Necessity of Automatic Kernel Selection in Machine Learning through Support Vector Machine

Gurdev Singh[1], Dr Anupam Bhatia[2]
[1]PhD Scholar, [2]Assistant Professor
CRS University, Jind, Haryana, INDIA

Abstract: Support Vector Machine (SVM) is widely used for classification and prediction. SVM maps non-linear hyperplanes to linear space. Kernel tricks are used to map high dimensional feature space to linear plane. Any mistake in selection of Kernel can severely affect support vector count, accuracy and many other parameters. Hence, Kernel Selection is an important task to achieve high accuracy in classification through Support Vector Machines. There is urgent need to develop mechanism to select the kernel automatically.

Keywords: SVM, Kernel Selection, Linear Kernel, Polynomial Kernel, Radial Basis Kernel, Sigmoid Kernel, High Dimensional Feature Space.

Introduction

SVM is based on supervised learning which classifies points to one of two disjoint half-spaces. It uses nonlinear mapping to convert the original data into higher dimension. Its objective is to construct a function which will correctly predict the class to which the new point belongs and the old points belong[1].

With an appropriate nonlinear mapping, two data sets can always be divided by hyperplane. Hyperplane separates the tuples of one class from another and defines decision boundary. There are many hyper planes that separate the data but only one will achieve maximum separation. The main reason behind maximum margin or separation because if we use a decision boundary to classify, it may end up nearer to one set of datasets compared to others. This was the case when data is linear but mostly we find data is non-linear and data set is inseparable then we use kernels[2].

Neural Networks are easier to apply than Support Vector Machine but sometimes it provides insufficient results. Even the perceptron learning algorithms are slower than SVM learning. Traditional Classification approaches perform weakly when working directly because of high dimensionality of data but Support Vector Machine can avoid the pitfalls of very high dimensionality representations. Support Vector Machine is the most promising technique and approach as compared to others. Limitations of SVM are speed, size both in training and testing. Discrete data presents another problem. Most severe difficulty with SVMs is the high algorithmic complexity and extensive memory requirements.

Support Vector Machine is used for many applications such as text categorization, pattern recognition, face recognition, handwriting analysis but especially for classification and regression applications. In short we can say that the development of SVM is an utterly different from standard algorithms used for learning and SVM provides a fresh insight into this learning. Support vector have proven track record in binary classification[2].

Selection of kernels for particular Data Set is a complicated and tricky choice for Data Miner Analyst as because Support Vector Machine is a kernel-sensitive in nature. Choosing the most appropriate kernel highly depends on the problem at hand because it depends on what we are trying to model and fine tuning its parameters can easily become a tedious and burdensome task. Automatic kernel selection is possible and has been discussed in some research work.

Kernel Methods

Kernel Methods are a class of Algorithms for Pattern Analysis and Classification whose optimum well-known element is the Support Vector Machine (SVM). The main characteristic of Kernel Methods, However, is their distinct approach to this problem. Kernel methods map the data into higher dimensional spaces in the hope that in this higher-dimensional space the data could become more easily separated or better arranged. There are also no restraints on the form of this mapping, which could even lead to infinite-dimensional spaces. This mapping function, however, hardly needs to be computed because of a tool called the kernel trick.

The kernel trick is a mathematical tool which can be applied to any algorithm which exclusively depends on the dot product between two vectors. Wherever a dot product is used, it is replaced by a kernel function. When properly applied, those candidate linear algorithms are transformed into non-linear algorithms. Those non-linear algorithms are equivalent to their linear originals operating in the range space of a feature space φ.

Here, we will explain about Kernel Specification. Slightly Change of value affect heavily on accuracy. There are 4 Types of kernels explained below:

- **Linear**

  The Linear Kernel is the simplest kernel function. It is given by the inner product <x, y> plus optional constant ‘c’. Kernel algorithms using a linear kernel are often equivalent to their non-kernel counterparts.

  $$k(x, y) = x^T y + c$$

- **Polynomial**

  The Polynomial kernel is a non-stationary kernel. Polynomial kernels are well suited for problems where all the training data is normalized. For degree-*d* polynomials, the polynomial kernel is defined as

  $$k(x, y) = (\alpha x^T y + c)^d$$

Where x and y are vectors in the input space, i.e. vectors of features computed from training or test samples, $c \geq 0$ is a constant trading off the influence of higher-order versus lower-order terms in the polynomial and when $c = 0$, the kernel is called homogeneous.

