

Designing a better Support Vector Machines Classification Model for Multi-Class Category

Muthukrishnan .R¹, Udaya Prakash .N²

¹Professor, Department of Statistics, Bharathiar University, Coimbatore, India

²Research Scholar, Department of Statistics, Bharathiar University, Coimbatore, India

Abstract: Support vector machines are most widely used classification technique and are computationally efficient. Even though, it provides better classification for high-dimensional data, it suffers a lot when the data contains extremes. Such situation kernel function plays a vital role to produce reliable results. An attempt has been made for selecting the best technique for a model from the multi-class groups in order to achieve the best accuracy level using Support Vector Machine (SVM) with a significant option among efficiency and predictiveness. Further, before scheming the best kernel, it has been tried to find the best type of SVM based on multiclass methods in order to select the superior one and save time in its application. When a poor type or method is used, it will result in a significant loss in its predictive accuracy, which will lead to high misclassification rate. In this context, this paper investigates the various types of SVMs and compares their accuracy levels by computing the misclassification rate on real and simulation environment. It shows that the multi-class SVM approach outperforms the others in terms of accuracy.

Keywords: Support Vector Machine, Multi-class classification, Kernels

1. Introduction

Classification techniques are one of the statistical learning techniques in the context of machine learning procedures. It play a vital role in the context of classifying the objects of one or more groups. Among the various classification techniques, the SVM is most popular method to achieve more accuracy of classifying the groups. The accuracy is mainly based on the nature of data, SVM types and selection of kernel function. The classical procedure for classification can be shown optimal only under a series of assumptions like normality, independence and homogeneity, which have violations that will nullify the optimality seriously. The literature has established that SVMs is one of the robust classification models which assurances about the reasonable performance while using the datasets. SVMs are essentially good to make a model for a binary classifier, it can also be prolonged to multiple classes that can be applied for many real world applications.

In this paper, an attempt has been made to study the performance of various methods used in the context of multiclass SVMs. The brief introduction about SVM is presented in Section 2. Section 3 explores the various types of class procedures used to perform classification task. The results of the experimental study under real and simulation are presented in the section 4. The summary of research work is presented in the last section.

1) Support Vector Machine (SVM)

In recent decades SVM was introduced into the field of machine learning and its related area, which has received wide range of attention to researchers and it had made excessive progress in various fields. SVM has a solid theoretical foundation and straightforward mathematical model and has got considerable development in function estimation, time series forecasting, pattern recognition and many other areas. It uses a nonlinear mapping method to map original data into high-dimensional space for finding the optimal

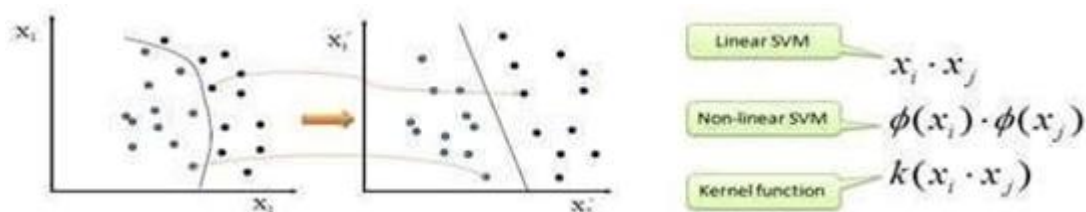


Figure 1: Types of SVM

classification by hyper plane that separates those data into different categories. SVM shown in figure 1 is based on the Vapnik–Chervonenkis dimension and structural risk minimization principle of statistical learning theory. This method gains great enhancement in generalization ability and also it plays an important role in the applications like forecasting financial statements, predicting the bankruptcy and analysis the credit risk.

The simplest way to separate two groups of data is with a straight line for one dimension, flat plane shows the two dimensions, for the n-dimensional hyperplane, situations where a nonlinear region can separate the groups more efficiently. SVM handles this by using a kernel function for the nonlinear to map the data into a different space where as for the hyperplane (linear) cannot be used to do the separation, a non-linear function is learned by a linear learning machine in a high-dimensional feature space while

the capacity of the system is controlled by a parameter that does not depend on the dimensionality of the space.

2) Multi-class Support Vector Machines

Multi-class SVMs are applied by merging more than a few binary SVMs for its methodology purpose, while applying the data, the modules has been generated based on its properties and limitation as we been considering to the research (binary and multi-class classification). Some of the methods used are One-Against-All (OvA_ls, OvA_hinge), All-Against-All (AvA_hinge, AvA_ls). Multiclass support vector machines has made a significant impact by its performance which are influenced by its methods and type of kernel and parameters. The Multiclass classifier has been checked for the standard datasets by the following procedures.

