BITS F464

MACHINE LEARNING

ASSIGNMENT 1A

Fischer's Linear Discriminant

This LDA algorithm uses Fisher's criterion to determine the weight vector and therefore the discriminant. We then project our points onto this single dimension and find the intersection of normal distribution curves to find the optimum classification threshold thus, minimizing the within-class variance and maximizing the between-class difference.

Let (X_i,y_i) where i $\in\{1,2,3,\ldots..,N\}$ be the N data points of the dataset where X_i is the set of values for the features of the i^{th} point and $y_i \in \{0, 1\}$ denoting the class they belong to.

Let *W* be the linear discriminator that separates the data points of the two classes called weight vector.

We find a direction W such that it separates the data points by minimizing within class variance $s_0^2 + s_1^2$ and maximizing the between-class difference $(m_1 - m_0)^2$.

$$=> \boldsymbol{J(W) = \frac{(m_1 - m_0)^2}{s_0^2 + s_1^2}} \text{ should be maximized.}$$

Let the mean values of the projected data be m_0 and m_1 for classes 0 and 1 respectively and s_0 and s_1 be the variances of the two classes in the projected data.

We know that,

$$M_0 = \frac{1}{n_0} \sum_{X_i \in C_0} X_i \, ; \qquad M_1 = \frac{1}{n_1} \sum_{X_i \in C_1} X_i$$

$$m_0 = W^T M_0 \, ; \qquad m_1 = W^T M_1 \qquad \text{-- equation (1)}$$

$$s_0{}^2 = \sum_{X_i \in C_0} (W^T X_i - m_0) \, ; \quad s_1{}^2 = \sum_{X_i \in C_1} (W^T X_i - m_1) \qquad \text{-- equation (2)}$$

<u>Numerator:</u>

From equation (1)

$$(m_0 - m_1)^2 = W^T (M_1 - M_0)(M_1 - M_0)^T W \qquad \text{-- equation (3)}$$

$$\Rightarrow \ (m_0 - m_1)^2 = W^T S_B W \qquad \text{-- equation (4)}$$

where $S_B = (M_1 - M_0)(M_1 - M_0)^T$

<u>Denominator:</u>

From equation (2)

$$s_k{}^2 = W^T \left[\sum_{X_i \in C_k} (X_i - M_k)(X_i - M_k)^T \right] W \qquad \text{-- equation (5)}$$

$$\Rightarrow \ s_0{}^2 + s_1{}^2 = W^T S_w W \qquad \text{-- equation (6)}$$

where $S_W = \sum_{class\ k} \sum_{X_i \in C_k} (X_i - M_k)(X_i - M_k)^T$

From the above simplification, we get

$$\Rightarrow J(W) = \frac{W^T S_B W}{W^T S_w W}$$

To maximize J(W) by differentiating and equating it to zero, we get

$$S_w W = p S_B W \text{ where p = constant} \qquad \text{-- equation (7)}$$

If S_w is invertible, then equation (7) can be re-written as

$$W = S_w^{-1} S_B W \qquad \text{-- equation (8)}$$

From equation (4) $S_B W = q(M_1 - M_0)$ where $q = (m_1 - m_0)$

$$\Rightarrow W = S_w^{-1}(M_1 - M_0) \qquad \text{-- equation (9)}$$

FLDA was implemented by computing Matrix difference of M1 and M0 for the data points of class 1 and class 2 respectively and within-class variance S_w.

