# Challenge 5: SVM Classification
Rishi Laddha

Introduction

The dataset for the challenge had two features and two distinct classes(0,1). Training as well as testing datasets had 200 data points. (Illustration 1) SVM (Support Vector Machine Classification) model could be used to classify the data. SVM is a type of supervised learning model which uses hyper planes to distinguish classes. Kernel functions are used to solve problems by transforming non-linear observation to higher dimensional space to make it separable.
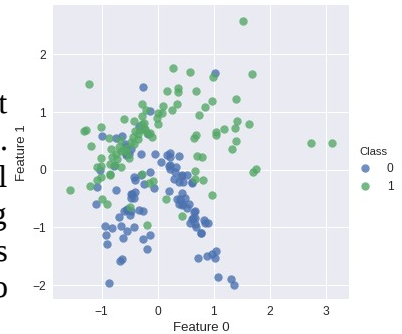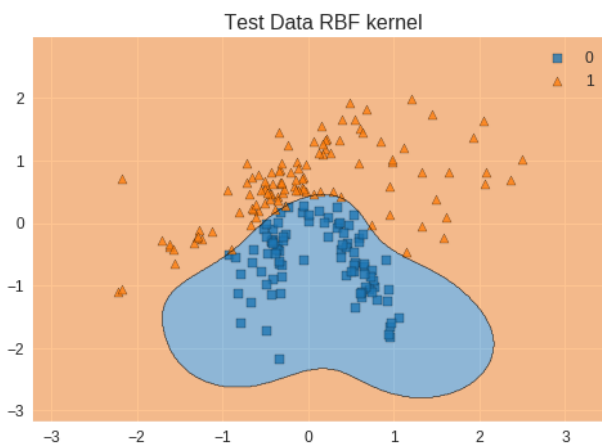

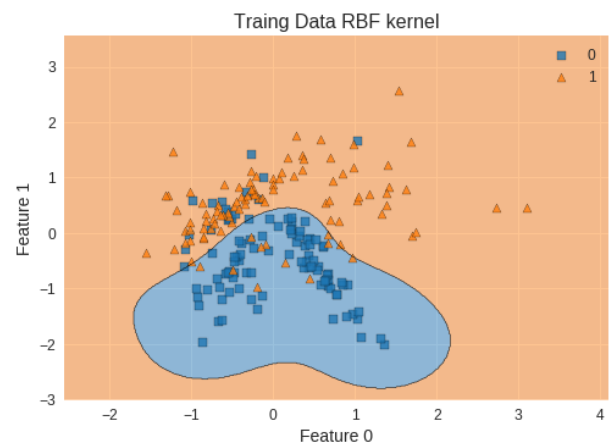*Illustration 1: Training Dataset*

Method

For the classification and prediction, different kernels were tested with varying penalty parameter C and gamma values so as to obtain best cross validation accuracy. Cross validation could be used readily because of the small data set size as it will be less time consuming. To test the best possible combination of C and gamma grid search method was performed. Different kernels were tested for the best model selection. the kernels are part of Scikit-learn library in python.

Result

RBF(Radial Basis Fuction) kernel outperformed all other with an accuracy of 85.5% . for C=3.4304 and gamma= 0.70716. The prediction was done on the test dataset using these hyper parameters.
The plot below shows the boundary that separates the two classes in test data (Illustraion 2) and Train data (illustration 1).

Illustration 2: Test Data RBF Boundary

Illustration 3: Train Data RBF Boundary

Conclusion

The non linear boundary outperformed the linear hyperplane for the dataset.