Volume 7, Issue 5, May 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Kernel's Impact on SVM Classification

Esha Bansal
Department of Computer Science
Dravidian University, Kuppam, India

Anupam Bhatia
Department of Computer Science & Applications
Chaudhary Ranbir Singh University, Jind, India

DOI: 10.23956/ijarcsse/SV7I5/0238

Abstract*— Support Vector Machine (SVM) has proven track record in Classification. Higher level of accuracy and reasonably good speed factor is attracting the analysts towards SVM classification. The core of success of SVM accuracy is its Kernel. The right choice of kernel can lead to highest level of accuracy whereas if the choice is not good then the accuracy may be too low to be considerable. In this paper, methods to choose the right kernel are listed with comparative analysis of multiple kernels on same data set. Multiple data sets are implemented on various kernels to establish the impact of kernel on SVM classification.

***Keywords*— Support Vector Machine; Kernel Functions; Feature Space.**

I. INTRODUCTION

The Choice of proper kernel method is the key to the success of SVM. There are a number of admissible kernel functions that can be applied but basic study reveals that these methods make use of information about the inner products between vectors (inputs) in some feature space which is indeed very complex.

An "ideal" kernel function assigns a higher similarity score to any pair of objects that belong to the same class than it does to any pair of objects from different classes. This is the case if the implicit mapping by the kernel function brings similar objects close together and takes dissimilar objects apart from each other in the induced feature space [3]. The kernel functions make it possible for the user to apply a classifier to statistics that have no apparent fixed-dimensional vector space representation.

Frequently used kernel functions are [11] [8] [2]:

Linear kernel: $K(X_i, X_j) = < X_i . X_j >$

Radial basis function (RBF) kernel: $K(X_i, X_j) = e^{-||Xi - Xj||x^2/2\sigma^2}$

Polynomial kernel: $K(X_i, X_j) = (X_i . X_j + 1)^h$

Sigmoid kernel: $K(X_i, X_j) = tanh(k\, X_i . X_j - \delta)$

II. SUPPORT VECTOR MACHINE

Support Vector Machine is a supervised learning tool used for data mining. It is basically a binary classification tool used to map binary data which can be linearly separable but can be extended to nonlinear mapping also by converting the data into higher dimension and then finding the Maximum Margin Hyperplane (MMH)[15]. Its goal is to formulate a function which will precisely predict the class to which the new point belongs using various kernel functions in case of multiclass classification.

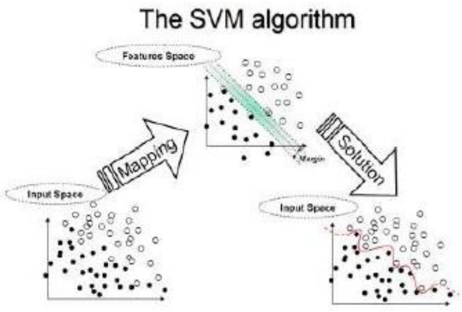

Fig. 1 An overview of SVM Process [18]

III. LINEAR KERNEL

A. Linearly Separable Data

Linear kernel is applicable on datasets which can be separated into two distinct classes by defining a hyperplane between them. The data with similar attributes falls on the same side of the hyperplane and those with distinct attributes falls on the other side based on sign of the linear discriminant function

$$f(x) = w^T x + b$$

© 2017, IJARCSSE All Rights Reserved

The vector w is known as the weight vector, such that w = {w_1, w_2, w_3,…., w_n }; n is the number of attributes and b is a scalar called bias. Any point that lies above the hyperplane satisfies the inequality $f(x) > 0$ and hence belongs to class +1 And any point that lies below the hyperplane satisfies $f(x) < 0$ and the class prediction in that case is -1.

This hyperplane chosen must be the Maximum Margin Hyperplane (MMH) that is, the distance between the nearest tuples must be maximum.

B. Linearly Inseparable Data

When data is linearly inseparable we need to map it into higher dimensional space and then have to locate the linear separating plane in this new space. Here instead of computing the dot product on the transformed tuples as its computation is complex we can indeed apply a kernel function, *K(Xi,Xj)*, to the original data

$$K(X_i, X_j) = \emptyset(X_i).\emptyset(X_j)$$

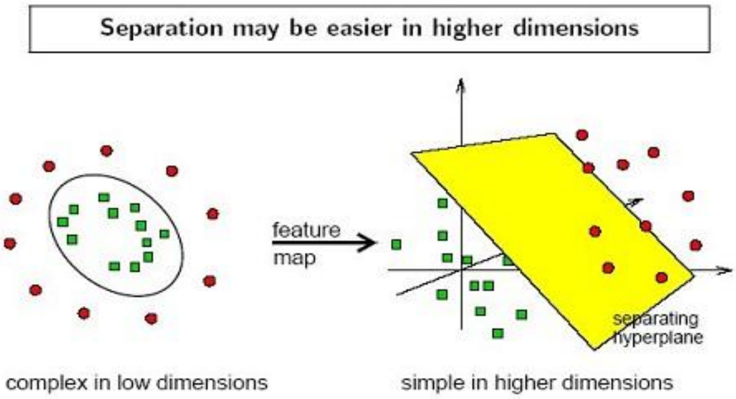

Fig. 2 Kernel Function transforming the data into higher dimensional space for separation of data [18]

