ECEN 689 : Challenge 5 report

Ashish Kumar

I. Objective

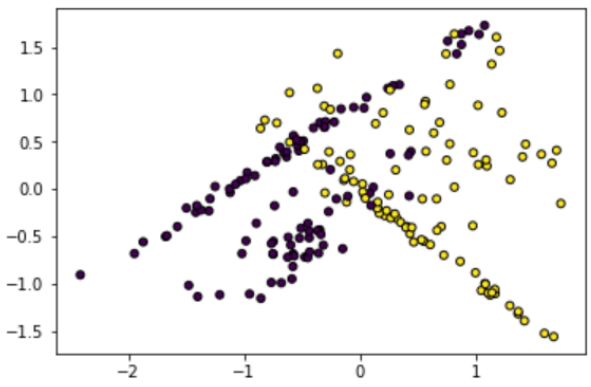
Build Support Vector Machine (SVM) based binary Classifier for the given data set. The task is to justify also why particular kernel is being used in this challenge.

II. Methodology

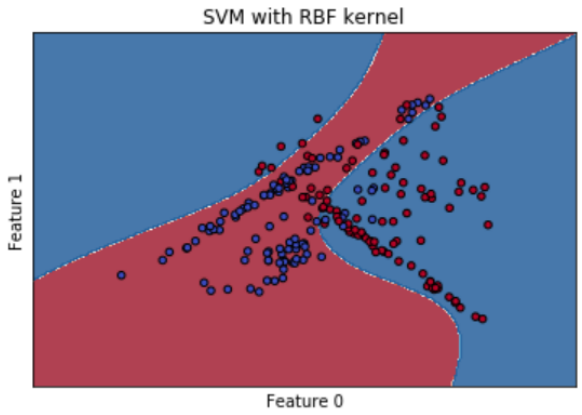
Sklearn library is being used and it has support of 4 kernels: Linear, Gaussian, Polynomial and Sigmoid. The first task is to analyze the dataset. It has two features while number of training dataset is 200. So, sigmoid kernel is ruled out as it works best when there are more features than the length of data set. From the figure 1a we can see that data is distributed randomly. Preliminary intuition is either Gaussian or polynomial kernel may be best fit. Linear should not be the best fir for the data set. In the implementation of code all 3 kernels except sigmoid is being searched. Polynomial function with degree 2 does not give the best accuracy on the data set. Same is true for linear kernel. Gaussian kernel gives the best accuracy.
Gaussian function is $exp(-\gamma||x - x\prime||^2)$. There is another parameter C which needs to searched. Accuracy is best for Gaussian which is also called RBF with $\gamma = 5600$ and $C = 0.1$.

III. Conclusion

Gaussian kernel gives the maximum accuracy of 88.5% on the training data set. Decision boundary plot is shown in the figure 1b.



(a) Training Data set



(b) Decision Boundary for SVM on training data set

Fig. 1: Scatter Plots