CS F364 | Machine Learning

JALAJ BANSAL (2017B3A71610H) | ADITYA AGARWAL (2017B1A71075H) | MANISHA KATARIYA (2017B3A70354H)

Assignment 1A: Fisher's Linear Discriminant

Description of Model and Implementation:

DESCRIPTION:

- Fisher's discriminant is a statistical and other fields analysis tool for determining a linear combination of features that characterizes or distinguishes two or more groups of objects or events.
- Two or more classes of the target variable are separated by the hyperplane which is found by using this algorithm.
- The threshold of the projections can be found by Fischer discriminant model, by finding a unit vector on which data points are projected.

IMPLEMENTATION:

- This notebook uses NumPy, pandas and matplotlib libraries to implement the algorithm
- Since the dataset has three features, we use matplotlib's 3D plotting for visualization of the data and the hyperplane.
- First, we loaded the data and split the data and labels into NumPy array
- Then we visualized the data in 3D.
- Next, we found M1 and M2 where they both are means of both classes.
- Next, we tried to find the matrix Sw by initializing it initially as zero matrix and going through both the classes by collecting data with that label.

  The formula used was:
  $S_w = \sum_{n \in C_1} (x_n - M_1)(x_n - M_1)^T + \sum_{n \in C_2} (x_n - M_2)(x_n - M_2)^T$ where M_1 and M_2 are means of data points.
- Next, we find inverse of Sw.
- Next, w is found by: $w \propto S_w^{-1}(M_1 - M_2)$ where M1 and M2 are the means of the projections of both classes calculated above and inverse of Sw also calculated above.

- Once the projections are calculated the respective projections of each class are then fitted to two normal distributions whose parameters are determined by maximum likelihood estimation.
- Then the two normal distributions are plotted together, the threshold is found by finding the intersection of the normal distributions. This is done by equating and finding the roots of a quadratic equation.
- Then we initialize the x and y equally distant coordinates and find z coordinate from the discriminant vector found above.
- Then we plot the hyperplane in 3D with the data points of both classes.
- Lastly, we will find the accuracy of our Fischer Discriminant Model.

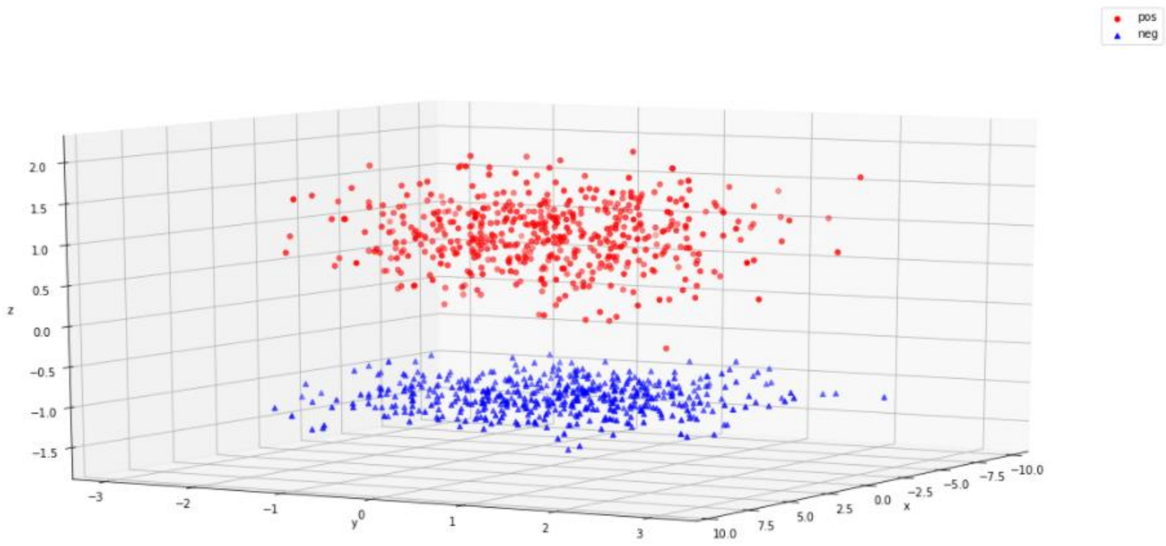
Output:

