

Support Vector Machine

Kishan Shah
Texas A&M University

Abstract-The objective of this challenge was to perform classification of given testing file using support vector machine. The challenge is to identify the most suitable parameters associated with support vector machines like kernel type, penalty parameter C, kernel coefficient gamma etc. Radial kernel gave the best classification accuracy on the training data set when grid search CV was performed on that, and it was used to predict the class of the testing data set.

I. INTRODUCTION

In machine learning, support vector machine is supervised learning algorithm which can be used for both classification and regression analysis. However, it is mostly used for classification [1]. Hyper plane is used to separate two or more classes in the feature space. The margin between the two classes should as maximum as possible to obtain more robust model. In some cases where two classes are not linearly separable, the learning of hyperplane in SVM is done by transforming the problem using the concept called kernel. There are various types of kernel called linear kernel, radial basis kernel, polynomial kernel etc. The choice of kernel can be automated by optimizing a cross-validation based model selection.

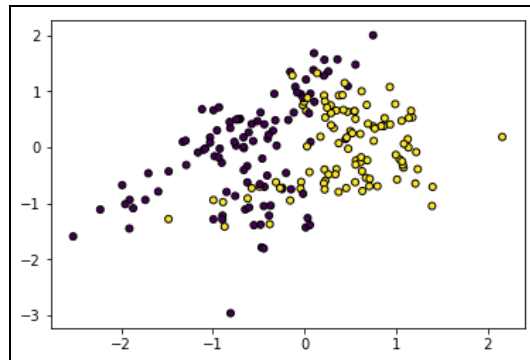


Figure 1. Scatter plot of training data

II. METHODOLOGY

From the Figure 1, linear classifier does not give better classification accuracy as data points are not linearly separated. To decide which kernel and related parameters (penalty parameter C, gamma) are the most suitable for the given data set, cross validation based model is used. Different types of kernel i.e. linear, polynomial, radial basis function, sigmoid are used. Grid search CV is performed using ten different values of penalty parameter C and Gamma for each type of kernel where larger the value of C gives smaller margin and the gamma parameter defines how far the influence of single training example reaches [2].

A commonly used kernel is the Radial Basis Function which is defined by as follows which means RBF kernel function on two sample points \mathbf{x} , \mathbf{x}' . σ is a free parameter.

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

III. RESULT & DISCUSSION

Table 1. Best model for each type of kernel

Kernel Type	C	Gamma	Accuracy (%)
Linear	3.334	0.001	84.5
Polynomial	5.556	1.1119	82
RBF	1.1119	1.1119	86
Sigmoid	5.556	0.001	82.5

From Table 1, we can say that the RBF kernel gives the highest classification accuracy for $C = 1.1119$ and $\text{Gamma} = 1.1119$.

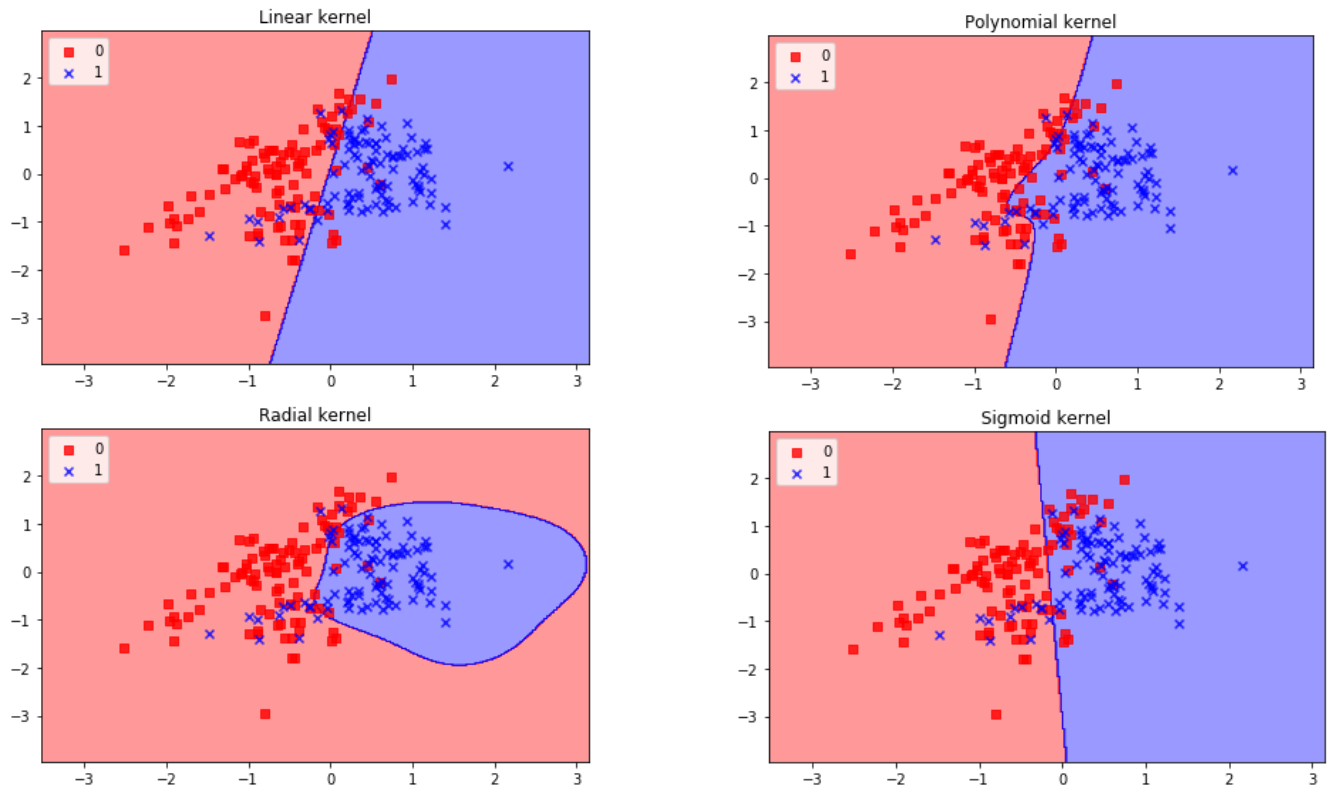


Figure 2. Training decision boundary for each type of kernel

From the boundary plot shown in Figure 2, RBF seems the best kernel to separate two different classes of training data set. So, the same parameters (kernel = RBF, $C = 1.1119$, $\text{Gamma} = 1.1119$) are used to predict the test data set.

IV. REFERENCES

- [1] https://en.wikipedia.org/wiki/Support_vector_machine
- [2] <https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72>