The three admissible kernel functions are-

Gaussian radial basis function kernel: $K(X_i, X_j) = e^{-||Xi-Xj||x^2/2\sigma^2}$

Polynomial kernel: $K(X_i, X_j) = (X_i.X_j + 1)^h$

Sigmoid kernel: $K(X_i, X_j) = tanh(k\,X_i.X_j - \delta)$

IV. GAUSSIAN RADIAL BASIS FUNCTION (RBF) KERNEL

This kernel function is the key kernel among all other kernel functions as all other kernel functions are just the special cases of RBF kernel since the linear kernel with a penalty parameter ~ C has the same performance as the RBF kernel with some parameters (C,♈). In addition, the sigmoid kernel behaves like RBF for certain parameters [9]. RBF maps nonlinear samples into a higher dimensional space such that the nonlinear relationship between class labels and attributes transforms into linear relationship.

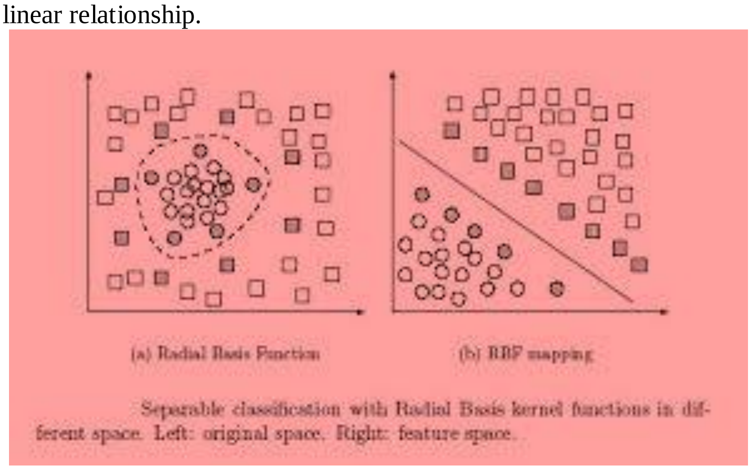

Fig. 3 Classification with RBF in feature space [18]

V. POLYNOMIAL KERNEL

For degree-*d* polynomials, the polynomial kernel is defined as

$$K(X_i, X_j) = (X_i.X_j + 1)^d$$

where X_i and X_j are vectors of features in the input space computed from training or testing sets. This method implicitly takes features combinations into account rather than combining of features explicitly. By setting different values of d in the above equation different number of feature combinations can be taken into account unlike linear kernel which do not take feature combinations into account in linear space. If d=1 then this polynomial kernel method behaves like linear kernel approach.

© 2017, IJARCSSE All Rights Reserved

The natural linear kernel simply uses the dot-product as

$$K(X_i, X_j) = < X_i . X_j >$$

While a polynomial kernel of degree *d* is given by

$$K(X_i, X_j) = (X_i . X_j + 1)^d$$

Although the efficiency of polynomial method is very high in terms of predictive analysis its limitation is that it is very costly in terms of training and testing time overheads[15]. The polynomial kernel in the normalized feature space is the best alternative to Gaussian RBF kernel.

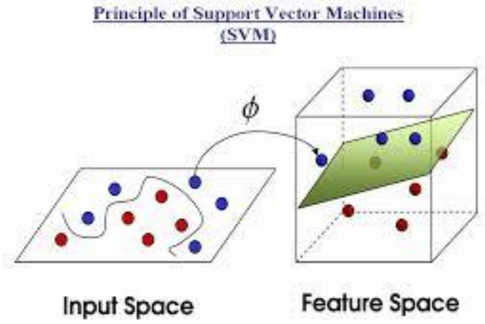

Fig.4 Mapping with Polynomial Kernel Function[16]

VI. SIGMOID KERNEL

The sigmoid kernel was relatively popular as its origin is from neural networks. The sigmoid kernel matrix is conditionally positive definite (CPD) in certain parameters but mostly it is a non-PSD (positive Semi Definite) kernel. However, it is already proven that sigmoid kernel is not better than the RBF kernel in general but still it's popular.

