ML Assignment 2 report

Harjeet Singh Yadav2020561

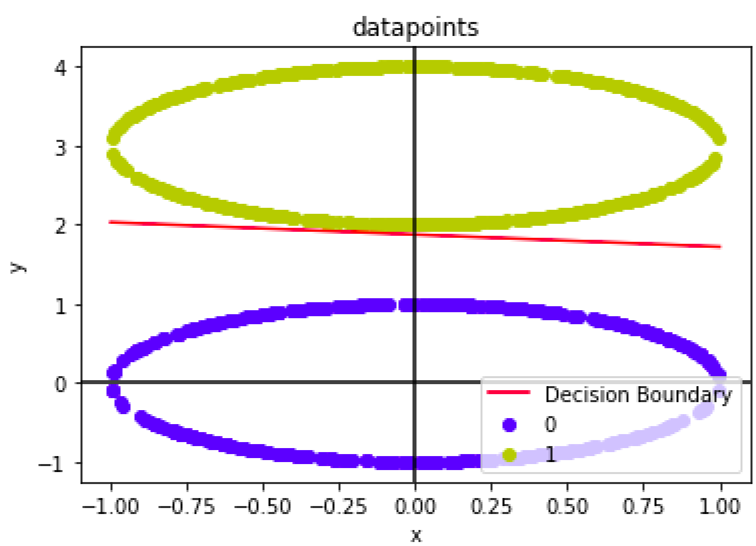
Section B

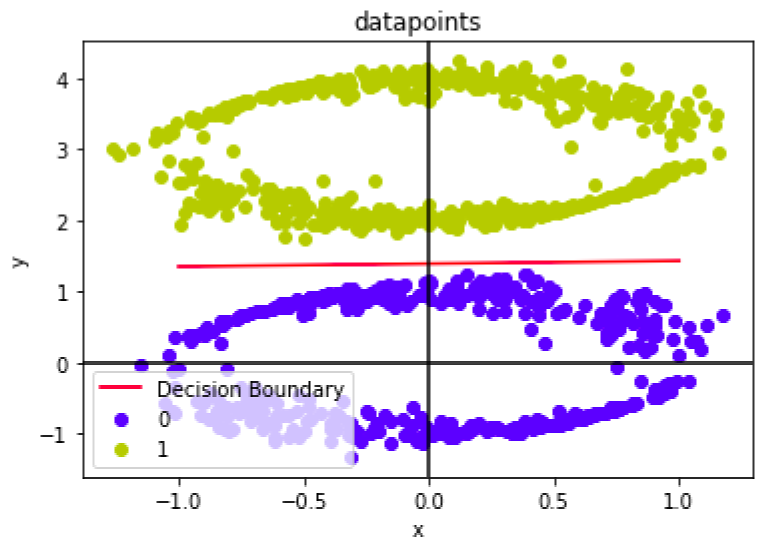
1) Adding noise

   Noise is coming from a gaussian distribution with a mean of 0 and a standard deviation of 0.1. That means noise values vary from -0.1 to +0.1. We add these values to the data points generated whenever add noise is True.

3)

Decision boundary exists in both cases of whether noise exists in the data or not. This is because the perceptron training algorithm (PTA) weights are moving in the right direction after every iteration. This is similar to the gradient descent algorithm, we subtract the error times the weights from the current weights. This will stop when there will be no change in the weights in the subsequent iterations which means we have reached the correct value of weights. The obtained hyperplane can classify the given data or the decision boundary exists.

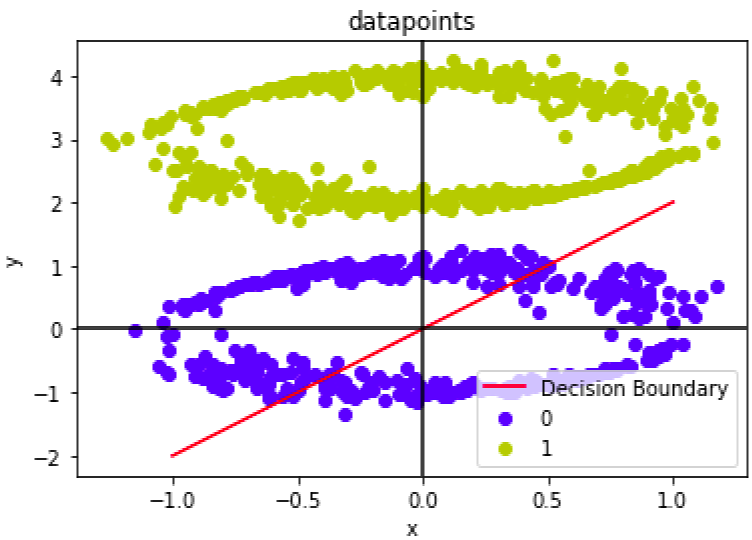


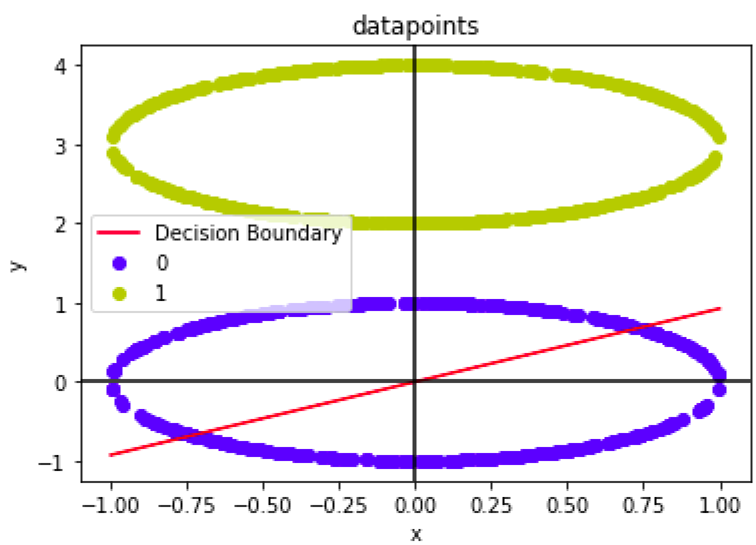


4)
As we can see in the previous part decision boundary exists when the bias is there.
This is because
the weight shows the effectiveness of a particular input. The more the weight of input, the more it will have an impact on the network.
On the other hand, Bias is like the intercept added in a linear equation. It is an additional parameter in the neurons which is used to adjust the output along with the weighted sum of the inputs to the neuron. Therefore Bias is a constant which helps the model in a way that can fit best for the given data.

Geometry analysis

Moreover, when the bias is zero the hyperplane will always pass through the **origin** no matter the dimension of the data i.e no of features in the data. But the data cannot be at the origin always as we can see when the centre of the circle is (0,3). This is because the weight decides the direction of the hyperplane whereas the bias shifts the hyperplane or the decision boundary parallelly. This can help to fit the data by shifting the hyperplane parallelly.





6) Given the hyperplane h(x) = 0, we can assume that class 1 on the h(x) > 0 side of the hyperplane and class 0 lies on h(x)< 0 side of the hyperplane.
Given the point, we put that point in the hyperplane equation and find the class if h(x) is less than or greater than 0. If the point lies on the plane itself then we can’t decide.

Section C

1)

The accuracy is better the entropy case for depth 10.

2)

The performance in the b) part is better than in the a) part because we are using 100 weak classifiers or learners in the b part and taking the majority voting. In this case, we have kind of pruned the three with a max depth of 3 in all the 100 stumps this avoids the overfitting of the data and reduces the variance. Hence it works better on the unseen data (test and validation set). In the a) there is only one decision tree and it very sensitive with testing data can overfit the data, and perform bad on the unseen data. This is the reason why accuracy is better on the ensembling learning.

3)

The accuracy on both technique is almost equal with a difference less than 0.1 this is because the nature of the dataset given, where labels of more the 80% of the data is same i.e “white”. This is reason why accuracy is so high (99.2 %)and almost similar in all the criterion and depths.

Ideally, AdaBoost is better because we do bootsting + random feature selection to reduce the variance and no of depths is 1 or 2 in all the stumps.
Random feature selection reduces the correlation between the feature and reduces the sensitivity to the training data unlike random forests where we use all ther feature for the stumps.