International Journal of Latest Research in Science and Technology **ISSN (Online):2278-5299**
Volume 4, Issue 6: Page No.37-39, November-December 2015 (special Issue Paper)
http://www.mnkjournals.com/ijlrst.htm
Special Issue on International conference on Methematics Sciences-2015(MSC-2015)
Confernece Held at Sadguru Gadage Maharaj College, Karad ,Maharastra,(India)

CLASSIFICATION OF TEACHING EVALUATION PERFORMANCE USING SUPPORT VECTOR MACHINE.

[1]Poonam Mhetre, [2]M. S. Bapat

[1]GIT,Ratnagiri,4115708,Maharashtra, India Email: mhetrepoonam@gmail.com
[2]Willingdon College, Sangli 416415, Maharashtra, India Email: msbapt@gmail.com

Abstract-***The support vector machine (SVM) is a method of classification introduced by Vapnik in 1992. The SVM is used in various fields due to its accuracy and ability to deal with high dimensional data. In SVM determination of parameter C is critical issue. In this paper the SVM method is applied to the multiclassification of teaching evaluation performance data using Matlab programming. Value of the parameter is determined for minimum error.***

Keywords: Support Vector Machine, training set, Confusion Matrix

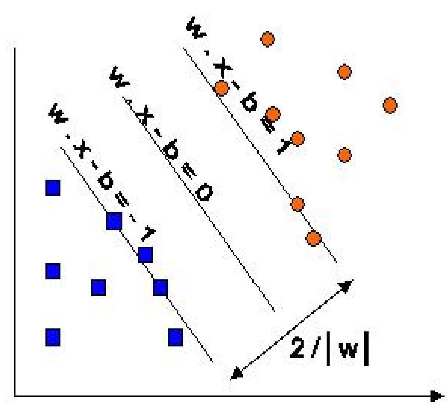
I. INTRODUCTION

Support Vector Machine (SVM) is a supervised learning method used for classification and regression. They belong to a family of generalized linear classification. A special property of SVM is that it simultaneously minimizes the empirical classification error and maximize the geometric margin. Therefore SVM is also called as Maximum Margin Classifiers. SVM is based on the Structural risk Minimization (SRM). SVM maps input vector to a higher dimensional space where a maximal separating hyper plane is constructed. Two parallel hyper planes are constructed on each side of the hyper plane that separate the data. The separating hyper plane is the hyper plane that maximizes the distance between the two parallel hyper planes. It is assumed that the larger the margin or distance between these parallel hyper planes the better the minimization of the error of the classifier will be.

We consider data points of the form
$\{(x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4)\ldots\ldots\ldots,(x_n, y_n)\}$
where y_n=1, -1 , a constant, denoting the class to which that point x_n belongs. n is the number of samples for training set. Each x_n is p-dimensional real vector. The scaling is important to guard against variable (attributes) with larger variance. We can view this training data , by means of the dividing (or separating) hyper plane of the type, $\mathbf{w} \cdot \mathbf{x} + b = 0$ (1)

where b is scalar and **w** is p-dimensional vector. The vector **w** points perpendicular to the separating hyper plane . Adding the offset parameter b allows us to increase the margin. Absent of b forces the hyper plane is to pass through the origin and restricting the solution. As we are interesting in the maximum margin , we are interested in support vectors and the parallel hyper planes. Parallel hyper planes can be described by equation,$\mathbf{w.x} + b = 1$, $\mathbf{w.x} + b = -1$

If the training data are linearly separable, we can select these hyper planes so that there are no points between them and then try to maximize their distance. By geometry, we find the distance between the hyper plane is $\frac{2}{|w|}$. So we want to minimize $|w|$. To excite data points, we need to ensure that for all i either $\mathbf{w.\ x_i} - b \geq 1$ or $\mathbf{w.\ x_i} - b \leq -1$. This can be written as$y_i$ ($\mathbf{w.\ x_i} - b) \geq 1$, $1 \leq i \leq n$ (2)



Samples along the hyper planes are called Support Vectors (SVs). A separating hyper plane with the largest margin M = 2 / $|w|$ that specifies support vectors, means training data points which satisfy $y_j [\mathbf{w}^T \cdot \mathbf{x}_j + b] = 1$, j=1,2,…k (3) Optimal Canonical Hyper plane (OCH) is a hyper plane having a maximum margin. For all the data points OCH should satisfy $y_i [\mathbf{w}^T \cdot \mathbf{x}_i + b] = 1$, i =1,2…n(4)

where n is number of training data point. In order to find the optimal separating hyper plane having a maximal margin, a learning machine should minimize $\|w\|/2$ subject to the inequality constraints

$y_i [\mathbf{w}^T \cdot \mathbf{x_i} + b] \geq 1$; i =1,2…….n (5)

This optimization problem solved by the saddle points of the Lagrange's Function

$LP = L(w, b, \alpha) = 1/2\|w\|^2 - \Sigma\ \alpha_i\ (y_i\ (\mathbf{w}^T\ .\mathbf{x_i} + b\ ) -1)$
$i=1,2\ldots = 1/2\ \mathbf{w}^T .\mathbf{w} - \Sigma\ \alpha_i\ (y_i(\mathbf{w}^T .\mathbf{x_i} + b\ ) - 1)$
$i=1,2,\ldots(6)$

where α_i is a Lagrange's multiplier .The search for an optimal saddle points (w_0, b_0, α_0) is necessary because Lagrange's must be minimized with respect to **w** and b and

has to be maximized with respect to nonnegative α_i ($\alpha_i \geq 0$). This problem can be solved either in primal form (which is the form of **w** & b) or in a dual form (which is the form of αi).Equations (4) and (6) are convex and are KKT conditions, which are necessary and sufficient conditions for a maximization of (4). Partially differentiating (6) with respect to saddle points (w_0,b_0, α_0).