- **DATA**

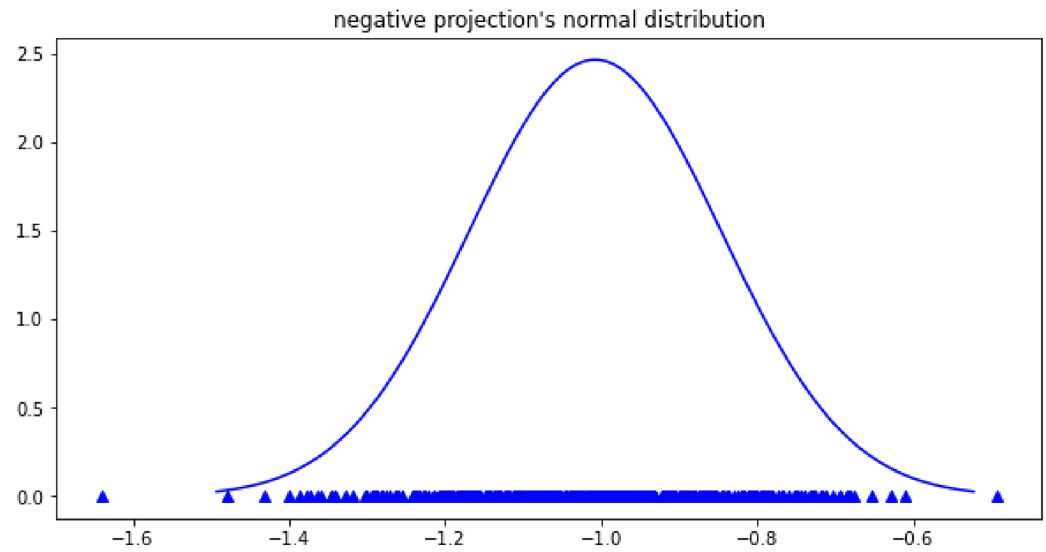
	feature 1	feature 2	feature 3	label
0	-6.672418	-1.206198	-1.081050	0
1	1.675598	0.614994	-0.971600	0
2	-4.039058	0.335102	0.544618	1
3	0.793526	-0.235277	0.551771	1
4	3.820273	-0.274691	0.454743	1
...	...	...	...	...
995	-3.680139	0.966962	-0.904337	0
996	-4.063900	0.802611	1.023708	1
997	-0.814430	-0.693945	0.876776	1
998	-0.325122	-0.759024	1.299772	1
999	-1.503431	-0.269458	-1.124390	0

1000 rows × 4 columns

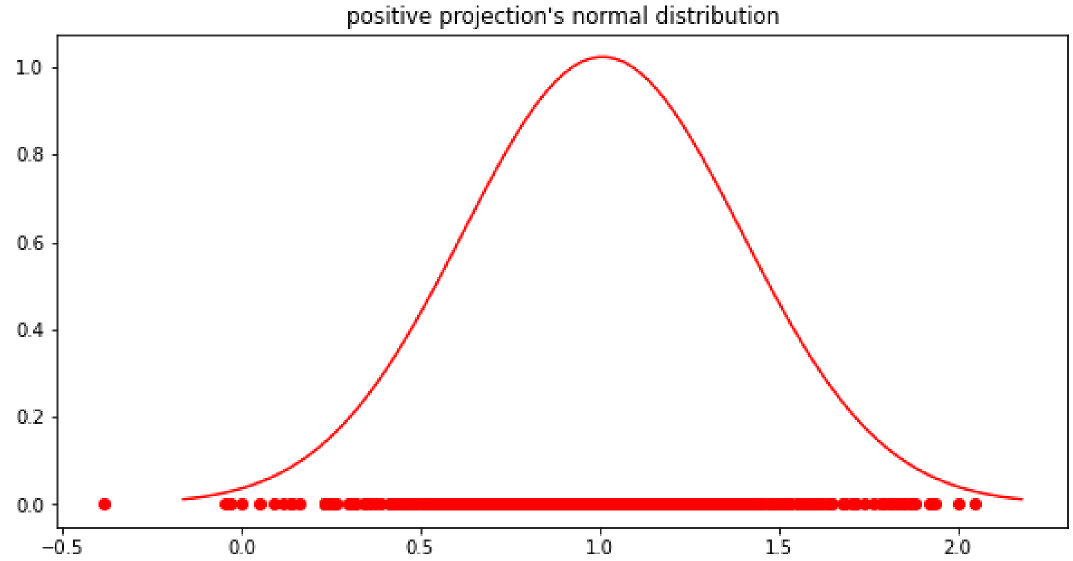
- **SCATTER PLOT OF POINTS IN 3 DIMENSIONS**



