Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 13 Issue 9 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

# SVM Classification in Multiclass Letter Recognition System

By Aarti Kaushik, Gurdev Singh & Anupam Bhatia
*Suresh Gyan Vihar University, India*

*Abstract -* In this paper, we have studied the concept of multiclass classification and support vector machine. After scrutinizing the problem of multiclass classification, we concluded the fact for how it is suited to describe the binary classification. Also illustrating various kernel functions associated with implementation for svm. By further proceeding, it would let you disclose the concept for fitness of support vector machine in multiclass letter recognition system.

*GJCST-C Classification : I.5.0*

SVM CLASSIFICATION IN MULTICLASS LETTER RECOGNITION SYSTEM

*Strictly as per the compliance and regulations of:*

# SVM Classification in Multiclass Letter Recognition System

Aarti Kaushik [α], Gurdev Singh [σ] & Anupam Bhatia [ρ]

*Abstract -* In this paper, we have studied the concept of multiclass classification and support vector machine. After scrutinizing the problem of multiclass classification, we concluded the fact for how it is suited to describe the binary classification. Also illustrating various kernel functions associated with implementation for svm. By further proceeding, it would let you disclose the concept for fitness of support vector machine in multiclass letter recognition system.

## I. Introduction

Each training point belongs to one of N different classes in multiclass classification. The classes are mutually exclusive. The goal is to construct a function which, given a new data point, will correctly predict the class to which the new point belongs. In this paper, we have to scrutinize the problem of multiclass classification and how can we extend the support vector machine to allow for multiclass classification is still a research issue. Support vector machine is inspired from statistical learning theory. The main feature of SVM is that accuracy is high owing to their ability to model complex nonlinear decision boundaries (margin maximization) but training can be slow. The main disadvantage of SVM is to choose a "good kernel function".

There are various methods of multiclass SVM classification:

- 1-vs-all
- 1-vs-1
  - DB2
  - Error Correcting Output Coding
- K-class SVM

## II. Support Vector Machine

A support vector machine is a design of classification of both linear and nonlinear data. It converts the original data into a higher dimension; from where it may SVM is based on supervised learning which SVM becomes popular because of its success in handwritten digit recognition, object recognition and speech recognition. 1.1% test error rate for SVM. This is the same as the error rates of a constructed neural network. Support Vector Machine is primarily designed for binary classification. The various properties of SVM are duality, robust, kernel, margin, convexity and sparseness. SVM is an important example of kernel methods.

SVM describe two cases:

The case when data are linear separable.
The case when data are linear inseparable.

*The case when data are linear separable*

Let training vectors: $\mathbf{z_k}$ , k=1…….i,
Consider a simpler case having two classes:
Define a vectory

$$Yi= \begin{Bmatrix} 1 \text{ if } z_k \text{ in class } 1 \\ -1 \text{ if } z_k \text{ in class } 2 \end{Bmatrix}$$

The idea of using a hyperplane to separate the data into two group sounds well when there are only two target categories .There are infinite numbers of decision boundary or separating hyperplane that separate the data. We have to choose the best one that has minimum classification error.

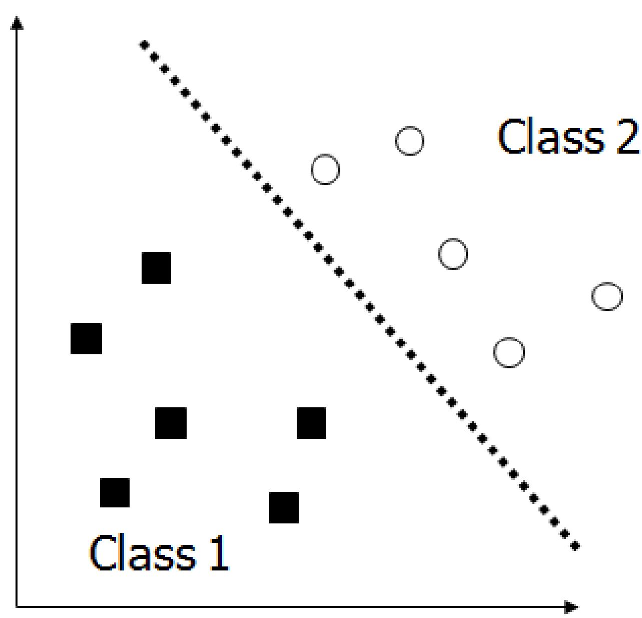


*Figure 1 :* Example of superior decision boundary separating two classes

*Author α : Research Scholar, Jind Institute of Engineering and Technology, Jind- Haryana, India. E-mail : kaushik.aarti@gmail.com*
*Author σ : Head, Department of Computer Science & Engineering, Jind Institute of Engineering and Technology, Jind.*
*Author ρ : Assistant Professor, Kuruskhetra University Post Graduate Regional Centre, Jind.*

Global Journal of Computer Science and Technology ( C ) Volume XIII Issue IX Version I Year 2013

The main purpose we need it is because if we use a decision boundary to classify, it may end up nearer to one set of datasets compared to others and we do not want this to happen and thus concept of **maximum margin classifier** or **hyper plane** as an apparent solution. Support vectors are the data points that lie closest to the decision surface.

