**ASSIGNMENT-1**

**Fisher Linear Discriminant Analysis**

GROUP MEMBERS:

VATSAL GUPTA: 2018A7PS0198H

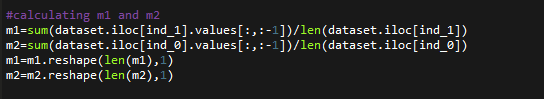
MIHIR BANSAL: 2018A7PS0215H

BANDARU ABHINAV: 2018A7PS0236H

**DESIGN:**

The design of the algorithm is based on the formulas given by Fisher’s Discriminant method where the separating vector is determined using Sw and means of the points of each class.

* m1 and m2 (the means of class 0 and class 1 respectively) are calculated by taking simple average of each factor.



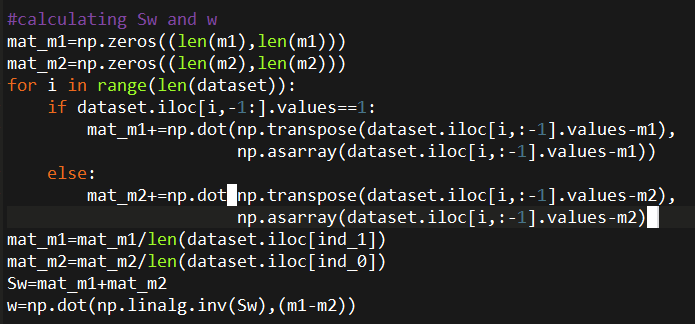
* Sw(co-variance matrix) is calculated using the formula:

SW=∑n∈C1(xn−m1)(xn−m1)^T +∑n∈C2(xn−m2)(xn−m2)^T

where c1 and c2 are the class sizes of class 0 and class 1 respectively.

* w is then calculated by taking the dot product of

Sw^-1 and (m1-m2)



* Then, the program coverts the data into 1D by multiplying the dataset values with w (this data is held by trans\_vec list) for both the classes
* This vector is then used to plot both the normal distribution for both the classes. This normal distribution can be plotted by calculating the mean and standard deviation for both the classes separately.
* Now to find the intersection point of the two normal curves we can use the below formula which gives the two intersection points. The point which lies in between the means of the curves is chosen as the intersection point and thus the threshold.

**A= −1/σ21 + 1/σ22**

**B= 2(−μ2/σ22 + μ1/σ21)**

**C= μ22/σ22 − μ21/σ21 + log(σ22/σ21).**

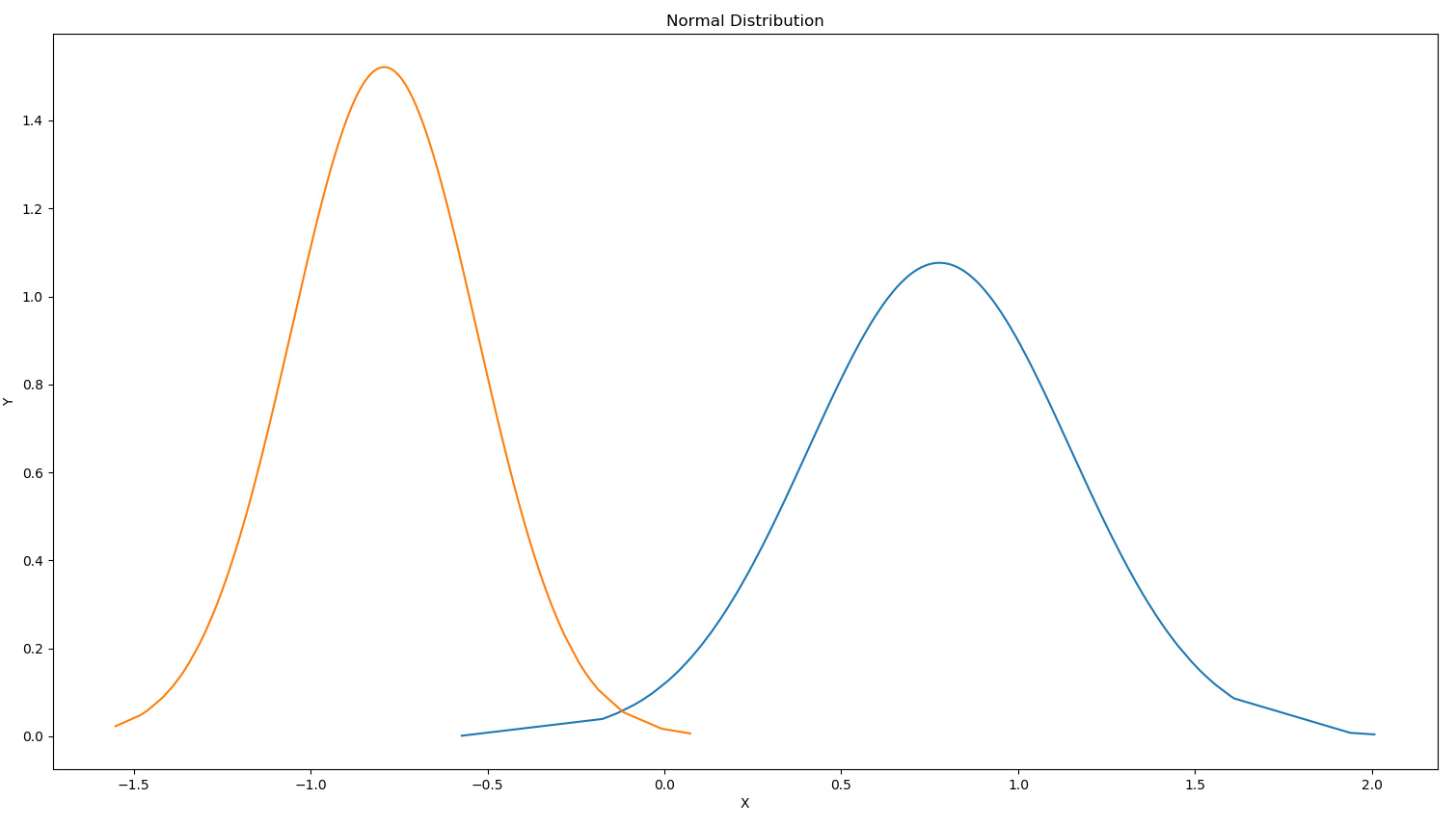
**Where A, B, C are the coefficients of the equation**