The weight vector is then computed using the equation (9) ->

$$W = S_w^{-1}(M_0 - M_1)$$

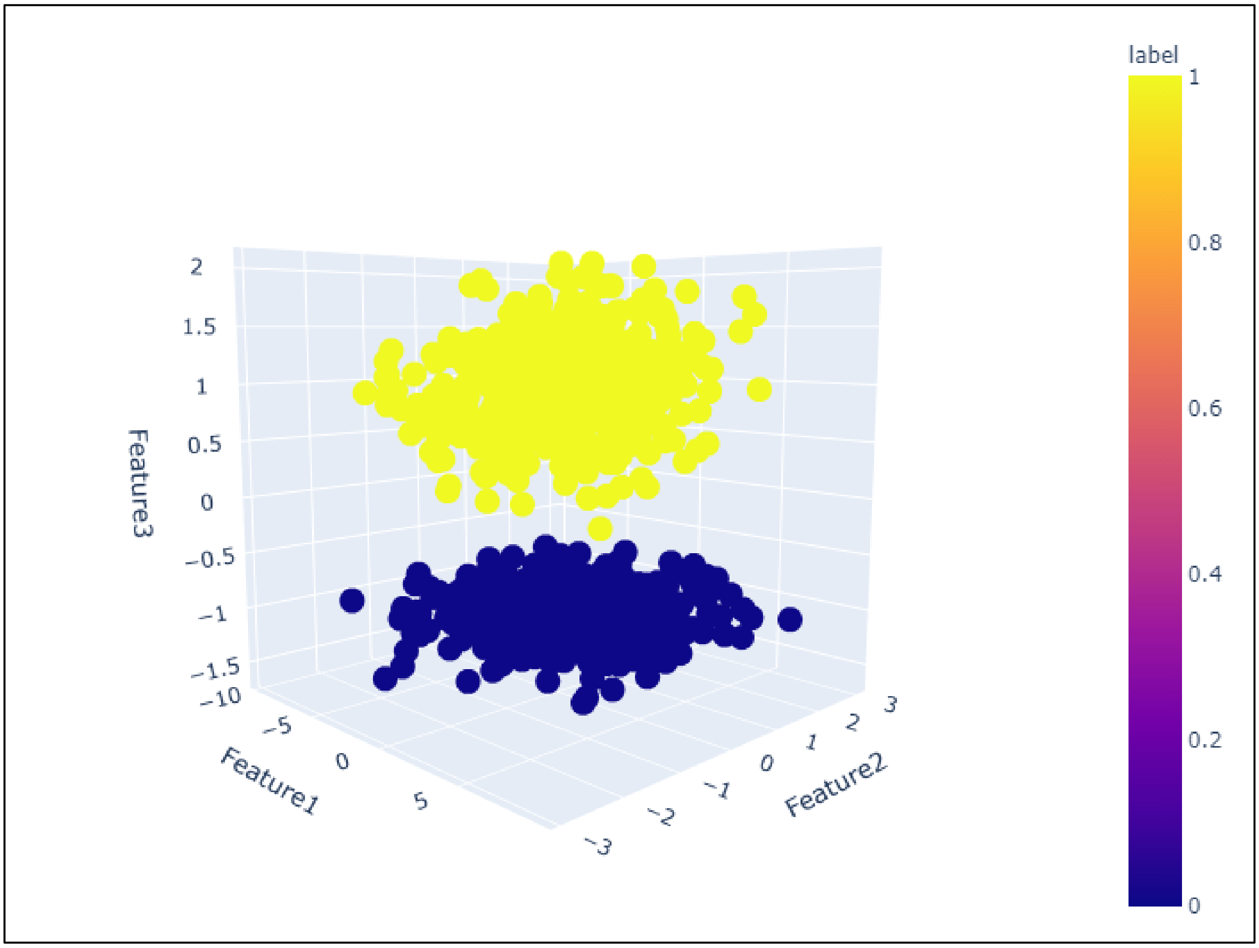
Analysis –

The given dataset is a binary classification problem. It has data points with 3 features with 0,1 as possible y values.