Maximum margin hyperplane

- The decision boundary should be as far away from the data of classes as possible.

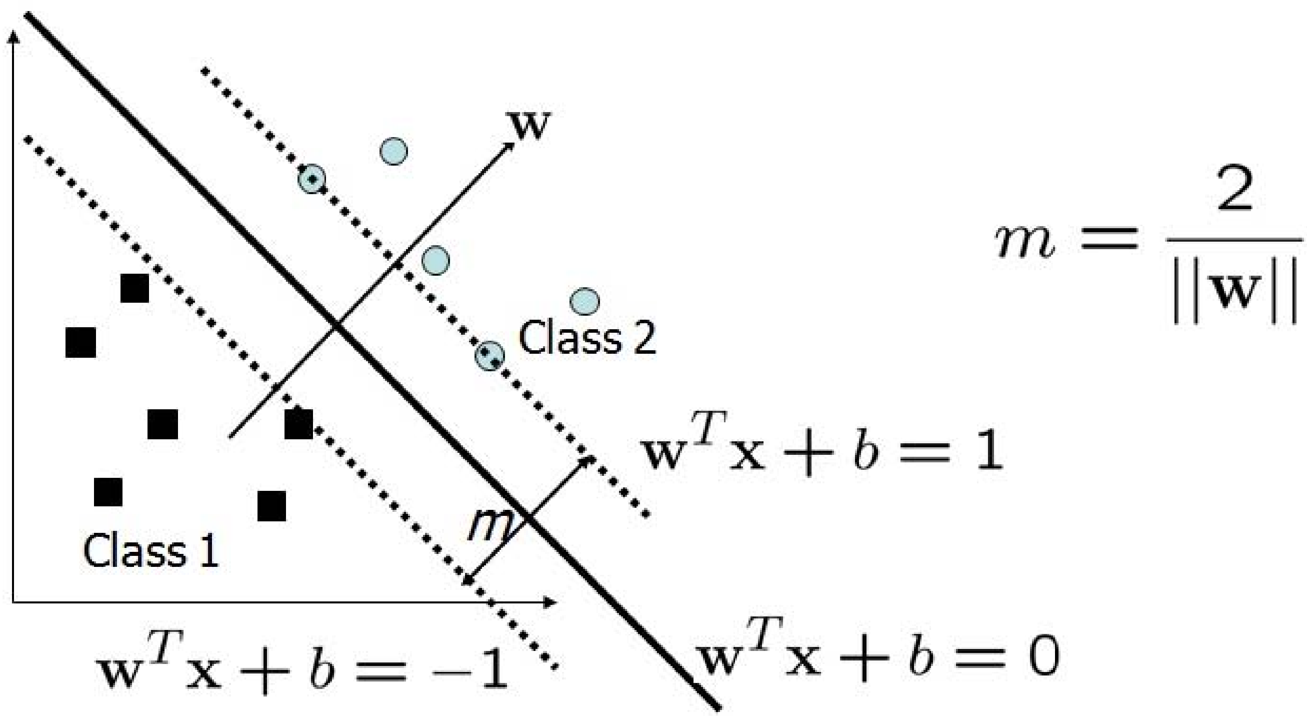


*Figure 2 :* Example of Maximum Margin Hyperplane

There are few steps for finding the decision boundary:

1. Let {$x_1$, ..., $x_n$} be our data set and let $y_i$ Î {1,-1} be the class label of $x_i$. The decision boundary should classify all points correctly Þ

$$w^T x_i + b \geq 1, \qquad ¥i$$

2. The decision boundary may be obtained by solving the following constrained optimization problem

Minimize $$\frac{1}{2} || w ||\text{^}2$$

Subject to $$y_i(w^T x_i + b \geq 1)$$

The Lagrangian formulation of this optimization problem is

$$\mathbf{L} = \frac{1}{2} || w ||\text{^}2 -$$

$$\sum \alpha_i y_i (w^T x_i + b - 1)$$

$$\alpha_i \geq 0$$

Where $y_i$ is the class label of support vector, $a_i$ and b are numeric parameters that were obtained automatically by the SVM algorithm and optimization. Langrangian formulation is very useful for finding the maximal margin hyperplane and support vectors when the given data are non linear separable. Because it contain the dot product of support vector $x_i$ and test tuple $w^T$.

In second case data are not linearly.