$$\partial L / \partial w_0 = 0, w_0 = \Sigma\ \alpha_i\ y_i\ x_i (7)$$

And $$\partial L / \partial b_0 = 0\ ,\quad \Sigma\ \alpha_i\ y_i = 0 \quad (8)$$

Substituting equation (7) and (8) in equation (6). We change the primal form into dual form.

$$Ld\ (\alpha) = \Sigma\ \alpha_i - 1/2\ \Sigma\ \alpha_i\ \alpha_j\ y_i\ y_j \quad \mathbf{x_i}^T\ .\ \mathbf{x_j} \quad (9)$$

In order to find the optimal hyper plane, a dual lagrangian (Ld) has to be maximized with respect to nonnegative α_i (i .e. α_i must be in the nonnegative quadrant) and with respect to the equality constraints as follow

$\alpha_i \geq 0$, $i = 1,2\ldots\ldots n$, $\Sigma\ \alpha_i\ y_i = 0$

1. Computational support - Libsvm

LIBSVM is an integrated software for support vector classification, (C-SVC, nu-SVC), regression (epsilon-SVR, nu-SVR) and distribution estimation (one-class SVM). It supports multi-class classification. LIBSVM was a library for support vector machines (SVM). Its goal was to make users use SVM as a tool easily.

2. Selection of Kernel function:

In SVM there are four types of the kernel such as Linear kerenel, Sigmoid kernel, Polynomial kernel, RBF kernel. But generally RBF is first choice among these kernels. Because whenever the relation between class labels and attributes is nonlinear ,the RBF kernel has fewer numerical difficulties. The linear kernel cannot handle multi-classification problems.

3. Example

The data consist of evaluations of teaching performance over three regular semesters and two summer semesters of 151 teaching assistants (TA) assignments at the *Statistics Department of the University of Wisconsin-Madison*. The scores were divided into 3 roughly equal-sized categories ("low", "medium", and "high"). In the data set there are totally 6 attributes such as

(1) Whether of not the TA is a native English speaker (binary),1=English speaker, 2=non-English speaker
(2) Course instructor (categorical, 25 categories)
(3) Course (categorical, 26 categories)
(4) Summer or regular semester (binary) 1=Summer, 2=Regular
(5) Class size (numerical)
(6) Class attribute (categorical) 1=Low, 2=Medium, 3=High.

4. Error Calculations

The confusion matrix is a 2X2 matrix $\begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix}$, where TP*(true positive)*- the proportion of positive cases that were correctly classified, TN*(true negative)* is the proportion of negative cases that were correctly classified ,FP*(false positive)* the proportion of negatives cases that were incorrectly classified as positive , FN*(false negative)* is the proportion of positives cases that were incorrectly classified

as negative .And error is given by , Error = (FP+FN)/(TP+FP+TN+FN).

5. Algorithm

To find error for different value of cost parameter C we used libsvm software to train data & classified it by selecting RBF kernel.

We used the following algorithm which may be implemented in Mat lab

1. Load teaching assistant evaluation data
2. Select maximum number of labels
3. Select no. of training data
4. No. of testing= no. of instant- no. of training data
5. Train data & label
6. Test data & label
7. Train data as model = svmtrain(training_label_vector, training_instance_matrix, 'libsvm_options');
8. Classify data [predict_label, accuracy, prob_estimates] = svmpredict(training_label_vector, training_instance_matrix,model);
9. Find error using confusion matrix.

The following table shows error corresponding to different values of C

Data points	Value of C	Error
1	0.01	34.4371
2	0.02	35.7616
3	0.03	42.38
4	0.04	37.0861
5	0.05	40.3974
6	0.06	37.0861
7	0.07	36.4238
8	0.08	41.7219
9	0.09	36.4238
10	0.1	35.7616
11	0.2	36.4238
12	0.3	35.7616
13	0.4	37.0861
14	0.5	40.3974
15	0.6	32.4503
16	0.7	33.1126
17	0.8	35.7616
18	0.9	35.7616

6. CONCLUSION:

In this paper, we have find error corresponding to different cost value using radial kernel. It can be seen that after some interval error value is repeated. Therefore the value of the constant C is to be determined properly to have minimum error in the classification.

REFERENCES

1. Chang, C.-C. and C. J. Lin (2001). LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm .
2. Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. "A Practical Guide to Support Vector Classification" Deptt of Computer Sci.National Taiwan Uni, Taipei, 106, Taiwan http://www.csie.ntu.edu.tw/~cjlin 2007.
3. Durgesh k. Srivastava and Lekha Bhambhu " Data Classification Using Support Vector Machine", Journal of Theoretical and Applied Information Technology © 2005 – 2009
4. L. Hamel, "Knowledge discovery with Support Vector Machines", John Wiley and Sons,2009
5. M.S.Bapat and M.N.Jawalgikar, " Determination Of Optimal Value Of Parameter In Fuzzy Support Vector Machine",Avishkar Journal, Solapur University ,Solapur,2014.
6. V. Vapnik."The Nature of Statistical Learning Theory". NY: Springer-Verlag. 1995