# Challenge 5 Report – Support Vector Machines

Shabarish Kumar Rajendra Prasad

## Problem

The challenge comprises of a binary classification problem, where we are asked to employ Support Vector Machine (SVM) to classify the entries based on two features (Feature 0 and Feature 1) and also to use the correct kernel so as to mark a clear decision boundary given the samples.

## Kernel Used

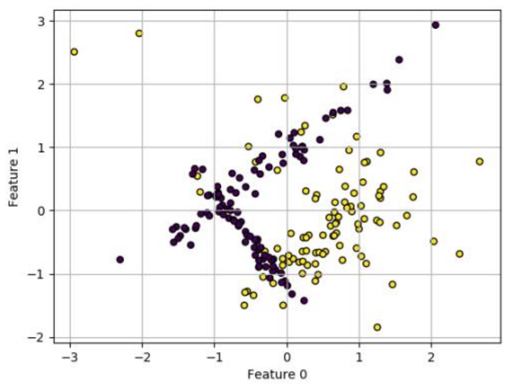
I have used the Radial Basis Function (RBF) kernel on the samples on my problem. The radial basis function kernel is a squared exponential kernel and the rbf kernel on two samples x and x' is given as:

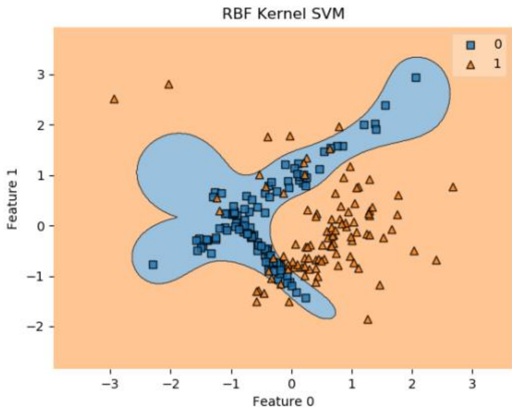
$$K(x,x') = \exp\left( - \frac{|x-x'|2}{2\ \sigma 2}\right)$$

Where $|x-x'|^2$ is the squared Euclidian distance between the two vectors and $\sigma$ free parameter. The kernel transforms the points in the linear space into a non-linear space and tries to find the decision boundary in the non-linear space.

## Why Gaussian Radial Basis Function (RBF) Kernel?

The diagram shows the scatterplot between the two features in the training set:




Clearly, there is no clear linear boundary between the two classes and hence the linear kernel would not work well. For the dataset which is not linearly separable, using a polynomial or radial basis function kernel would be more fitting to mark a decision boundary. The polynomial kernel would be a parametric method of approaching the problem whereas the rbf kernel is a nonparametric method. Using a nonparametric method means that the complexity of the model is potentially infinite, the complexity of the model grows with the data. Whereas when a parametric method, the model tends to saturate after a certain point. While that is what we would ideally want, our model to be simple and not really complex, given a classification problem where the classes are tangled within each other, we may want to fit the model close to one of the classes.

The diagram shows the decision boundary of an SVM employed on the training dataset using the RBF kernel marked on the cartesian co-ordinate system. We can see that the RBF kernel forces the decision boundary close to class 0 and hence achieves a training loss of 0.91.