In second case data are **separable** i.e. in such cases no such straight line can be found that would divide the classes. Linear SVM can be extended to generate Non Linear SVM'S for classification of linearly inseparable data. Such SVM are capable of finding nonlinear decision boundaries kernel functions.

In practical use of SVM, only the kernel function (and not f (.)) is specified. Kernel function can be thought of as a similarity measure between the input objects. Not all similarity measure can be used as kernel function, however Mercer's condition states that any positive semi-definite kernel $K(x, y)$, i.e. can be expressed as a dot product in a high dimensional space.

Global Journal of Computer Science and Technology ( C ) Volume XIII Issue IX Version I Year 2013

$$\sum_{j}^{i}(x_{i,}x_{y})c_i c_j \geq 0$$

*a) Types of Kernel Functions*

- Polynomial kernel with degree *d is* defined as

$$K(X_{i,}X_{j}\ )=(X_i\ .X_j+1)\text{^}h$$

- Radial basis function kernel is defined as

$$K(X_{i,}X_{j}\ )=e^{-\|X_{i-}\ X_j\|^2/2\sigma\text{^}2}$$

  – Closely related to radial basis function neural networks.
- Sigmoid with parameter k and δ

$$K(X_{i,}X_{j}\ )=tanh(kX_i\ .X_j-\delta)$$

  – It does not satisfy the Mercer condition on all k and q.

*b) Steps for Classification*

1. Prepare the pattern matrix
2. Select the kernel function to use
3. Steps for classification
4. Prepare the pattern matrix
5. Select the kernel function to use
6. Select the parameter of the kernel function and the value of C
7. You can use the values suggested by the SVM software, or you can set apart a validation set to determine the values of the parameter
8. Execute the training algorithm and obtain the $a_i$
9. Unseen data can be classified using the $a_i$ and the support vectors.

## III. Implementation

In this section we present experimental result on several problems from the UCI repository of machine learning databases. We choose the Letter recognition dataset from the UCI Repository. The data set downloaded from UCI Repository was in .txt format and for Matlab; we transformed it in .xlsx format. Note that we scale all training data to be in [1, 1]. Then test data are adjusted to [-1, 1] accordingly. There are 26 classes, 20 fonts and having 17 columns.

By Merging LIBSVM and Matlab, we have done Multi- class classification by using Support Vector Machine. LIBSVM maintain the following learning tasks.

1. SVC
2. SVR
3. One class SVM

Generally, LIBSVM contains two steps: first, training a data set to obtain the model and second, using the model to predict information of a testing data set. LIBSVM supports various SVM formulations such as C-support vector classification, distribution estimation and v support regression.

## IV. Result and Discussion

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. We estimate the generalized accuracy using different kernel parameters ¥ and cost parameters €. For dataset letter recognition where both training and testing tests are available, for each pair of (¥, €) the validation performance is measured by training 70% and testing the rest of the training set. Then we train the whole training set using the pair of that achieves the best validation rate and predict the test set.

The resulting accuracy is different kernel is shown below:

| S. No. | Kernel Type | Accuracy |
|---|---|---|
| 1. | Linear Kernel Function | 84.511% |
| 2. | Polynomial Kernel Function | 52.222% |
| 3. | Radial Basis Kernel Function | 84.4000% |
| 4. | Sigmoid Kernel Function | 80.089% |

*Table 5.1 :* Result of all Kernel Functions

Here, we can see that Support Vector Machine with the use of Linear kernel and radial basis kernel function provide the best classification accuracies. Thus, it is able to better represent the data classification.

## V. Conclusion

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. Support vector machines have shown their great promise in many multitudinous areas and in few cases they have surpassed other methods. There is also an increasing number of modifications of the SVM, one of the most important generalization being the use support vector methodology in regression, e.g. for a very good presentation. The important disadvantage of SVM is computational appeal because it include quadratic optimization problem. The SVMs has solved various realistic problems varying from economics to genetics.

## References Références Referencias

1. Jiawei Han and Micheline Kamber "Data mining concept and technique", 2nd Edition, Maurgan & Kauffman Publishers, 2001.
2. http://nlp.stanford.edu/IR-book/html/htmledition/multiclass-svms-1.html

3. Yashima Ahuja, Sumit Kumar Yadav, "Multiclass Classification and Support Vector Machine", Global Journal of Computer Science and Technology, Vol. 10 Issue 11, 2012.
4. Christopher J.C. Burges "A tutorial on Support Vector Machine for Pattern Recognition", Data Mining and Knowledge Discovery, Vol. 2, Kluwer Academic Publishers. 1998.
5. http://www.svms.org/tutorials/Erasto2001.pdf
6. http://research.microsoft.com/en-us/people/tyliu/atc-lecture--tyliu.ppt
7. http://www.csie.cyut.edu.tw/~shwu/PR_slid
8. http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf

Global Journals Inc. (US) Guidelines Handbook 2013

www.GlobalJournals.org