

---

# ML Assignment 1: Fischer's LDA

---

Anirudh Agrawal  
2018A7PS0099H

Aviral Agarwal  
2018A7PS0192H

Vikramjeet Das  
2018A7PS0280H

## 1 Introduction

In this assignment, we implement Fischer's Linear Discriminant, it is used for supervised learning in solving classification problems. Given the data points in  $M$  dimensional space, we project all those points to  $D$  dimensions and then try to find out a discriminant function to classify our points using a threshold found by calculating the intersection point between the normal distribution followed by the projected points. In this assignment we were given data points in  $3-D$  and we projected it to  $1-D$  and found out the threshold.

## 2 Implementation

After solving the optimisation problem:

$$\max \frac{(W^T M_1 - W^T M_2)^2}{s_1^2 + s_2^2}$$

We get

$$W \propto S_W^{-1}(M_1 - M_2), \text{ where}$$

$$S_W = \sum_{k=1}^2 S_k, \text{ and}$$

$$S_k = \sum_{n \in c_k} (x_n - M_k)(x_n - M_k)^T$$

The threshold can be found by solving the quadratic equation  $Ax^2 + Bx + C = 0$ , where

$$A = -\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

$$B = 2\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right)$$

$$C = \frac{\mu_2^2}{\sigma_2^2} - \frac{\mu_1^2}{\sigma_1^2} + \log\left(\frac{\sigma_2^2}{\sigma_1^2}\right)$$

The algorithm can be roughly outlined as:

1. Find the individual means for different classes

2. Find  $S_W$

3. Calculate unit vector  $W$

4. Project all the points on to the unit vector by taking the dot product  $W^T X$

5. Find the Normal distribution fitting the projected points

6. Find the **threshold** using the intersection point of the normal distribution.

## 3 Results

Figure 1 shows the original data plot in 3 dimensions along with the decision boundary (in green). All points above the green plane are classified as belonging to *class 1*, while those below the plane are classified as belonging to *class 0*. Figure 2 shows the data points projected to 1-dimension along with normal distributions fit on them and the threshold point.

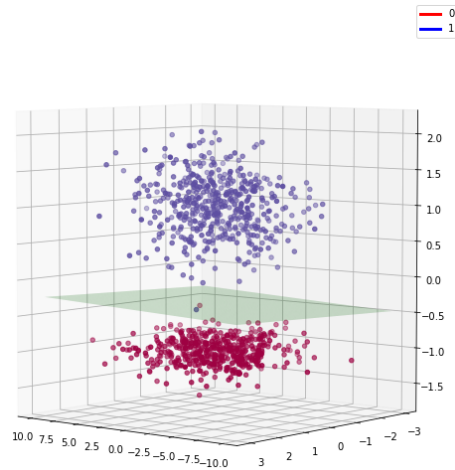


Figure 1: Scatter plot of the original data in 3-dimensional space along with the decision boundary

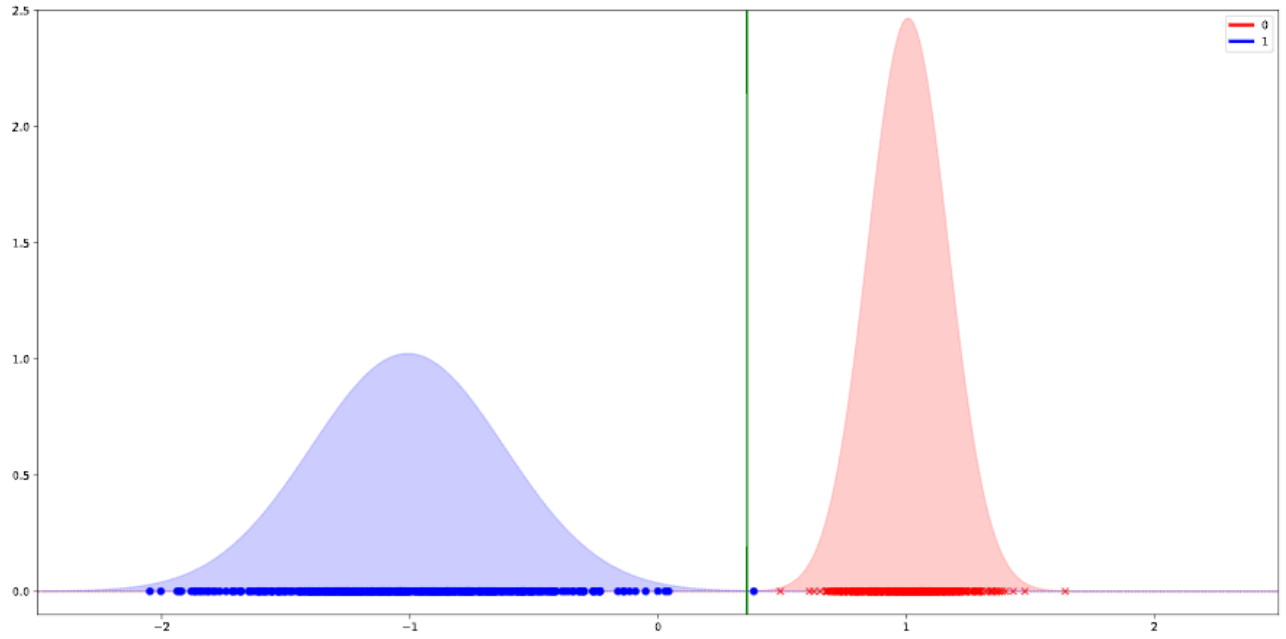


Figure 2: Normal distribution fit to each projected cluster. The green line shows the threshold point

We got an accuracy of : 99.9%

The threshold point in **1-D** is 0.359.

Equation of discriminating plane is :  $x = 0.359$

The discriminating plane in original dimensions(**3-D**) is:  $0.00656x_1 + 0.0812x_2 - 0.9998x_3 = 0.359$

Normal unit vector to plane is :  $[ 0.00656, 0.0182, -0.9998 ]$