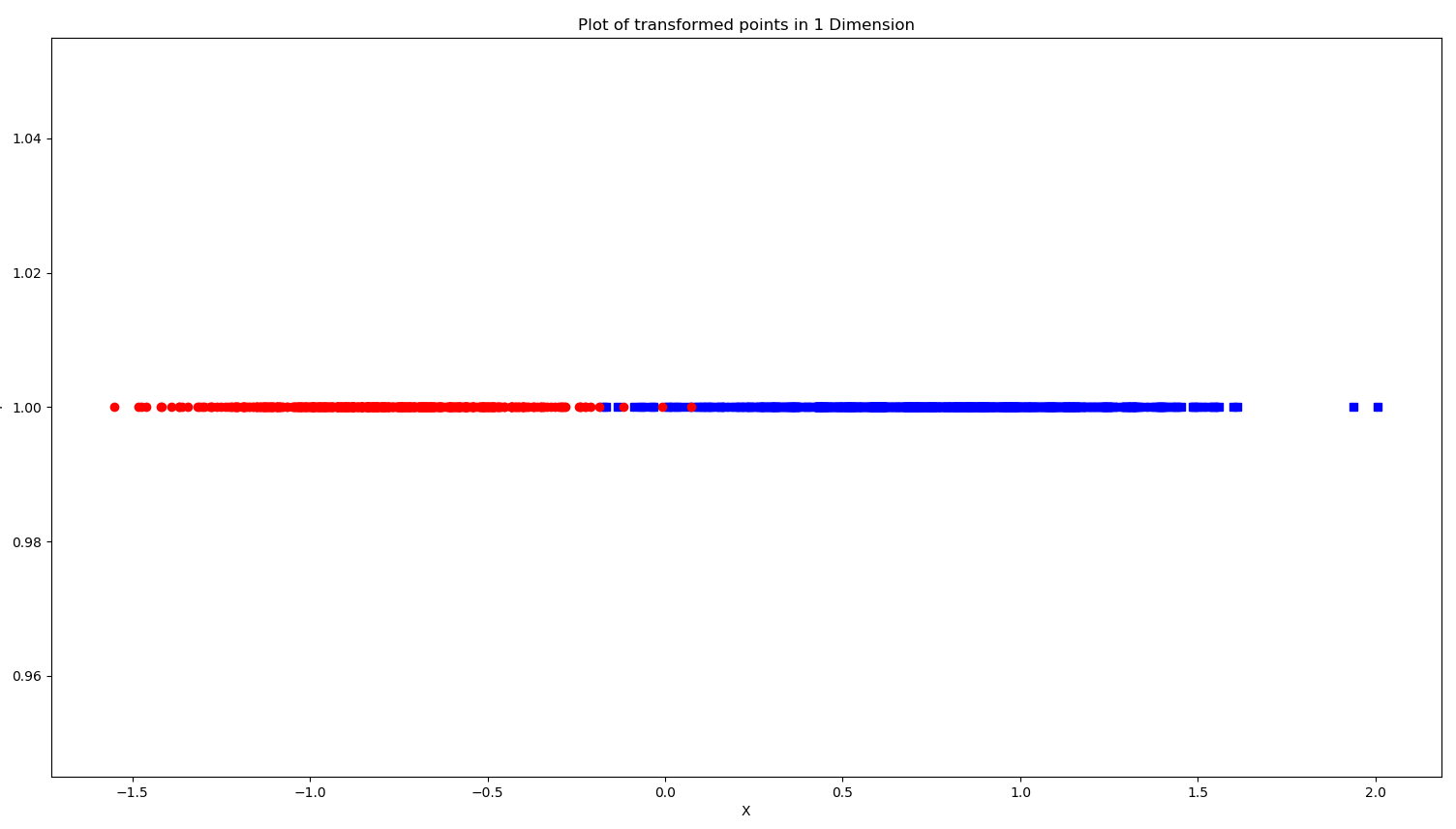
**Ax2+Bx+C=0. Whose solution can be easily found out by Shri Dharacharya’s method**

