BITS F464 - Machine Learning

Assignment-1A: Fisher's Linear Discriminant

Assignment done by:

Jhaveri Aditya Alok - 2018A7PS0209H
Vaishnavee Nautiyal - 2018A7PS0286H
Aryesh Koya - 2018A4PS0637H

Idea of Fisher's Linear Discriminant?

The idea proposed by Fisher is to maximize a function that will give a large separation between the projected class means while also giving a small variance within each class, thereby minimizing the class overlap.

In other words, FLD selects a projection that maximizes the class separation. To do that it maximizes the ratio between-class variance to the within class variance.

In short, to project the data to a smaller dimension and to avoid class overlapping, FLD maintains 2 properties:

- A large variance among the dataset classes
- A small variance within each of the dataset classes

Here we consider only two classes as C0 and C1 with their class labels 0 and 1 respectively.

To find the projection with the following properties, FLD learns a weight vector ***W*** with the following criterion.

$$\boldsymbol{J(W) = (m_2 - m_1)^2 / (s_1^2 + s_2^2)}$$

Where green color indicates Between-class variance and red indicates within class variance

We already know that:

$s_k^2 = \sum (y_n - m_k)^2$

$y_n = \boldsymbol{W^T} x_n$

Substituting these values we get J(W) as:

$$J(W) \ = \ (W^T(SB)W)/(W^T(SW)W)$$

$$W \propto (SW)^{-1}(m2 \ - \ m1)$$

Where **m2-m1** indicates per class mean, **yellow** highlighted text indicates within class variance and **green** highlighted text indicates projection equation.

Therefore to maximize J(W) we consider the following constrained optimization problem:

$\min_w = -½(w^T S_B w)$
s.t. $w^T S_W w = 1$

Implementation of FLD:

There will be only one decision boundary as the number of unique classes are 2 only, C_0 and C_1 having labels as 0 and 1.

Our team has used google colab to do this assignment along with mounting google drive to import the dataset.

First of all we input our whole dataset and we used this whole dataset for training as there is no requirement of splitting the dataset.

Then we compute the within class variance matrices (Scattered matrices) for both of the classes and add them to get the final S_W matrix which represents the within class variance. Its inverse is taken and stored in the SW_INV matrix. Then Weight vector (W) is found out by the dot product of SW_INV and mean1-mean2 matrix.

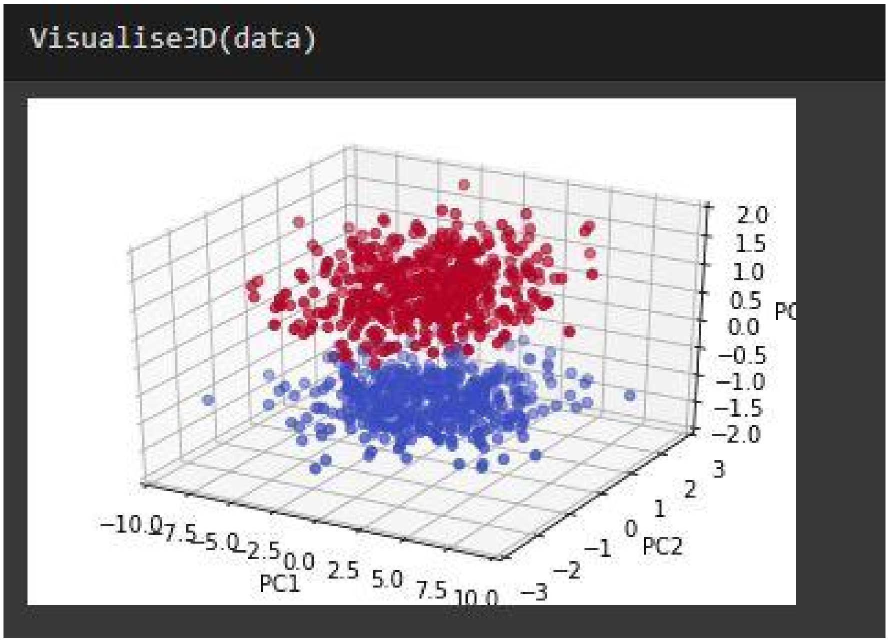
Separating the data with labels 0 and 1 into the class2 and class1 data structure are then used to find the projection of each class on the 2-Dimension plot.
The value of w_0 is also to be found out with the help of means and standard variations of the projection classes and the values are to be plotted using a normal distribution.

