Fischer Discriminant Analysis

Brief description of the model and its implementation.

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 the between-class variance to the within-class variance.

A large between-class variance means that the projected class averages should be as far apart as possible. On the contrary, a small within-class variance has the effect of keeping the projected data points closer to one another.

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

Between-class variance

Within-class variance

W (our desired transformation) is directly proportional to the inverse of the **within-class covariance** matrix times the difference of the class means.

$$s_k^2 = \sum_{n \in C_k} (y_n - m_k)^2 \qquad y_n = \boldsymbol{W}^T x_n \tag{2}$$

$$J(\boldsymbol{W}) = \frac{\boldsymbol{W}^T S_B \boldsymbol{W}}{\boldsymbol{W}^T S_W \boldsymbol{W}} \tag{3}$$

$$\boldsymbol{W} \propto S_W^{-1}(m_2 - m_1) \tag{4}$$

Per-class mean

Within-class variance

projection equation

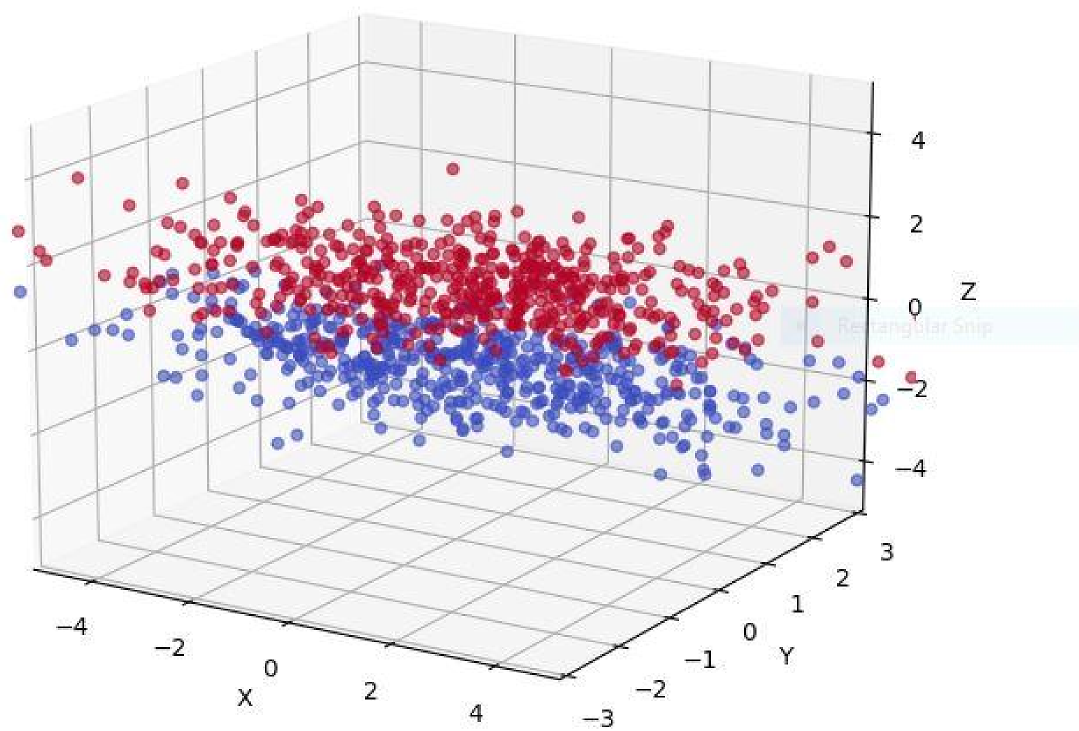
Implementation:

The weight vector is found such that the mean difference between the clusters is maximized and the variance within the clusters is minimized.