* This threshold point will then classify our dataset into two classes storing the output in the array pred, whose accuracy and F-score can easily be found out.

----------------------------------------------------------------------------------------

**RESULTS:**

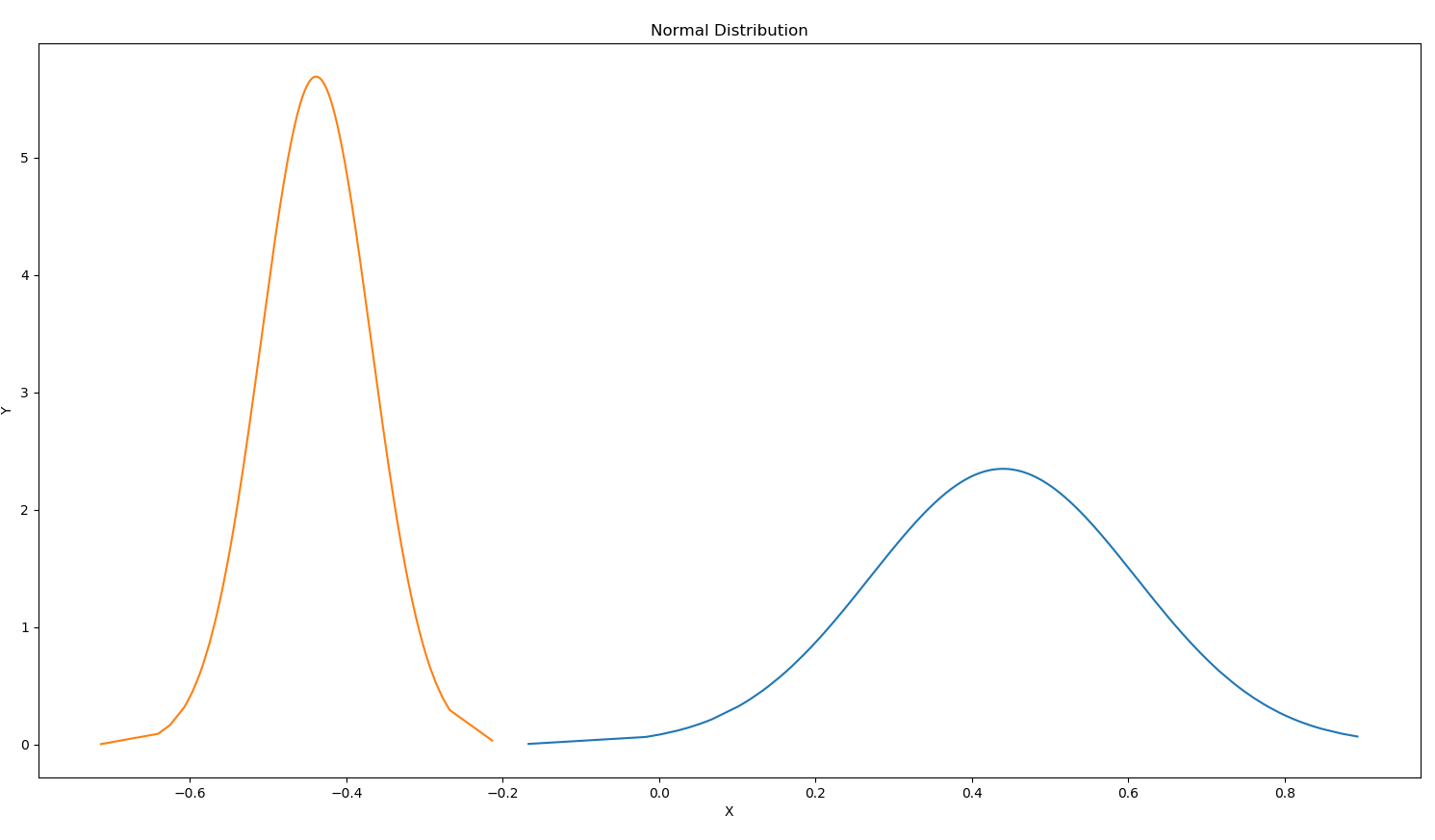
**TRAINING SET-1**

Red – class 0, Blue – class 1

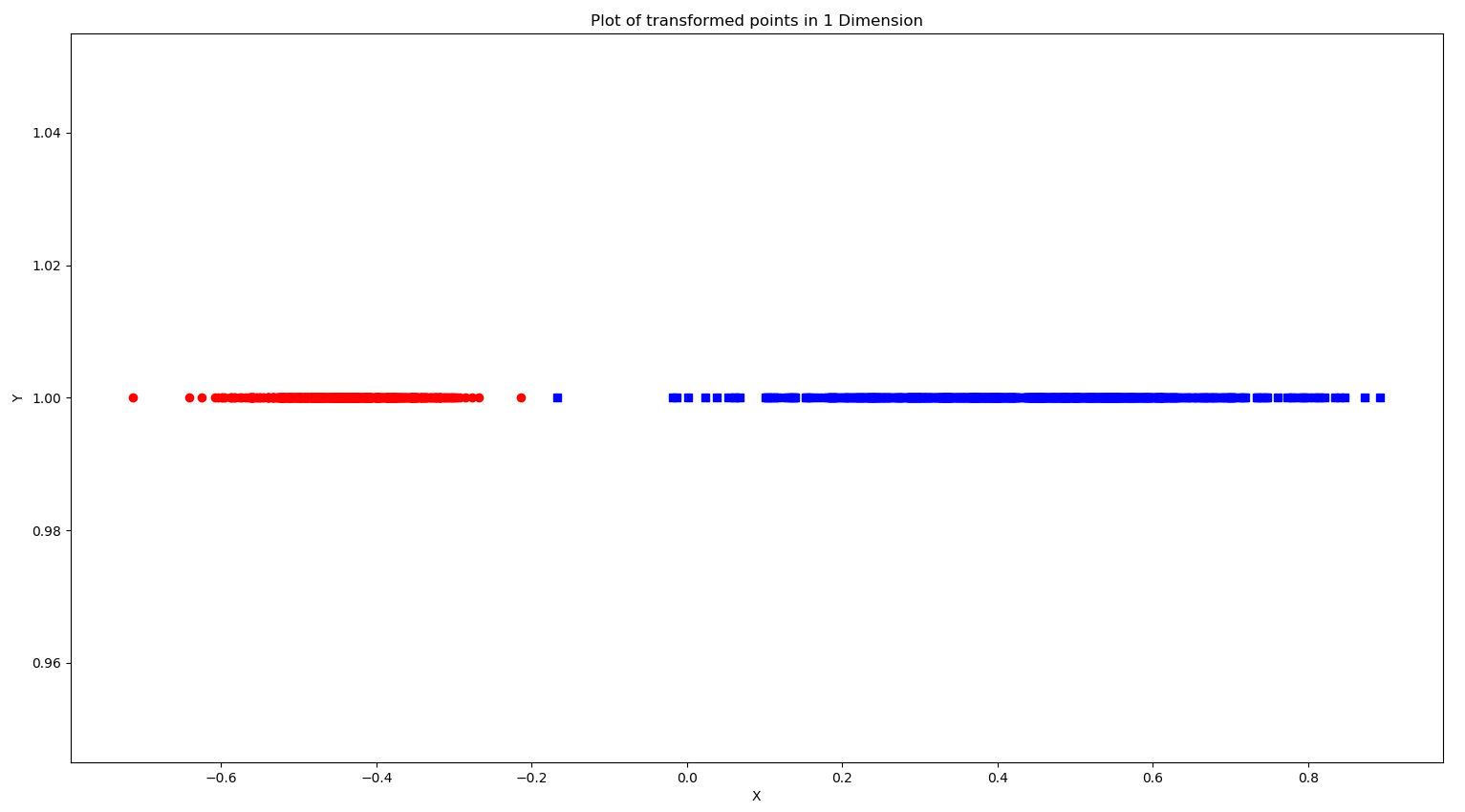
Testing Accuracy for dataset 1: 99.3 %

F-Score for dataset 1: 99.2992992992993 %

**TRAINING SET-2**



Red – class 0, Blue – class 1



Testing Accuracy for dataset 2: 100.0 %

F-Score for dataset 2: 100.0 %

**CONCLUSIONS:**

1.Normal Curves:

We observe an intersection in the first training set, but not in the second set. As the second set is completely linearly separable, we observe 100% accuracy and F-Score.

2. Using the above observations, we can conclude that the Fisher Discriminant Method can be used to accurately predict the data for approximately linearly separable data only.