The accuracy of the model is to be found out by knowing how many true and falsely classified values there are.

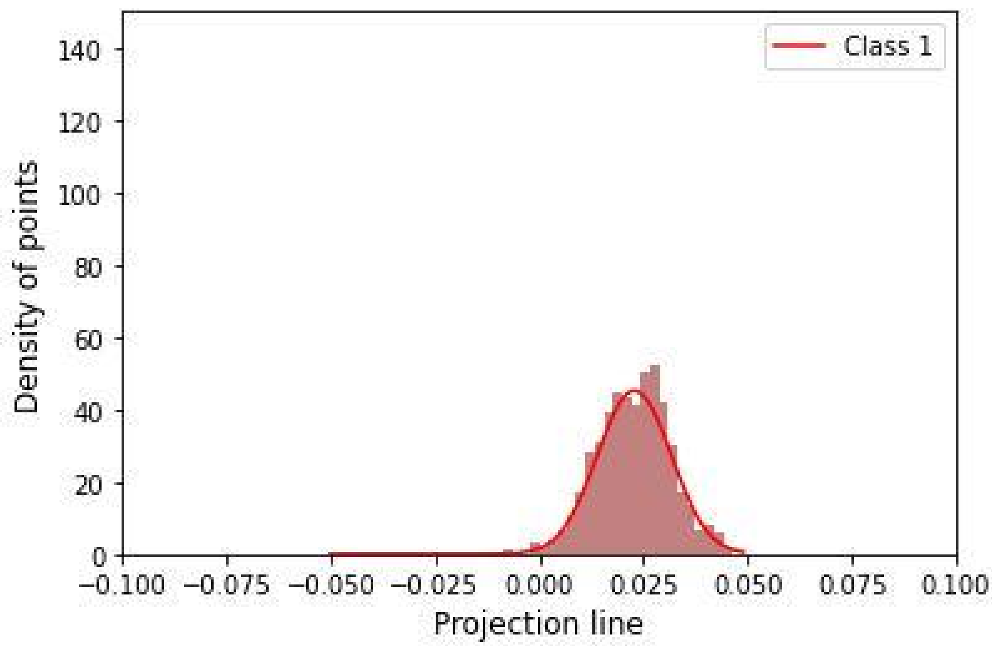
The plane is also to be found out which separates the classes