One-Against-All (OvA) In case of p-class problems ($p > 2$), p-binary SVM classifiers are constructed, the ith SVM samples class are considered as positive samples and all other remaining are taken as negative samples. In this step, a test sample is attained for all ‘p’ SVMs and it is been taken based on its extreme output amongst the ‘p’ classifiers, the classical methodology taken to solve this multiclass classification problems is to consider as a collection of binary classification problems. This method constructs p-classifiers, one for each class.

All-Against-All (AvA): The process for binary classifier for all the pair wise combination to the given p-classes are given $p \times (p-1) / 2$ binary classifiers, the first procedural step for the classifier, is by taking the first class as positive and second class as negative, then these classifiers are being combined, based on the Max Wins algorithm that will lead to find the subsequent class by selecting the class voted by the majority of the classifiers when it is smaller. This procedure is being followed for all versus all p-classes.

The above methods can be applied for one class or more than one class when it is in exclusive classes, also if the classification issues are more complex in nature it is one of the suitable methods. Furthermore, this method will give the competence of diverse methods in the multi class concepts when kernels are used for classification.

2. Experimental Study

This section presents the results of experimental study which were carried out under various methods of multiclass classifier on real and simulated data sets.

1) Real Data

The two real data sets was considered for the study, namely anorexia dataset and hemophilia data set. The hemophilia data (Habemma et al. (1974)) has two measured variables, AHF activity and AHV antigen on 75 women, belonging to two groups as the first group contains 30 observations belong to normal group and the second group contains 45 observations that belong to obligatory carrier. In the second experiment, the anorexia data (Hand et al.1993) which contains two variables of three groups with a frame of 72 observations. The data describes the weight change data for young female anorexia patients. The two variables are

weight of patients before study periods (prewt) and weight of patients after study periods (postwt). The three groups, namely Cognitive-behavioral treatment (CBT), Control (Cont), Family treatment (FT). The general description of the datasets based on groups is shown in (Figure 2).

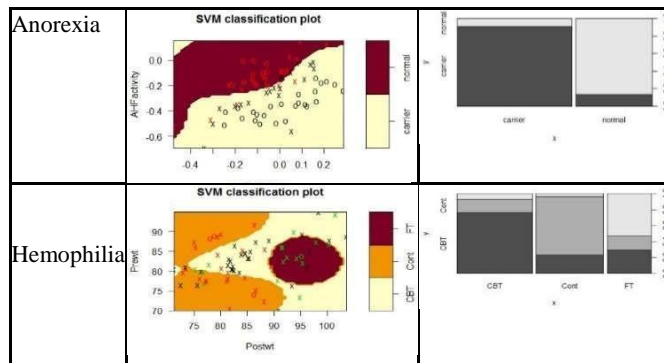


Figure 2: Graphical Description of Data sets based on the classifiers of groups

Table 1: Accuracy rate of classification under various types of SVM

Datasets	SVM	mcSVM	Multiclass SVM
Anorexia	0.694	0.819	0.861
Hemophilia	0.893	0.867	0.893

The above table shows the accuracy results for both datasets, it is observed that multiclass SVM gives more accuracy than the conventional SVM and mcSVM, along with that AVA method (i. e) All-Against-All is doing good for the real dataset. The detailed accuracy value of the dataset with all the methods both OVA and AVA are given in the appendix.

3. Simulation Study

The simulation study has been carried out by simulating the data of two and three groups of two and three dimensions respectively along with the various level of contaminations, specifically location, scale and both the location and scale contaminations. The data were generated under multivariate normal distribution.

For two groups, the mean vectors, (1, 1) and (3, 3), and the covariance matrix $1.5 \times I_2$ and $2 \times I_2$ were considered and generated 50 observations of each. For location contamination, the mean vectors, (-4,-4) and (-5,-5) were considered. For scale contamination, the covariance matrices are $3 \times I_2$ and $4 \times I_2$. The mean vectors, (-4,-4) and (-5,-5), and the covariance matrix $3 \times I_2$ and $2 \times I_2$ were considered for contamination of location and scale with the various levels such as 5%, 10%, 20%, and 30%.

For three groups, the mean vectors (0, 0, 0), (3, 3, 3) and (5, 5, 5) and the covariance matrix I_3 , $2.5 \times I_3$, $3 \times I_3$ respectively were considered and generated 50 observations of each. The location contaminations are applied as described using the mean vectors (-4,-4,-4), (-5,-5,-5) and (-7,-7,-7). For scale contamination, the covariance matrices are $2.5 \times I_3$, $5 \times I_3$ and $6 \times I_3$. In the case of location and scale contaminations, the mean vectors (-4,-4,-4), (-5,-5,-5) and (-7,-7,-7) with the covariance matrix $1.5 \times I_3$, $4 \times I_3$, $6 \times I_3$. The obtained accuracy of the classification under various levels of contaminations

is summarizes and given in the following table.