No. of features = 3

No. of instances = 1000

No. of classes = 2



Results obtained --

```
The value of M1 is:
[[ 0.10229483]
 [ 0.10337021]
 [-1.00513   ]]
The value of M2 is:
[[0.22310309]
 [0.00255859]
 [1.01027072]]
The weight vector is:
[[-0.07422404]
 [-0.20644839]
 [11.31793517]]
Mod of vector is:    11.320061250442755
The unit weight vector is:
[[-0.00655686]
 [-0.01823739]
 [ 0.99981218]]
The value of m1 is:
-1.0074971533352854
The value of m2 is:
1.0085714575861833
```

```
Std Dev Class 1:  0.16168139595107733
Std Dev Class 2:  0.3898154645217035
```

```
The threshold is:
-0.38930280209937684
The accuracy is:
100.0 %
The F-score is:
1.0
```

The equation of the discriminant plane is –

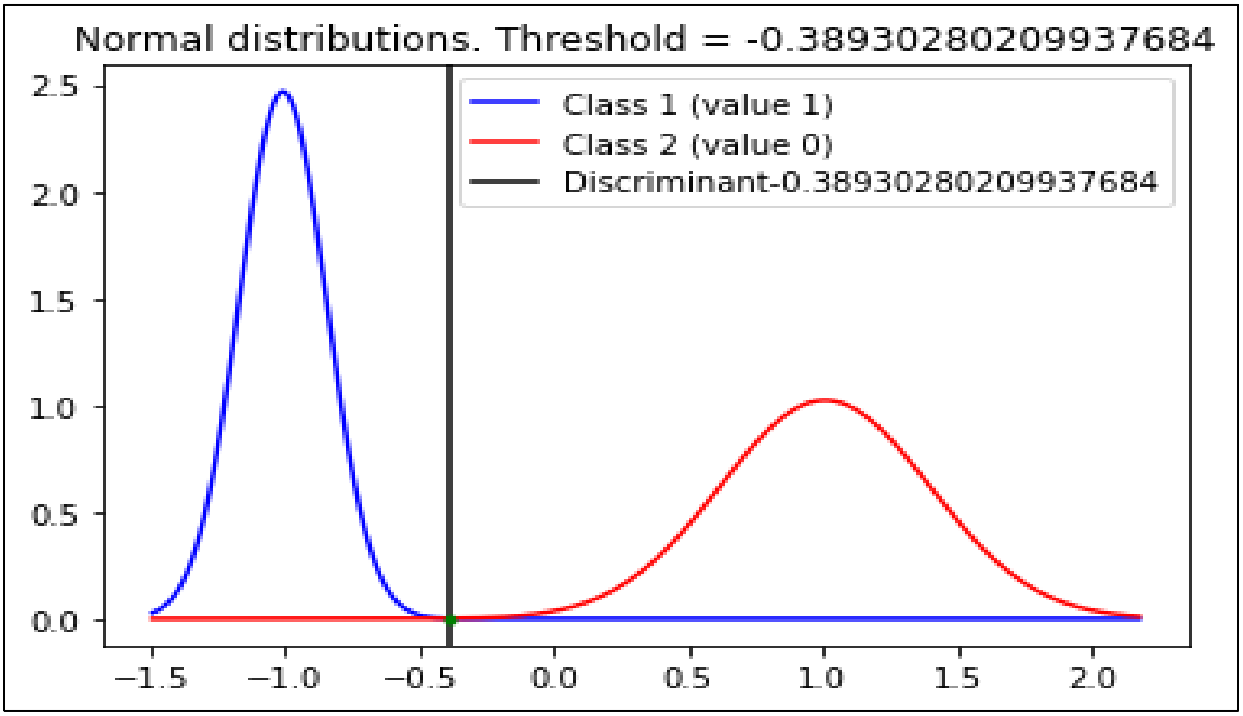
w[0]x + w[1]y + w[2]z = threshold.