- **Radial Basis Function**

The RBF is by far the most popular choice of kernel types used in Support Vector Machines. This is mainly because of their localized and finite responses across the entire range of the real x-axis.

There are different types of Radial Basis Kernel Function such as Linear Radial Basis Function, Gaussian Radial Basis Function and Multiquadrics Radial Basis Function.

- **Sigmoid**

The Sigmoid Kernel comes from the Neural Networks field. It is interesting to note that a SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network. This kernel was quite popular for Support Vector Machines due to its origin from neural network theory. Also, despite being only conditionally positive definite, it has been found to perform well in practice.

$$k(x, y) = \tanh(\alpha x^T y + c)$$

There are two adjustable parameters in the sigmoid kernel, the slope **alpha** and the intercept constant **c**.

Existing Work

Classification through Support Vector Machine has shown a significant difference in various parameters. Number of Support Vectors ranges from 439 to 535 for different kernels. Number of Kernel Evaluations ranges from 19131 to 282572. Root Mean Squared Error ranges from 0.476 to 0.5907. Accuracy also had shown significant difference from 65.1042 to 77.3438. This significant difference clearly indicates that Kernel Selection is an important task while dealing with Support Vector Machine. Wrong Selection of Kernel will not only reduce the accuracy but also affect number of support vectors, kernel evaluation and root mean squared error[3].

A classifier is a function that mimics the relationship between the data vectors and their class labels. Support vector machines (SVMs) are a popular classifier based on the principle of structural risk minimization (SRM), which selects the optimal classifier based on a trade-off between empirical risk minimization and margin maximization. They proposed a new approach to multiclass SVM classification by introducing the concept of target space into the problem formulation. They define the trained machine as a map from input (data) space to a dT-dimensional target space divided into non-overlapping proper conic regions termed class regions, each of which is associated with a single class.[4]

Kernel Methods have great importance in Machine Learning process. Kernel methods are efficient to map original training points into multidimensional feature space. Linear Kernel functions focus on neighborhood training points and do not consider far away training points. Polynomial Kernel Functions give equal weightage to all training points across the dataset.

Various Kernels are available for Support Vector Machine. There is need to develop mechanism which can guide the Kernel Selection Process. [5]

It is an established fact that Support Vector Machine offers excellent performance because of its method to apply linear algorithm on high dimensional data space. There is big problem related to choice of parameters used by SVM and its Kernels. The performance is dependent on both parameters as well as Kernel Selection, as these facilitate SVM to achieve optimal separating hyperplane in high dimensional feature space.[6]

Conclusion

One of the major challenge is that of choosing a suitable kernel for given application. The motivation behind the choice of a particular parameter can be very intuitive and straightforward depending on what kind of information we are expecting to extract about the data. A radial basis function allows you to pick out circles (or hyperspheres) - in contrast with the linear kernel, which allows only you to pick out lines (or hyperplane).In comparison, a polynomial kernel, for example, allows us to model feature conjunctions up to the order of the polynomial.
Hence, It is observed in this paper that Support Vector Machine is kernel-type sensitive and Hence, Data Miner Analyst must ensure the choice of correct kernel parameter for particular data set.

References

[1] S.B. Kotsiantis (2007), “Supervised Machine Learning: A Review of Classification Techniques”. *Informatica 31 pp 249-268.*

[2] Laura Auria, Rouslan A. Moro (2008), “Support Vector Machines (SVM) as a Technique for SolvencyAnalysis”.*http://www.diw.de/english/products/publications/discussion_papers/27539.html*

[3] Baitharu, Pani, Dhal (2015); Comparison of Kernel Selection for Support Vector Machine using Diabetic Data Set; *Journal of Computer Science and Applications, Vol. 3, No. 6.*

[4] Shilton Alistair and Daniel T. H. (2012).The Conic-Segmentation Support Vector Machine -a Target Space Method for Multiclass Classification. *WCCI 2012 IEEE World Congress on Computational Intelligence.*

[5] Valerio, VIlalta (2014); A Data Complexity Approach to Kernel Selection for Support Vector Machine, *Proceedings of Twenty Eighth AAAI Conference on Artificial Intelligence.*

[6] Amami, Eyed, Ellouze (2013); Practical Selection of SVM Supervised parameters with Different Feature Representation for Vowel Recognition; *International Journal of Digital Content Technology and Its Applications, Vol. 7, No. 9.*