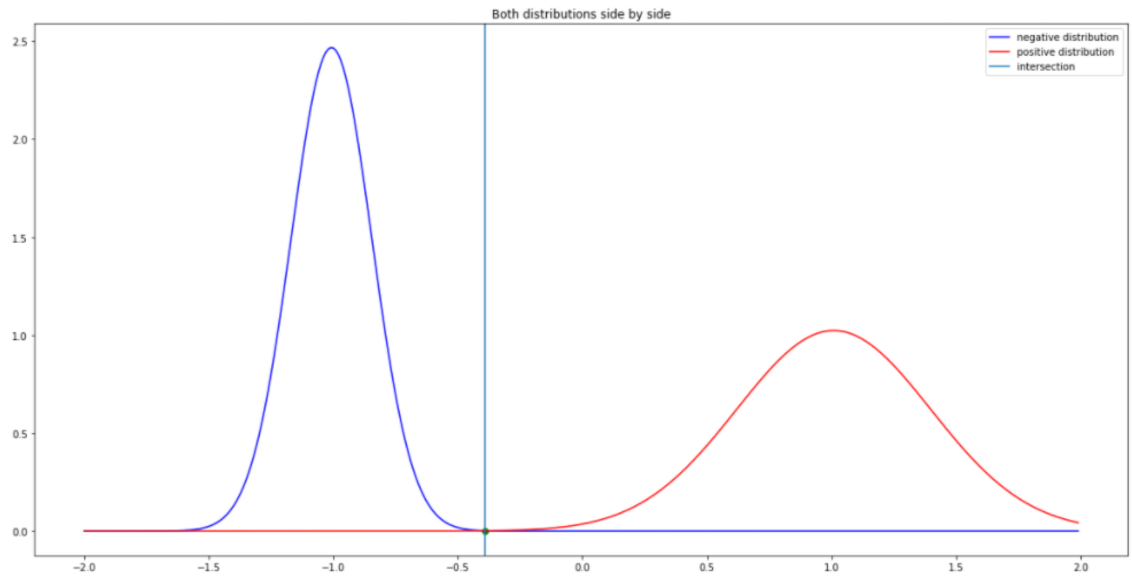
- **NORMAL DISTRIBUTION OF DATA (LABEL 0)**



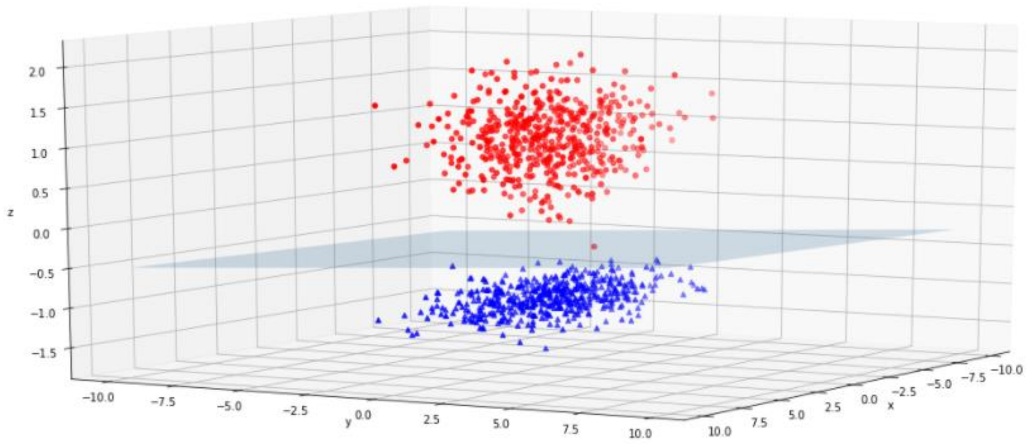
- **NORMAL DISTRIBUTION OF DATA (LABEL 1)**



- **BOTH DISTRIBUTION SIDE BY SIDE WITH INTERSECTION LINE IN 1D**



- **CLUSTER AND THE SEPERATING HYPERPLANE**



- **We obtained 100% accuracy which shows that our dataset is Linearly separable.**

LIMITATIONS:

The fact that FLD is limited to problems with linearly separable features seriously restricts its usefulness. Fisher's quadratic discriminant analysis is an extension of FLD that allows problems to be separated based on quadratic class boundaries.