Performance of Support vector Machine

Shrikant patro,

School of Computer Science and Engineering, VIT Chennai, Tamil Nadu, India 600127

Email:shrikantjagannath.2018@vitstudent.ac.in

Abstract

The Classification is one of the fundamental task in case of the machine learning algorithm. The Support Vector machine is very powerful and mathematically proven model for the linear classification task. This paper will cover the fundamental of the Support vector machine. The mathematical background and the related mathematical calculations of the Support Vector Machine. The Principal of the working of the support vector machine. The brief description of the algorithm how the SVM is applied on the woman's fertility dataset.

Introduction

Support vector Machine is one of the widely used algorithm that is extension of the Perceptron Learning algorithm. The goal of the perceptron algorithm is to minimize the misclassification errors. The goal of the SVM is to maximize the margin. The margin is defined as the distance between the separating hyperplane (decision boundary) and the training samples that are closest to this hyperplane, which are the so-called support vectors. The reason behind maximization of the difference between the separating hyperplane and the margin is that it reduces the generalization error and while the model with the smaller margin are more prone to fitting. The figure 1 will illustrate the hyperplane and Maximization of the Margin using SVM.

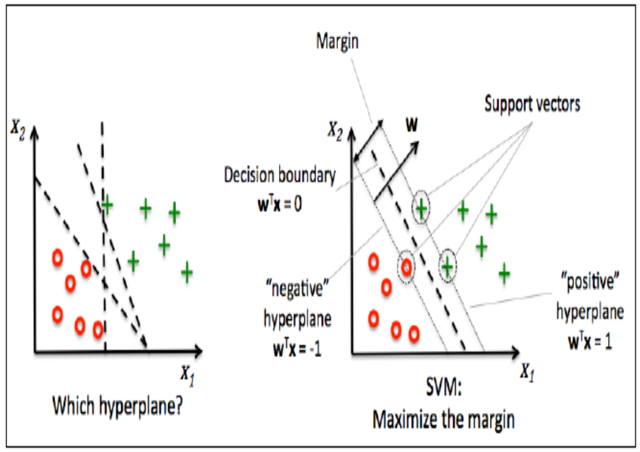

Fig 1 Support Vector Machine

Database- Woman's Fertility dataset

There are different CSV file for different phenomenon associated with woman fertility dataset. This dataset is collected by WHO with the help of 100 volunteers a semen sample analyzed according to WHO 2010 criteria. Sperm concentration are related to socio-demographic data, environmental factors, health status, and life habits. Dataset characteristics: Multivariate, Attribute characteristics: Real, Associated task: Classification, Regression, Number of instances: 100, Number of attribute: 10, Missing values: NA, Area: Life, Date Donated: 2013-01-17 and number of web hits: 142847. The further description is specifically regarding the "Blood Transfusion Service Center". This include field like Recency (Months), Frequency (Times), Monetary (c.c. blood), Time (months) and donated. Donated is binary attributed stating in terms of 1 and 0 to represent whether donated or not.

Methodology

The Mathematical Model for the explaining the concept of the hyperplane i.e. positive and negative hyperplane and the decision boundaries is discussed in detail in following section:

$$w_0 + \boldsymbol{w}^T \boldsymbol{x}_{pos} = 1 \qquad (1)$$

$$w_0 + \boldsymbol{w}^T \boldsymbol{x}_{neg} = -1 \qquad (2)$$

If we subtract those two linear equations (**1**) and (**2**) from each other, we get:

$$\Rightarrow \boldsymbol{w}^T \left(\boldsymbol{x}_{pos} - \boldsymbol{x}_{neg} \right) = 2$$

We can normalize this by the length of the vector *w*, which is defined as follows:

$$\|\boldsymbol{w}\| = \sqrt{\sum_{j=1}^{m} w_j^2}$$

So we arrive at the following equation:

$$\frac{\boldsymbol{w}^T \left(\boldsymbol{x}_{pos} - \boldsymbol{x}_{neg} \right)}{\|\boldsymbol{w}\|} = \frac{2}{\|\boldsymbol{w}\|}$$

The left side of the preceding equation can then be interpreted as the distance between the positive and negative hyperplane, which is the so-called margin that we want to maximize.

Now the objective function of the SVM becomes the maximization of this margin by maximizing 2/||w|| under the constraint that the sample are classified correctly,

$$w_0 + \boldsymbol{w}^T \boldsymbol{x}^{(i)} \geq 1 \ \mathit{if} \ y^{(i)} = 1$$

$$w_0 + \boldsymbol{w}^T \boldsymbol{x}^{(i)} < -1 \ \mathit{if} \ y^{(i)} = -1$$

For non-linearly separable data, Slack variable is introduced. The motivation for introducing the slack variable was that the linear constraints need to be relaxed for nonlinearly separable data to allow convergence of the optimization in the presence of misclassifications under the appropriate cost penalization.

$$\boldsymbol{w}^T \boldsymbol{x}^{(i)} \geq 1 \ \mathit{if} \ \ y^{(i)} = 1 - \xi^{(i)}$$

$$\boldsymbol{w}^T \boldsymbol{x}^{(i)} < -1 \ \mathit{if} \ \ y^{(i)} = 1 + \xi^{(i)}$$

So the new objective to be minimized (subject to the preceding constraints) becomes:

$$\frac{1}{2}\|\boldsymbol{w}\|^2 + C\left(\sum_i \xi^{(i)} \right)$$

Using the variable C, we can then control the penalty for misclassification. Large values of C correspond to large error penalties whereas we are less strict about misclassification errors if we choose smaller values for C. We can then we use the parameter C to control the width of the margin and therefore tune the bias-variance trade-off as illustrated in the following figure:

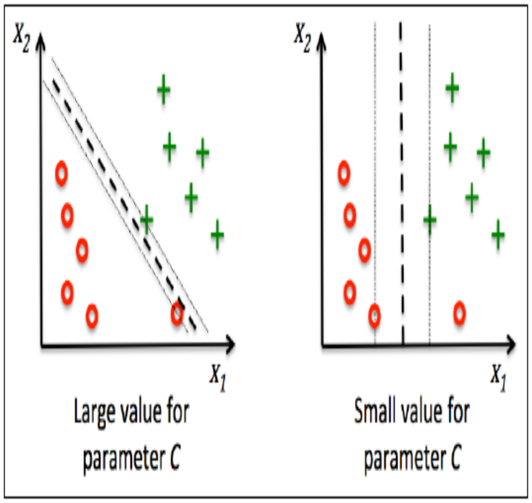

Fig 2: Penalty for non-linear classification

The another form of SVM is Kernel SVM. It is highly popular and can easily be kernelized. It is capable of drawing the non-linear decision boundary. For example XOR data can separated using the kernel SVM.

Experiment

To implement SVM we need to import the scikit learn package use the SVC Class. We can use the parameter kernel='linear' for linear classification while the kernel='rbf' for kernel SVM.

Conclusion

The classification using Support vector machine is reliable and used for linear and non-linear separable dataset. This performance of SVM is better than other classification model. It also include the Kernel SVM for Non-linearly separable data point. It is free of the dimension of the size of the data. So , no curse of dimensionality. Due to Kernel SVM, we need not to map the nonlinearly separable data point in X-Plane to high dimensional space in Z-Plane.

Reference

[1] Sebastian Raschka, "Python Machine Learning "