avoid zero probability (Every word had at least one minimum occurrence in both positive and negative words)

Without preprocessing the accuracy was around 69%.

After preprocessing Mean Accuracy:80.70%

Final Result:

Accuracies: [77.0, 75.0, 85.0, 77.0, 86.0, 81.0, 84.0, 85.0, 76.0, 81.0]

F-Scores: [0.8, 0.7311827956989247, 0.8598130841121494, 0.7676767676767676, 0.8541666666666666, 0.8376068376068375, 0.8333333333333333, 0.8623853211009175, 0.7777777777777778, 0.8224299065420562]

Mean Accuracy: 80.700%

Accuracy stddev: 4.191

Mean F-Score: 0.815

F-score stddev: 0.044

Accuracy=0.807 ± 0.0491

F-score=0.815 ± 0.044