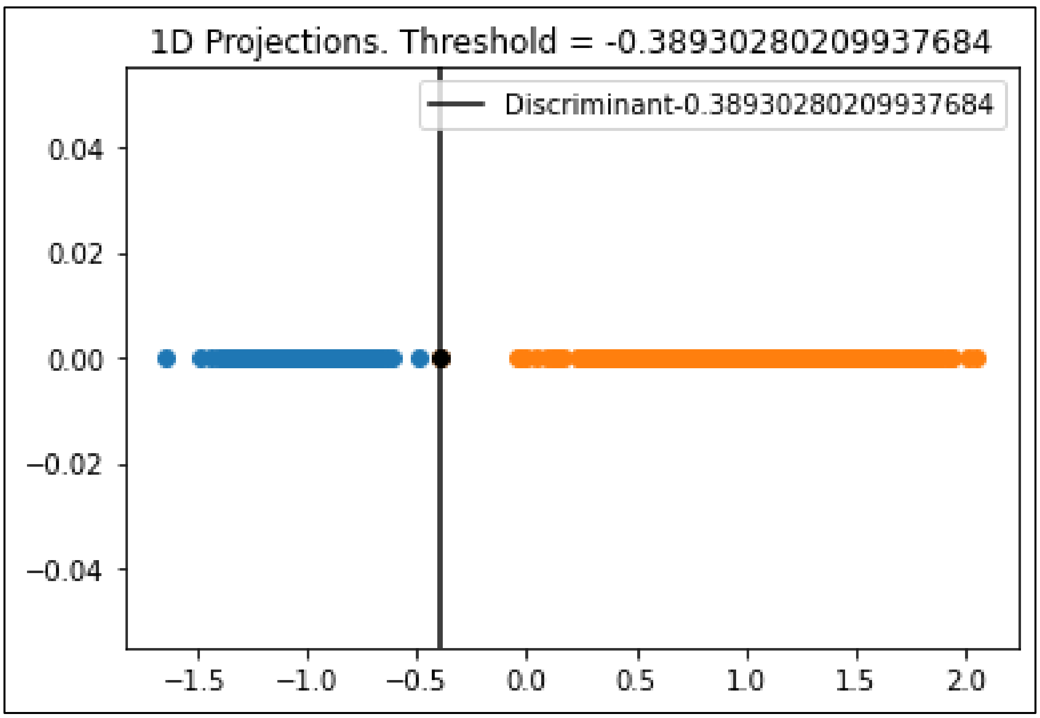
Where w[0] = -0.00655686

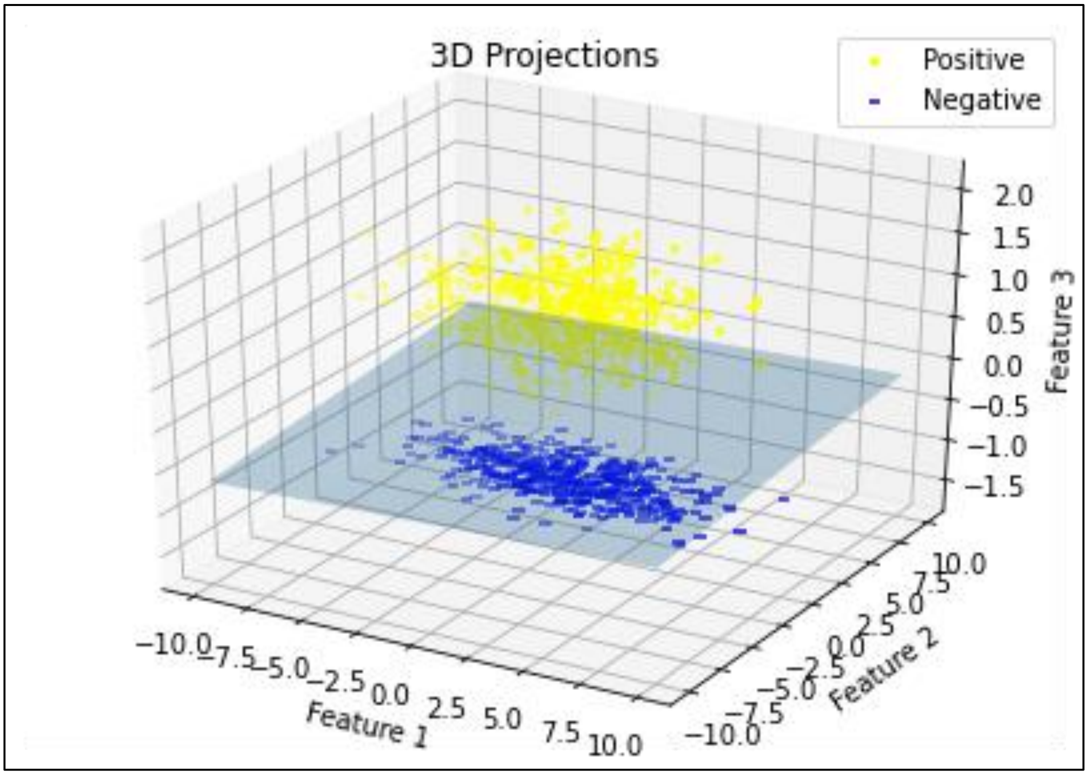
w[1] = -0.01823739

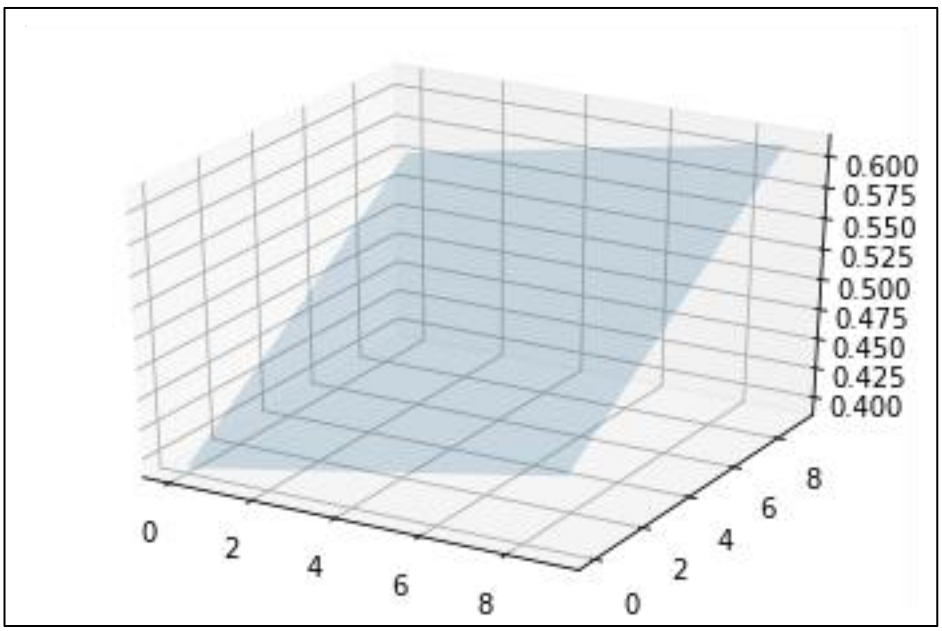
w[2] = 0.99981218

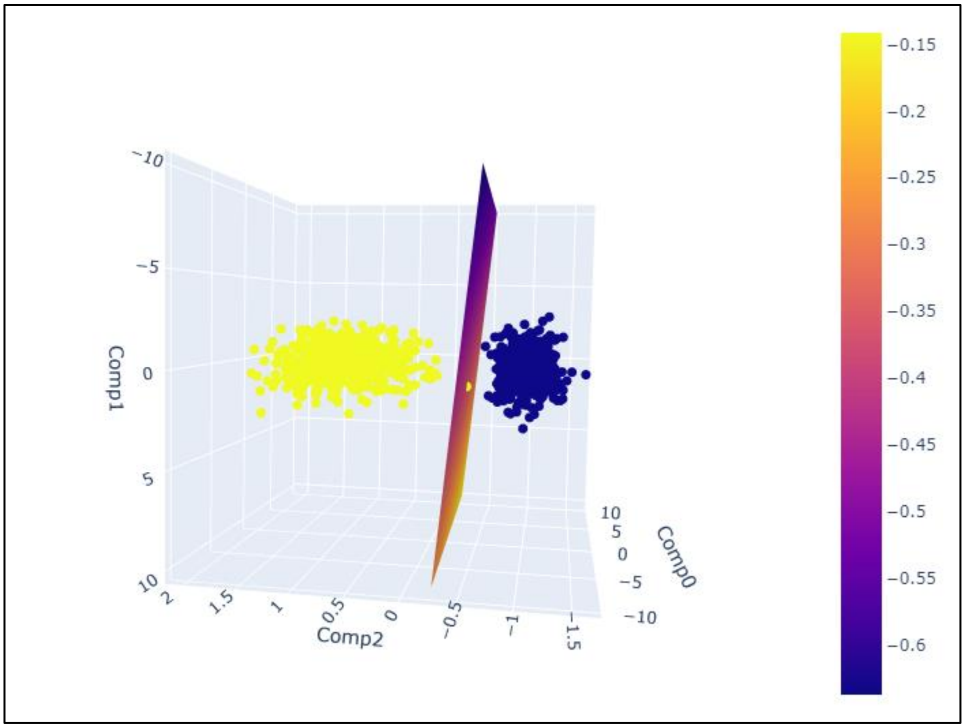
threshold = -0.38930280209937684


Normal distributions. Threshold = -0.38930280209937684