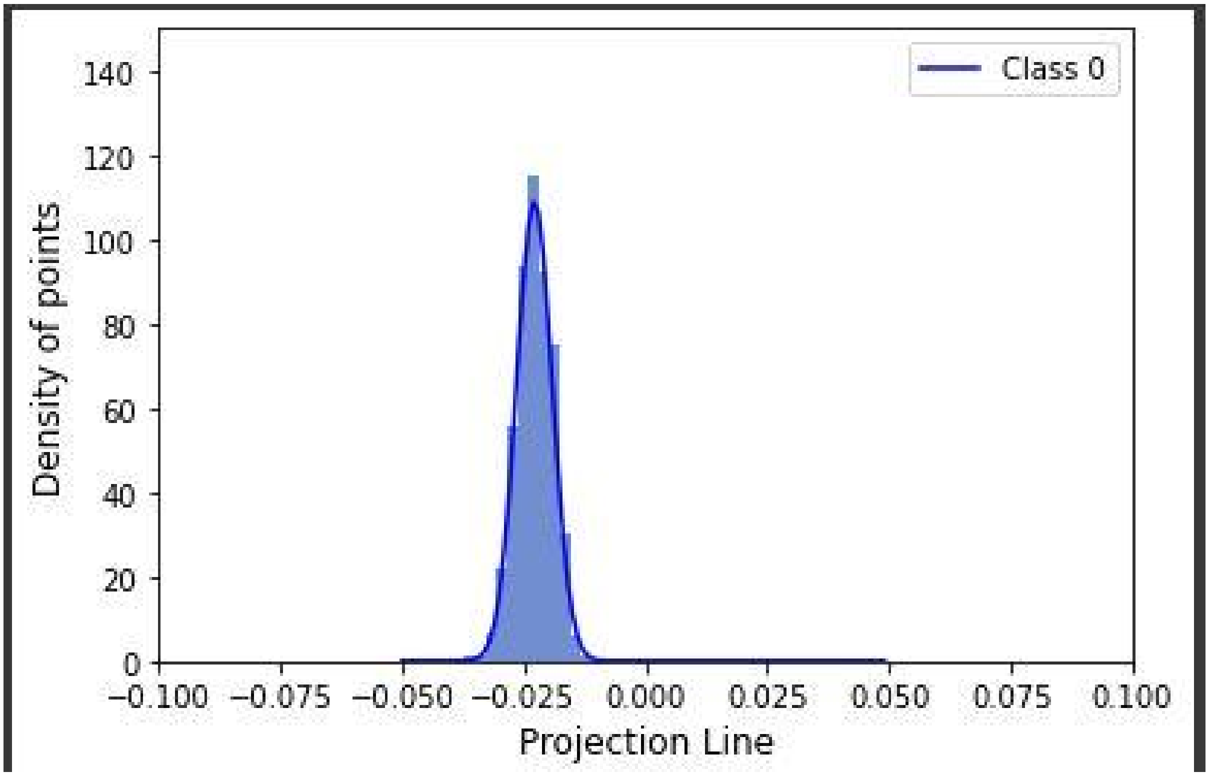
The plots are here as follow:

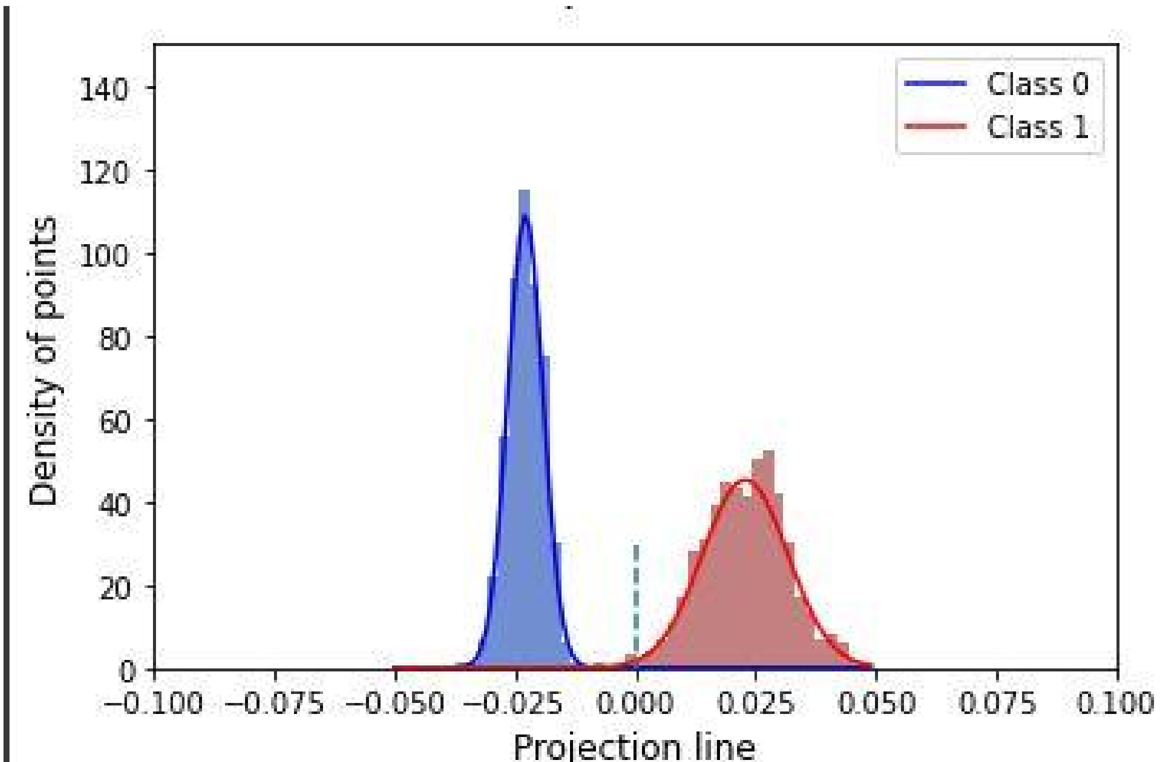
Scattered plot of all points colored according to their classes:

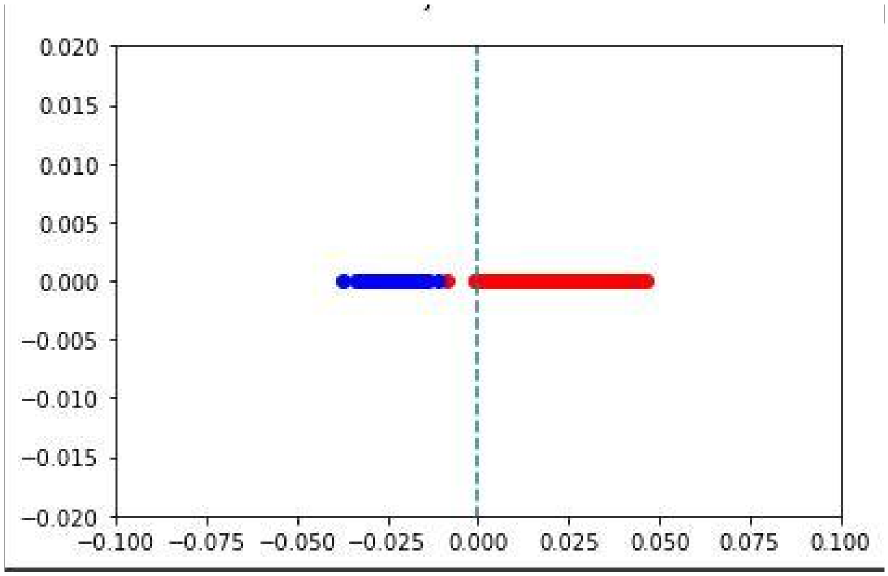


Normal distribution plots obtained for class 0 and class 1 and their intersection as follows:



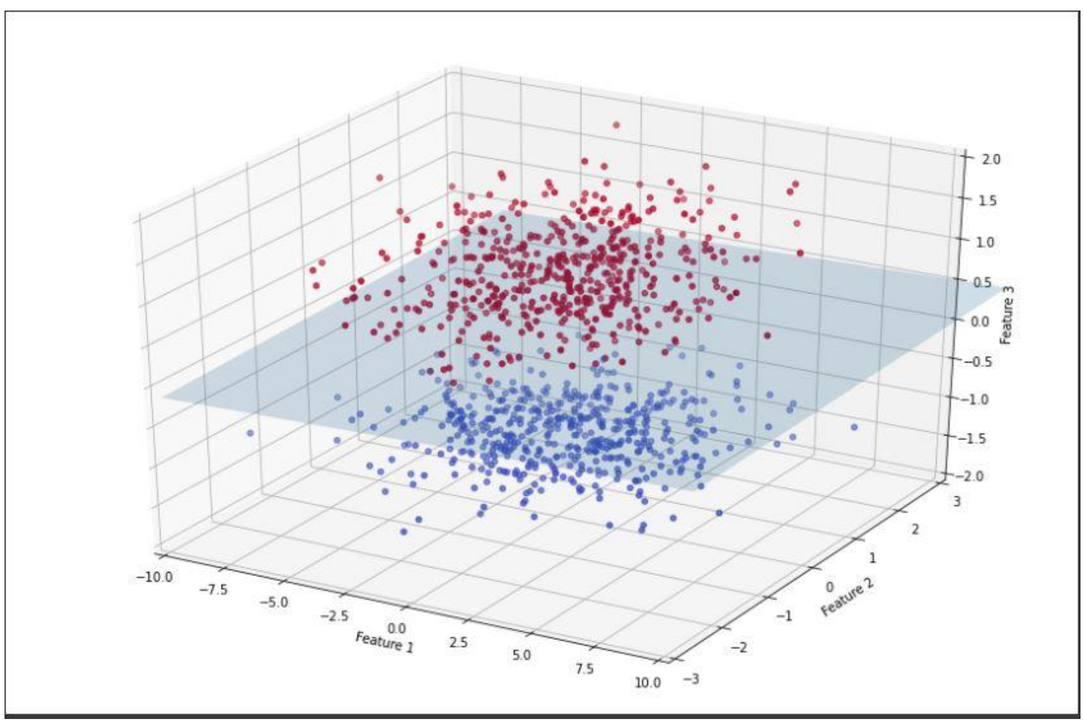

Class 0
Density of points
Projection Line
140
120
100
80
60
40
20
0
−0.100
−0.075
−0.050
−0.025
0.000
0.025
0.050
0.075
0.100


Class 0
Class 1
Density of points
Projection line
140
120
100
80
60
40
20
0
−0.100
−0.075
−0.050
−0.025
0.000
0.025
0.050
0.075
0.100



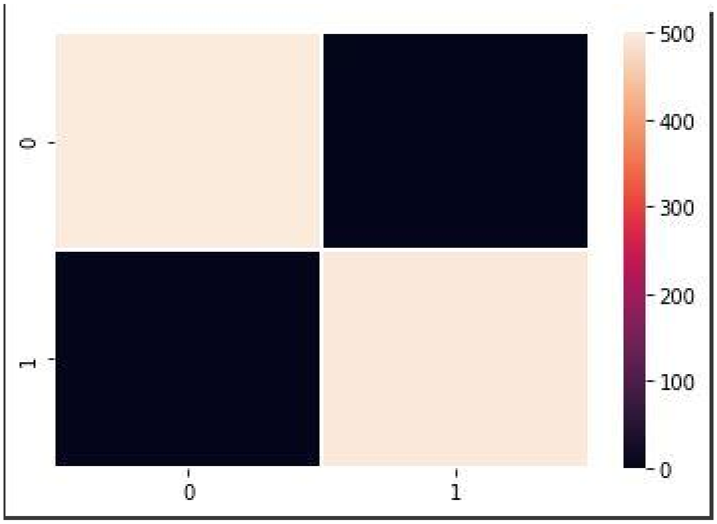
The upper image is 1D representation of the clusters with discriminant line

Here from all the images above we can see that the Deascriinistic value, where the two classes intersect for our given dataset is very near to 0.000



The Classes with their respective separating plane in 3D

The accuracy received with this model is 99.7 percent which is close to 100 percent as the data given is linearly separable



We can see that (0,0) and (1,1) are white while the other blocks (misclassified blocks) as black (indicating zero belong to them) say that the data is correctly classified.

Limitations:

FLD can only solve problems which are linearly separable which says that its utility has limits and they can be overcome by Fisher's quadratic discriminant which is an extension of FLD which separates problems based on quadratic boundaries.