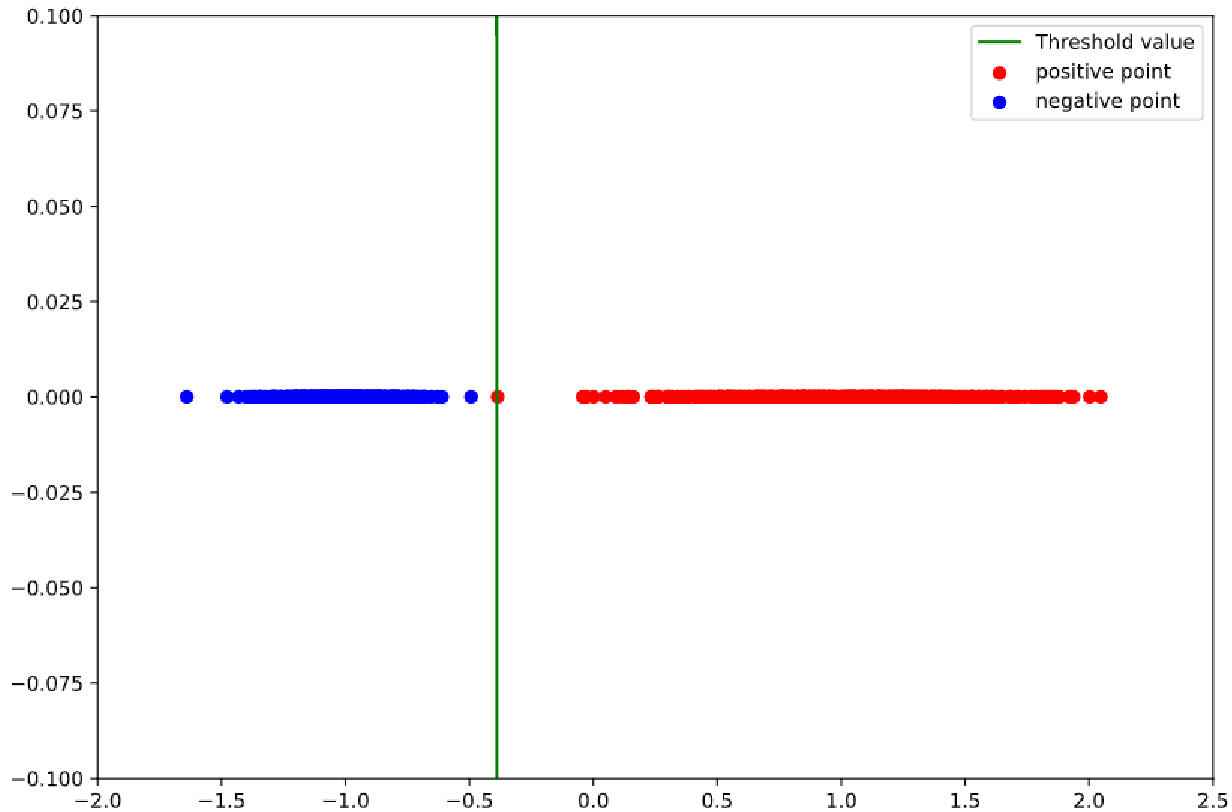
Firstly, the dataset was divided into two parts, one part for each class. Then, the covariance matrix is calculated by using the formula mentioned above. Then , by using the covariance matrix, the weight array is calculated and normalized. Then, the points are projected in 1-D and the point of intersection of their normal curves is calculated. Further, the equation of hyperplane is calculated using the weight array and the corresponding unit vector is calculated and points are projected to that unit vector in 3-D.

Finally, the **training accuracy** was calculated to be 100%. It was calculated by classifying the points and comparing with their corresponding original class to find the number of true positives, true negatives, false positives and false negatives.