Class 1 (value 1)
Class 2 (value 0)
Discriminant-0.38930280209937684

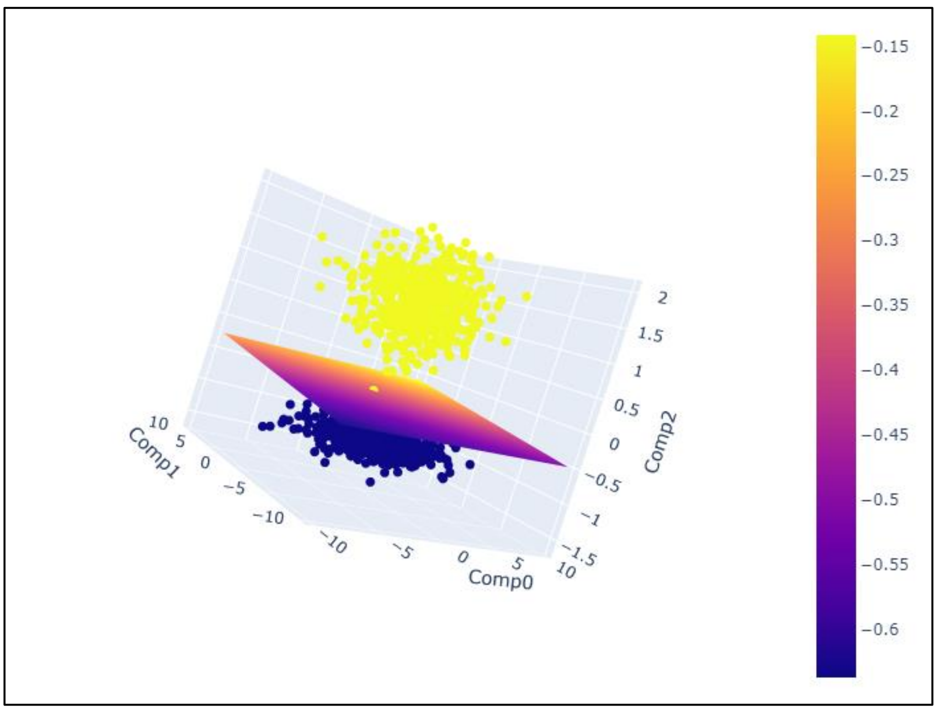

1D Projections. Threshold = -0.38930280209937684
Discriminant-0.38930280209937684


3D Projections
Positive
Negative
Feature 1
Feature 2
Feature 3
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
−1.5
−1.0
−0.5
0.5
1.0
1.5
2.0


0.600
0.575
0.550
0.525
0.500
0.475
0.450
0.425
0.400
0
2
4
6
8



Comp1
Comp2
Comp0



Comp1
Comp0
Comp2

Team Members –

- **PARTH KRISHNA SHARMA – 2017B3A70907H**
- **SIDDHI MAHESH BURSE – 2017B3A70972H**
- **PRAJJWAL VIJAYWARGIYA – 2017B3A70954H**