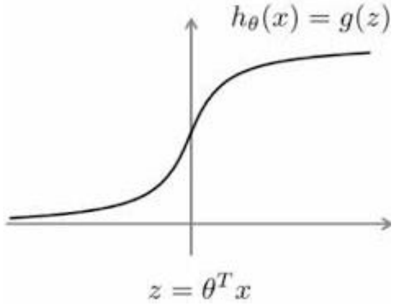

Fig. 5 Mapping with Sigmoid Kernel Function [17]

We consider the sigmoid kernel $K(X_i, X_j) = tanh(k\, X_i . X_j - \delta)$, which takes two parameters: *k* and δ.

For k > 0, we can view k as a scaling parameter of the input data, and δ as a shifting parameter that controls the threshold of mapping. [4]

For k < 0, the dot-product of the key in data is scaled as well as reversed. It concludes that the case, k > 0 and δ < 0, is more appropriate for the sigmoid kernel and is close to RBF kernel.

VII. DATA ANALYSIS

We have used different kernel functions on the same data set keeping all other parameters like degree, cost, cachesize, weight etc constant. Comparative analysis of multiple kernels on same data set is done and also multiple data sets (Handwritten Digit Data Set and Letter Recognition Data Set) are implemented on various kernels to establish the impact of kernel on SVM classification.

A. Data Set

The data set Handwritten Digit Data Set [20] and Letter Recognition Data Set [19] has been downloaded from UCI data repository.

Table-1. Accuracy Table for two data sets

Datasets	Linear	Polynomial	Radial	Sigmoid
Handwritten Digit Data Set	87.28	97.68	11.68	11.2
Letter Recognition Data Set	83.79	33.65	80.75	76.51

VIII. CONCLUSION

From the above Table-1 it is concluded that -

In case 1 –Handwritten Digit Data Set accuracy varies from 11.2% to 97.68%

In case 2- Letter Recognition Data Set accuracy level varies from 33.65% to 83.79%

So choice of right kernel to be applied on a dataset is the foremost priority for higher accuracy in results.

REFERENCES

[1] Asa Ben-Hur et al. "A User's Guide to Support Vector Machines",Methods in Molecular Biology, Vol . 609 (2010), pp.223-239
[2] B. Scholkopf and A.J. Smola (2002) " Learning with Kernels." MIT Press, Cambridge, MA
[3] D.Huson(2007) "Algorithms in Bioinformatics II " SoSe'07, ZBIT.
[4] Hsuan-Tien Lin and Chih-Jen Lin (2003), "A Study on Sigmoid Kernels for SVM and the Training of non-PSD Kernels by SMO-type Methods" Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Technical Report.
[5] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques" book
[6] John Shawe-Taylor(2009) "Kernel Methods and suppport Vector Machines",Chicago/TTI Summer School
[7] Koby Crammer,Yoram Singer(2001) "On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines",Journal of Machine Learning Research 2,pp. 265-292
[8] K.-R .Muller et al. (2001)" An introduction to kernel-based learning algorithms", IEEE Transactions on Neural Networks, 12(2):181–201
[9] Lin, K.-M. and C.-J. Lin (2003). A study on reduced support vector machines. IEEE Transactions on Neural Networks.
[10] Martin Hofmann(2006) " Support Vector Machines — Kernels and the Kernel Trick ",An elaboration for the Hauptseminar "Reading Club: Support Vector Machines"
[11] V. Vapnik (1995) "The nature of statistical learning theory." Springer Verlag, New York.
[12] V.Vapnik(1998)"Statistical Learning Theory,John Wiley & Sons"
[13] Yashima Ahuja ,Sumit Kumar Yadav (2012) "Multiclass Classification and Support Vector Machine"Global Journal of Computer Science and Technology,Vol 12 issue 11
[14] Yu Chi Wu (2007) " An Approximate Approach for Training Polynomial Kernel SVMs in Linear Time " ,Proceedings of Association for Computational Linguistic (ACL) Prague, pp. 65-68
[15] E.Bansal, A. Bhatia, (2013). Support Vector Machine For Multiclass Handwritten Digits" . *International Conference On Advanced Information Communication Technology In* Engineering (Icaicte- 2k13), (Pp. 239-241).
[16] imtech.res.in/raghava/rbpred/svm.jpg
[17] holehouse.org/mlclass/12_Support_Vector_Machines.html
[18] dtreg.com/svm.htm
[19] Letter Recognition: http://archive.ics.uci.edu/ml/datasets/Letter+Recognition
[20] Hand Written Digit Dataset : http://archive.ics.uci.edu/ml/datasets/Semeion+Handwritten+Digit.

© 2017, IJARCSSE All Rights Reserved