ML Assignment-1

Group Details

Tangeda Sai Sharan	**2017A7PS0241H**
Sanjiv Yelthimar Shenoy	**2017A7PS0224H**
Manish Kumar Bachhu	**2017A7PS0036H**

Fisher's Linear Discriminant Analysis – Question-1

We performed Fisher Discriminant Analysis on the 2 datasets provided. The results are as shown below.

To run the code:

- In order to not repeat code, the program first asks us to enter the relative path of the data file. Based on what path is entered, the algorithm is run on the particular dataset.

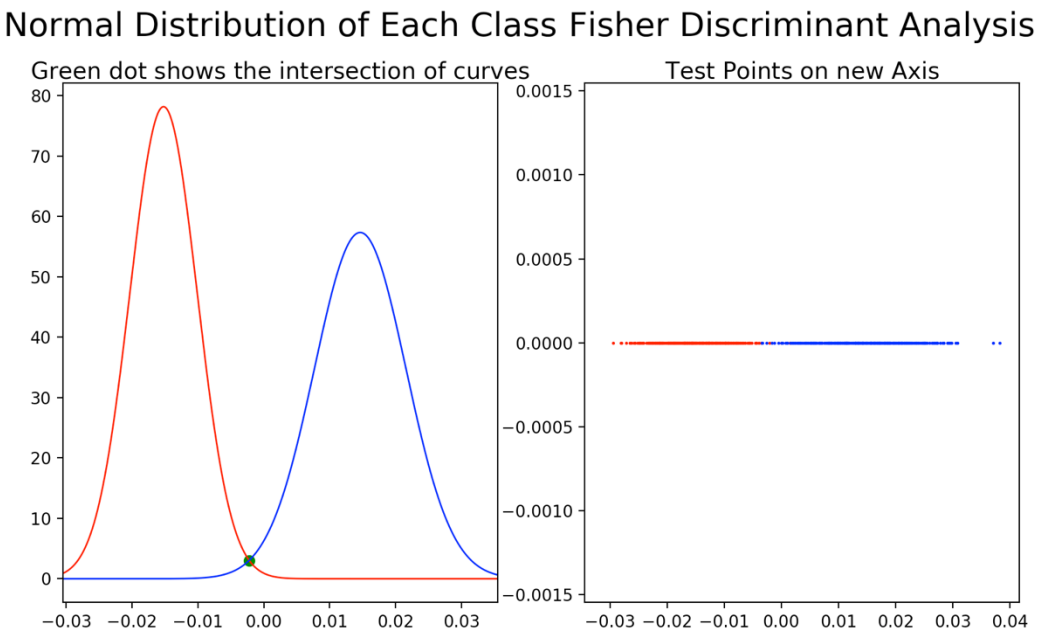
Relative paths for the datasets based on the directory structure we submitted will be:
For file 1: ./a1_data/a1_d1.csv
For file 2: ./a1_data/a1_d2.csv

Design Decisions/ Optimization Techniques used

- Numpy library was used everywhere for performing arithmetic operations using matrix multiplications, due to this optimization the code complexity is decreased, hence leading to faster output.
- No loop was used in any part of the application except in calculation of the Confusion matrix

1.1. Data Set 1

Precision: 0.9939
Recall: 0.992
Accuracy: 0.993
F - Score: 0.99299299



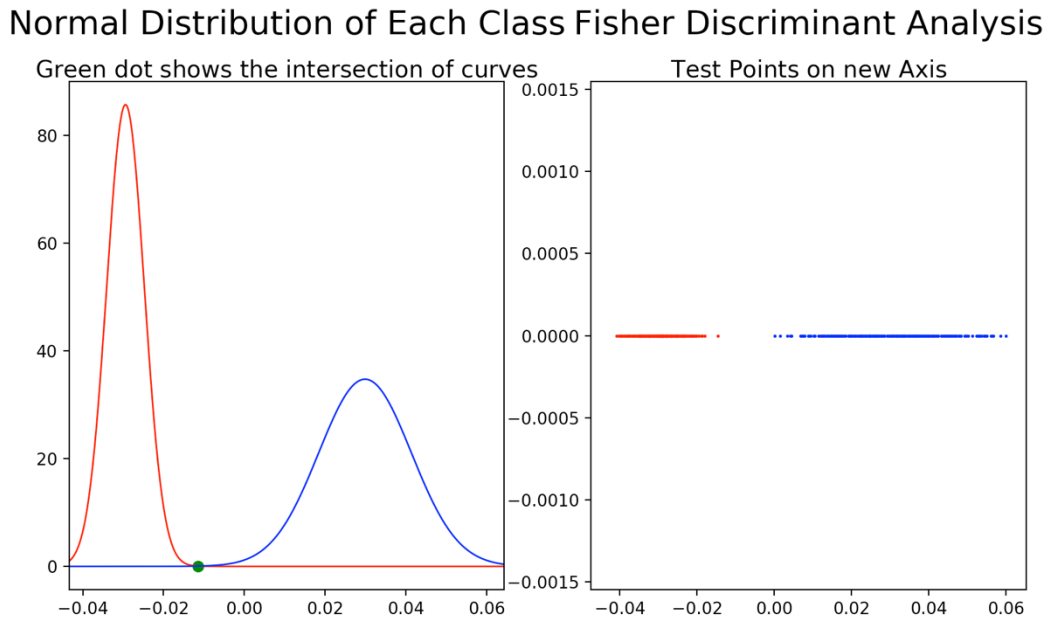
2 graphs are shown here.

The left graph shows the intersection of the normal distributions of the positive and the negative points on the training data set. The threshold point for separation is taken as the intersection of the normal distributions.

The right graph shows the visualisation of the testing points projected on a line got by the fisher discriminant analysis. As you can see, the positive and negative points are nearly separated out into 2 regions with just a little overlap.

1.2. Data Set 2

Precision: 1.0
Recall: 1.0
Accuracy: 1.0
F - Score: 1.0



2 graphs are shown here.

The left graph shows the intersection of the normal distributions of the positive and the negative points on the training data set. The threshold point for separation is taken as the intersection of the normal distributions.

The right graph shows the visualisation of the testing points projected on a line got by the fisher discriminant analysis. As you can see, the positive and negative points are clearly separated out in this dataset thereby giving us 100% accuracy.

Conclusion:

For both the data sets, our model performed quite well almost entirely separating out the points accurately.