Table 2: Accuracy rate of classification under various types of SVM

(a) Location Contamination

Datasets	Types	0.00	0.05	0.10	0.20	0.30
Two Groups	SVM	0.900	0.860	0.850	0.840	0.820
	mcSVM	0.910	0.860	0.860	0.850	0.860
	Multiclass SVM	0.920	0.910	0.880	0.870	0.870
Three Groups	SVM	0.933	0.926	0.893	0.940	0.913
	mcSVM	0.940	0.953	0.953	0.946	0.940
	Multiclass SVM	0.950	0.966	0.960	0.953	0.953

(b) Scale Contamination

Datasets	Types	0.00	0.05	0.10	0.20	0.30
Two Groups	SVM	0.900	0.890	0.880	0.850	0.880
	mcSVM	0.920	0.900	0.890	0.840	0.870
	Multiclass SVM	0.930	0.910	0.900	0.860	0.880
Three Groups	SVM	0.920	0.940	0.933	0.920	0.900
	mcSVM	0.946	0.946	0.946	0.940	0.913
	Multiclass SVM	0.966	0.960	0.953	0.960	0.920

(c) Location and Scale Contamination

Datasets	Types	0.00	0.05	0.10	0.20	0.30
Two Groups	SVM	0.96	0.96	0.89	0.90	0.90
	mcSVM	0.94	0.90	0.92	0.88	0.85
	Multiclass SVM	0.94	0.98	0.96	0.89	0.87
Three Groups	SVM	0.973	0.926	0.926	0.926	0.906
	mcSVM	0.993	0.953	0.953	0.933	0.933
	Multiclass SVM	0.993	0.960	0.960	0.946	0.940

Based on the above Table 2 the accuracy level for the different contamination level done for Location and scale has been compared and it is evidence that Multiclass SVM is doing good, more over on precise it is observed the the method AVA has a clearly accuracy value shown in the appendix when it has been compared with OVA method.

4. Conclusion

Support vector machines are most widely used classification technique and are computationally efficient. Even though, it provides better classification for high-dimensional data, it suffers a lot when the data contains extremes. This paper explores the working algorithms of SVM and studied its performance under real and simulation environment with/without contaminations. The experimental study shows that, multiclass SVM algorithm performs better, by considering the good accuracy than conventional SVM and mcSVM algorithms under with and without contaminations. Further the study can be extended as subclass of classifiers that can be utilized to divide the data based on the inclusion of the robust kernel function for the given groups which will lead the model of SVM to achieve the best accuracy of classification.

References

[1] Chen. Y, Wen. Z, He. B and Chen. J, "Efficient Decomposition Selection for Multi-class Classification, " in IEEE Transactions on Knowledge and Data Engineering, 2021

[2] Cheng, S. O., Smola, A. J., Williamson, R. C., "Learning the kernel with hyper-kernel, " Journal of Machine Learning Research, Vol.6, pp.1043-1071, 2005.

[3] Cortes, C. and Vladimir, N. V. Support-vector networks. Machine Learning.20, 273– 297, 1995.

[4] Deepika, K., Jyostna, D. B and Veeranjaneyulu, N., "Effects of Different Kernels on the Performance of an SVM based Classification", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue – 5S4, February 2019.

[5] Fu, Y., Ren, D., "Kernel function and its parameters selection of support vector machines, " Science and Technology Innovation Herald, Vol.9, pp.6-7, 2010.

[6] Guo, L, Sun, S, X. Duan, X, "Research for support vector machine and kernel function, " Science Technology and Engineering, Vol.8, No.2, pp.487-489, 2008.

[7] Hong J-H, Cho S-B. A Probabilistic Multi-Class Strategy of One-vs.-Rest Support Vector Machines for Cancer Classification. Neurocomput.; 71: 3275-81, 2008

[8] Hsu C-W, Chang C-C, Lin C-J. A Practical Guide to Support Vector Classification. Department of Computer Science, National Taiwan University; 2003.

[9] Hsu C-W, Lin C-J. A comparison of methods for multiclass support vector machines. Neural Networks, IEEE Transactions.2002; 13: 415-25, 2002

[10] James. G., Witten. D., Hastie, T., and Tibshirani, R. An Introduction to Statistical Learning with applications in R, Springer, 2017.

[11] Li, T., Wang, X., "A semi-supervised support vector machine classification method based on cluster kernel, " Application Research of Computers, Vol.30, No.1, pp.42-45, 2013.

[12] Lijuan Liu, Bo Shen, Xing Wang, "Research on Kernel Function of Support Vector Machine", National Natural Science Foundation of China under Grant, and Beijing Natural Science Foundation. October 2013.

[13] Manevitz, L., and Yoursef, M., One-class svms for document classification. Journal of Machine Learning Research 2.2001.

[14] