SVM Classification (challenge 5)

Jatin Kamnani

Problem statement*— Given a training data, use a SVM classifier and predict the classes of the test data, using a specific kernel function and explain why that kernel function is used.** ***Keywords*— *Radial Basis function(RBF), SVM, polynomial, decision boundary

I. Introduction

The training data consists of data divided into two classes(binary classification), and fig1. Shows how the data looks.

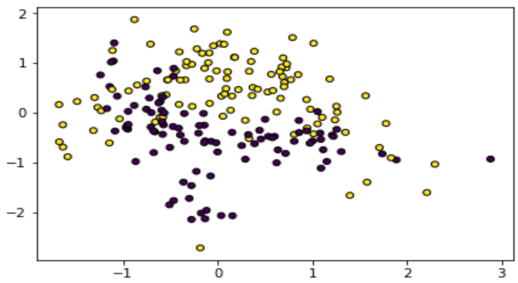

Fig1. Scatter plot of the training data

II. Decision Bondary

It is easier to visualize figure 1 and reach to a conclusion that a 'linear boundary' cannot separate the two classes optimally, so we need to look for some classifiers that can separate the data using some non linear decision boundaries.

The problem statement, asks us to use a SVM classifier; and so we can easily test and train different kernel functions(linear and non-linear decision boundary estimators) and see which kernel gives the highest training accuracy. Sklearn's SVM api, gives us three kernels essentially, viz RBF, Linear, Polynomial

III. Scores of different kernels with default params

I trained the SVM classifier using three kernel functions through the default parameters, and tried to a gain a perspective, that which kernel can give the highest accuracy. The scatter plots of those with the respective decision boundaries are shown below. Linear kernel showed a training accuracy of 77.5% whereas Polynomial kernel showed only 58% training accuracy. The highest accuracy was achieved by the RBF kernel function 82.5%; which no doubt generates a non linear decision boundary with ease for complex datasets.

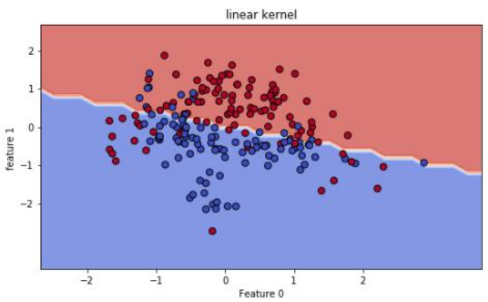

Fig2. Linear separator kernel

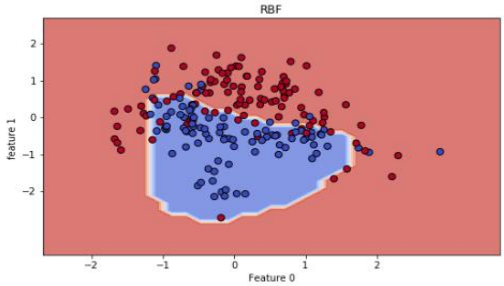

Fig3. RBF kernel

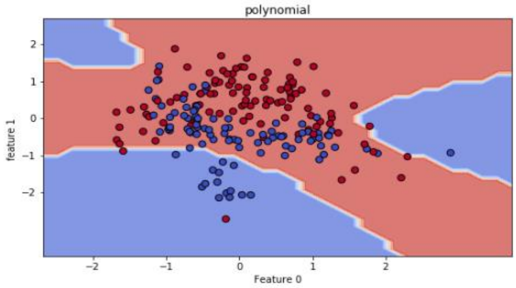

Fig4. Polynomial kernel

IV. RBF Kernel

The RBF kernel is given by the formula:

$$\exp(-\gamma\|x - x'\|^2).$$

V. Optimizing parameters

Fine tuning the hyper parameters is an essential procedure in machine learning problems, and once I tuned the hyper parameters viz gamma and C(regularization constant), I was able to achieve a 4% increase in the accuracy. The scatter plot with the decision boundary of the final classifier is shown below

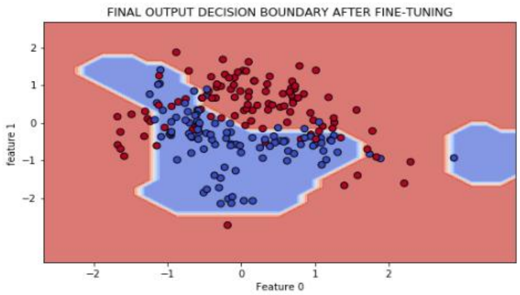

Fig5. RBF kernel after fine tuning

VI. Conclusion

In the given problem, I used a RBF kernel with gamma equal to 1.2 and regularization parameter as 1.8. The score that was achieved on the training data(86%) was much more than any other kernel and thus, the test data classes were predicted using the same.