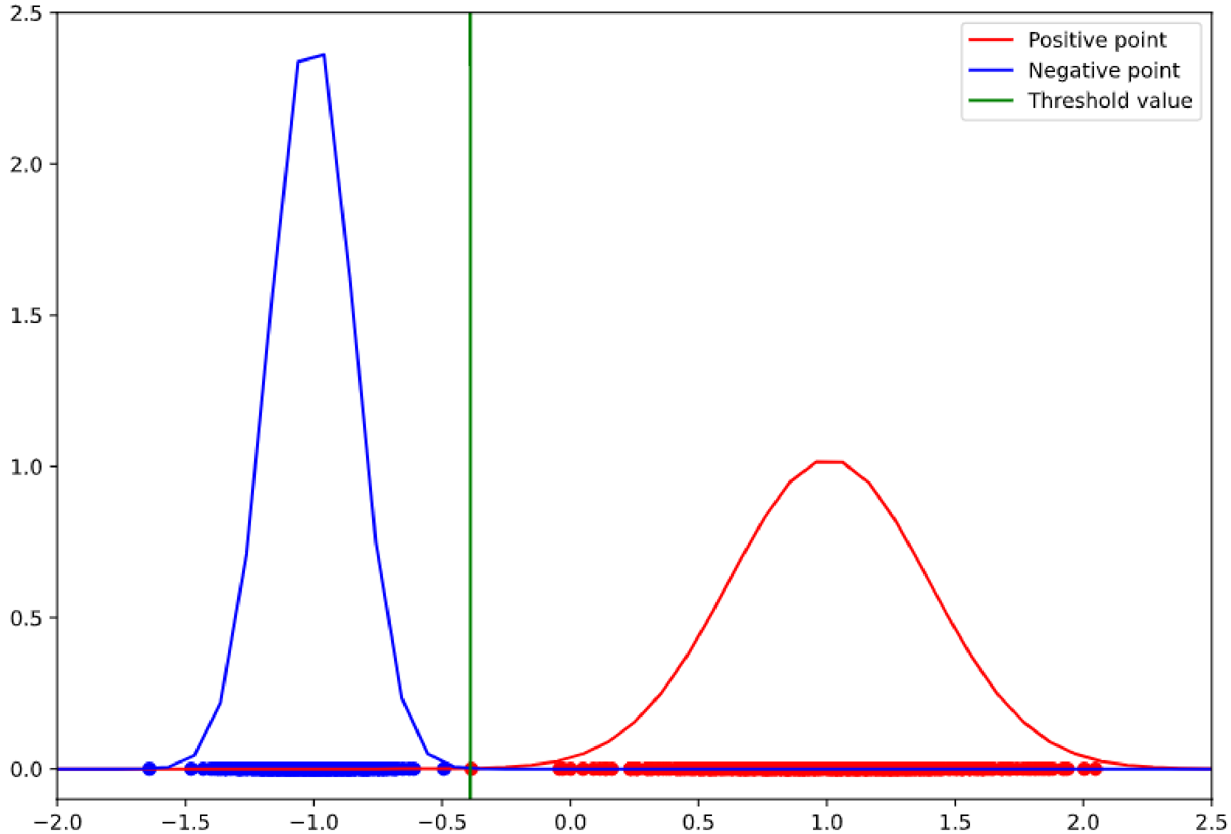
Plot of the higher dimensional data:



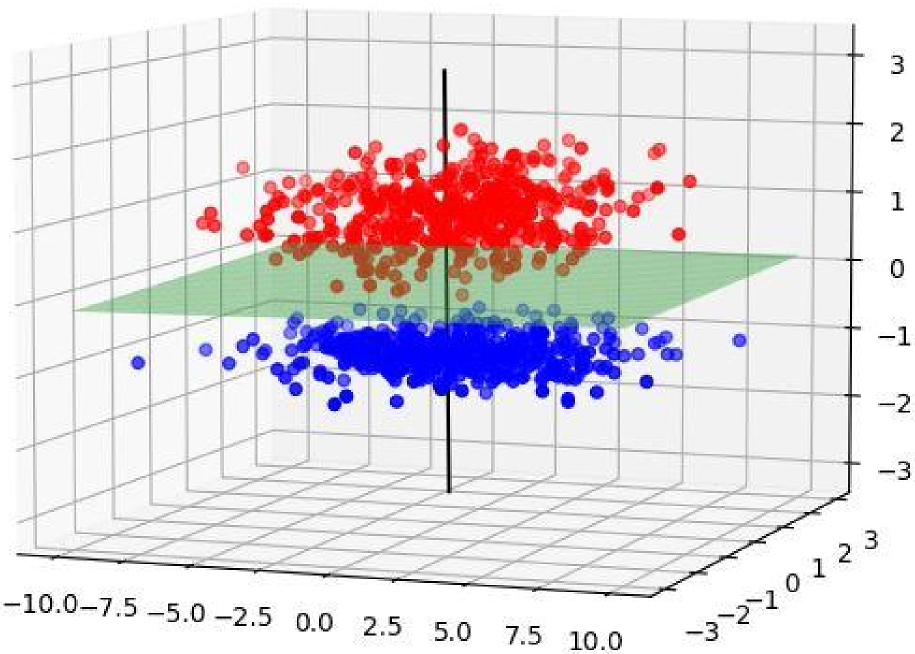
Plot of the reduced clusters and line of intersection in 1-D:



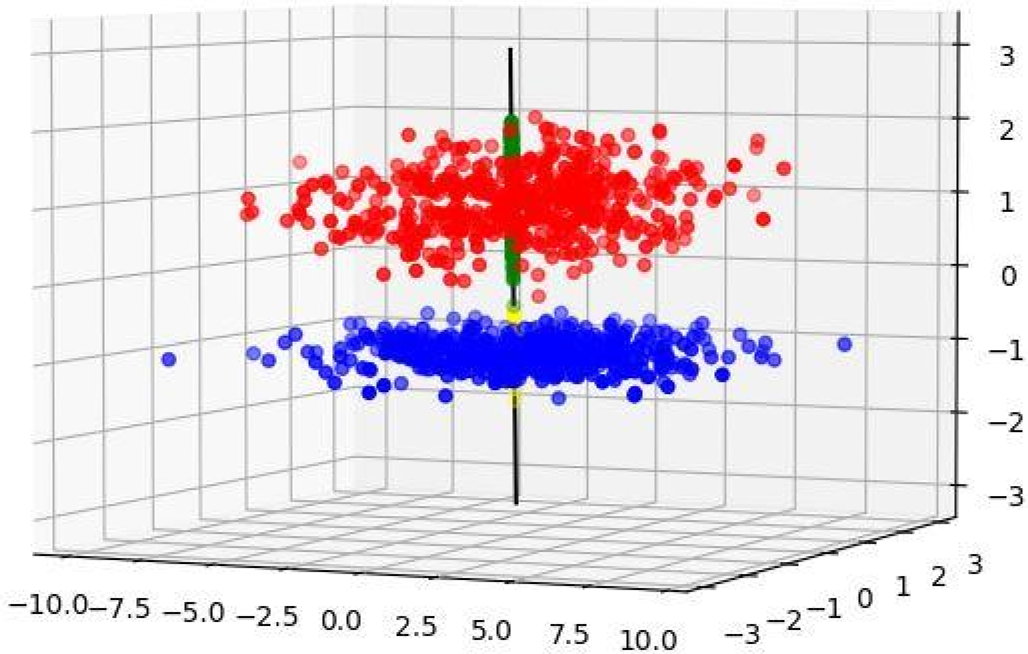
Corresponding Normal Distribution and Line of Intersection:



Unit vector Perpendicular to the Discriminant Plane in 3-D:



Projection of the points on the unit vector in 3-D:



Results

- **The weight array was found to be [-0.00655686, -0.01823739, 0.99981218]**

- **The covariance matrix was calculated to be**
  [18.12275486, -0.19720153, 0.1259276]
  [-0.19720153, 1.97240312, 0.02577773]
  [0.1259276 , 0.02577773, 0.17936747]

- **The point of intersection of the normal curves in 1D is -0.3893028020993765**
- **The equation of hyperplane is 0.006556858371172118x - 0.01823739128274949y + 0.9998121849465029z = 0**
- **Accuracy of the model with training data is 100.0 %**