BITS F464 Machine Learning
Assignment - 1

Dev Gupta - 2017B3A71082H
Garvit Kashyap - 2019A7PS0078H
Kshitij Upadhyay - 2019A7PS0105H

1 Fishcher's Linear Disrcriminant

1.1 Model Description

Fischer's Algorithm is a dicriminative algorithm that finds projections to a line such that different classes are well separated. It does this by maximizing the difference between the means of the classes and reducing the variance(scatter). Quantity we have to maximize is :

$$\frac{(m_1 - m_2)^2}{(s_1 + s_2)}$$

$$S_B.W \propto S_W.W$$

$$w \propto Inv(S_W) \, * \, (M_2 - M_1)$$

M_1 : Average of points of class 1
M_2 : Average of points of class 2
S_W : Sum of variances of both classes
S_B : $(M_1 - M_2)(M_1 - M2)^T$
w: Direction *on which we project the points*
m$_1$: Average of points of class 1 after projection
m_2 : Average of points of class 2 after projection
s_1 : Standard Deviation of points of class 1 after projection
s_2 : Standard deviation of points of class 2 after projection

1.2 Accuracy

Accuracy achieved for the dataset : 100%

1.3 Plots of Data

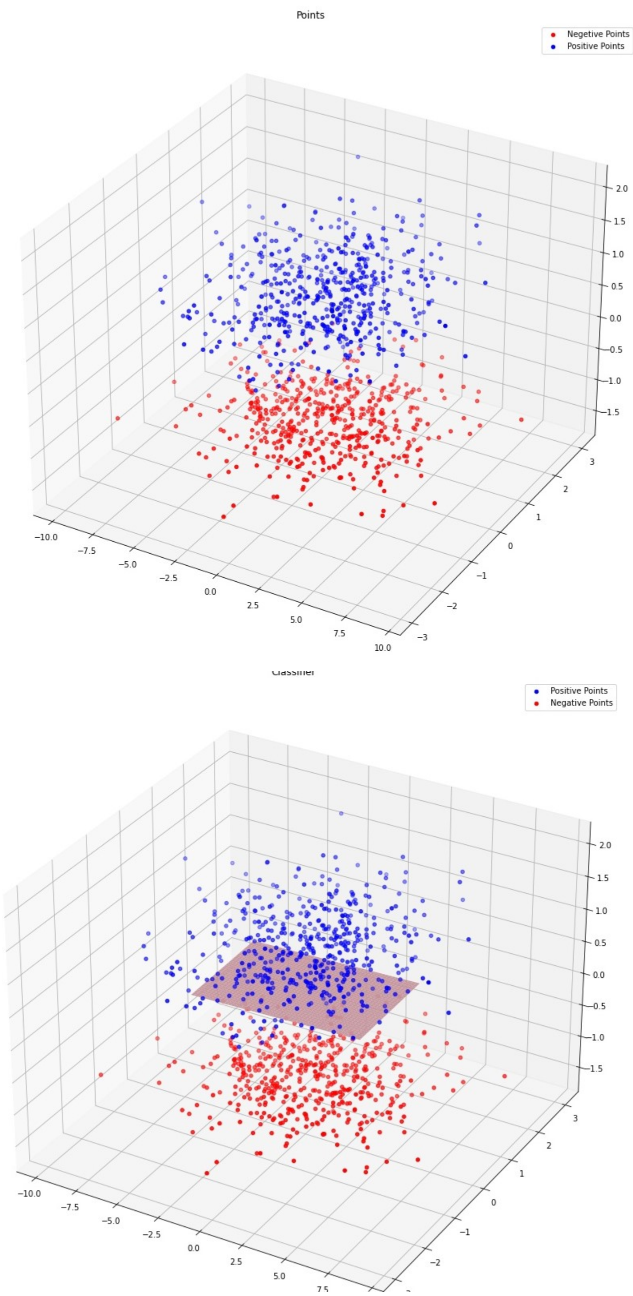

Figure 1: 3-D Plots of data with and without hyperplane

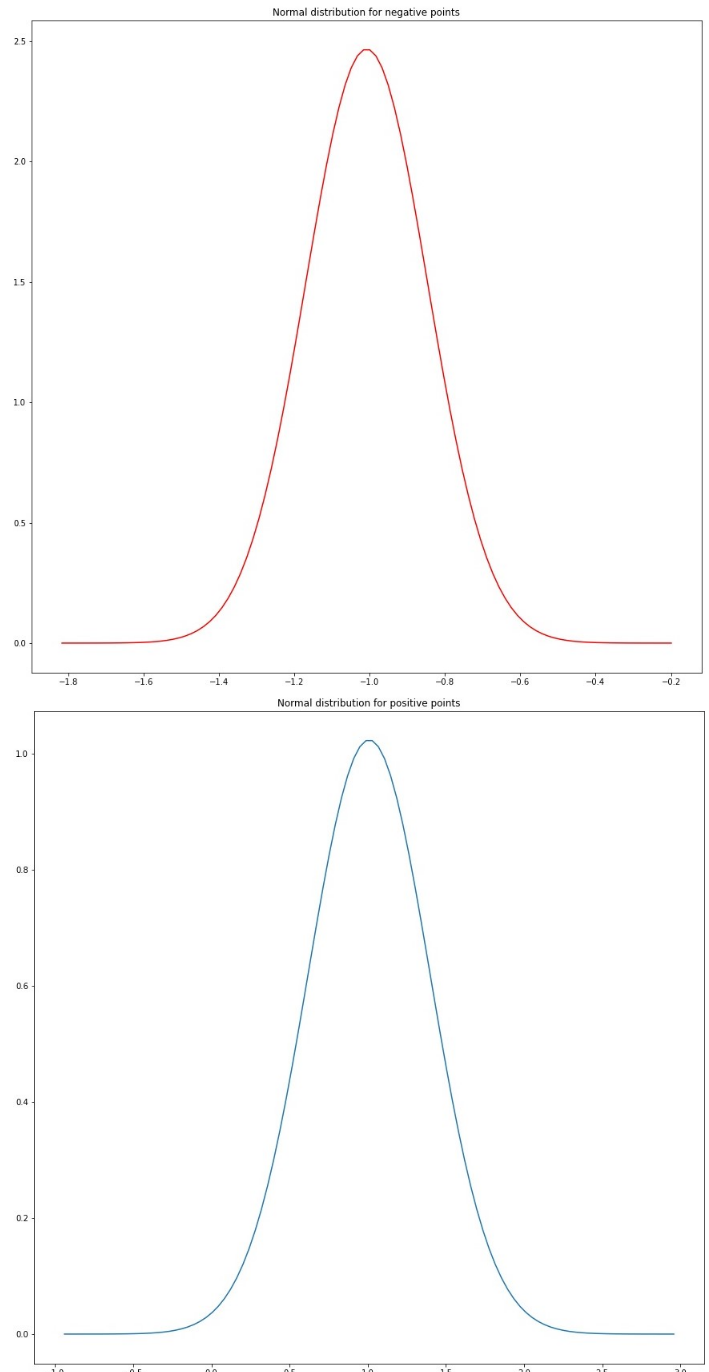

Figure 2: Normal Distributions of both the classes

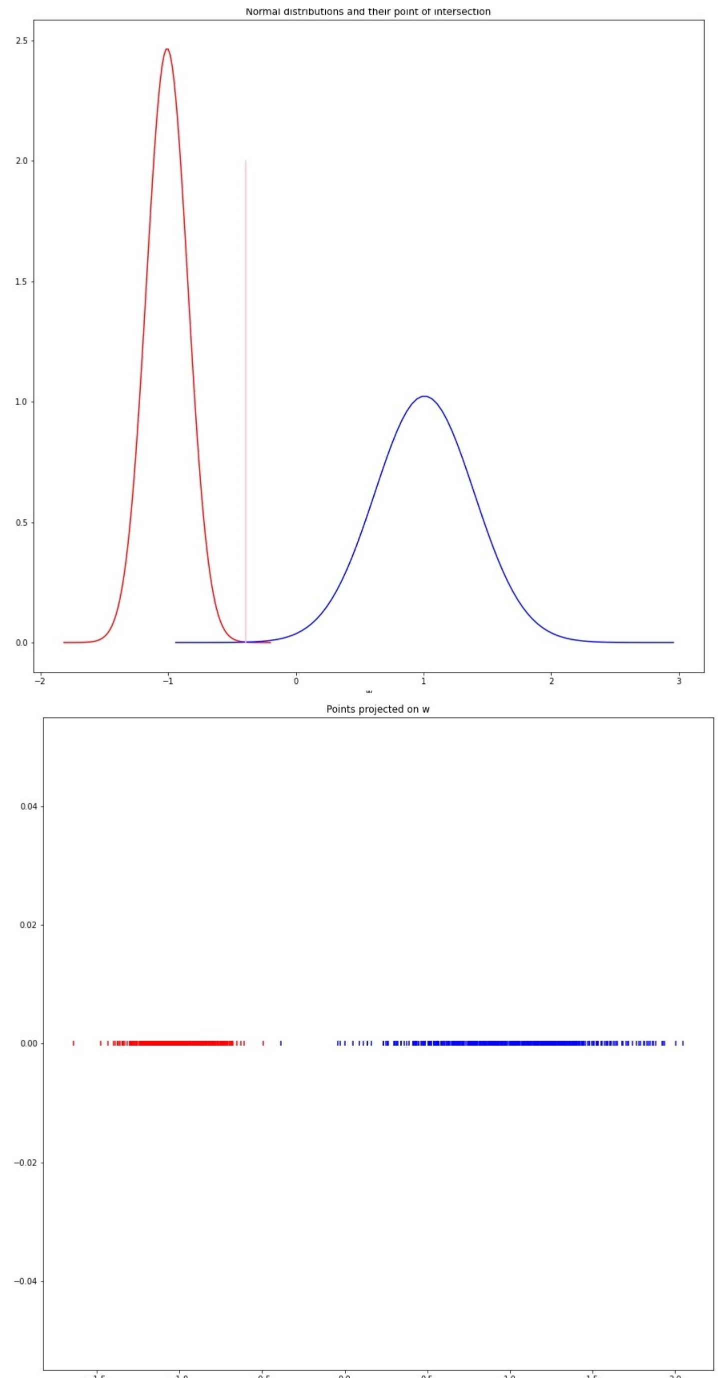

Figure 3: Projection in 1D and their distributions with Discriminant line

2 Naive Bayes

2.1 Model Description

The model executes 3 steps:

1. The first step is preprocessing of the data where we remove the certain stop words and change the data into a dataframe with the second column as classification
2. The Second step involved splitting the dataset into 7 parts for the 7-fold cross validation
3. The final step was training and testing of the model. For this we first created a set of all words in the dataset, then we created the dictionary where the keys were initialized to 1 and updated to indicate the probability of a word appearing in positive and negative classes in the training data. Using this we classify a new data depending on which classification is more likely

2.2 Accuracy

The accuracies are consistently around 80% for each fold. Since the dataset is shuffled each time but in the given run, the accuracies were, 0.79 0.74 0.80 0.79 0.78 0.81 0.83 Giving an average of 0.79

2.3 Limitations

The major limitation of the Naïve Bayes classification is that it assumes conditional independence of each parameter. Thus, in the current model, the probability of "huge" and "offer" are taken to be independent even though it is quite likely that $P(huge/offer) > P(huge)$

3 Linear Perceptron Algorithm

3.1 Model Summary

In the linear Perceptron algorithm, we assume that the data is linearly separable and for every misclassified example we decrease the error using the gradient descent algorithm.

$X = \ input\ data \quad size = (number\ of\ examples,\ number\ of\ features)$
$w = \ weights \quad size = (number\ of\ features)$
$y = \ labels \quad size = (number\ of\ examples)$
$t = \{-1, 1\} \quad (-1\ for\ y = 0\ and\ 1\ for\ y = 1)$
$Loss\ for\ misclassified\ example\ =\ t(w^t.x)$

$Update\ step\ for\ weights$
$w := w\ +\ learning\_rate * t\ *\ (x)$

3.2 Accuracy

1. Dataset - 1
   - Training Data - 98.125%
   - Test Data - 98.058%
2. Dataset - 2
   - Training Data - 100%
   - Test Data - 100%

3.3 Inference from Data

The accuracy for the second dataset is 100% for both training and testing data, hence it is more lienarly separable as compared to the first dataset.

3.4 Limitations

Perceptron algorithm can only successfully classify data that is linearly separable. So if no line/hyperplane exists to separate the two classes, the Perceptron algorithm won't be able to give an exact solution.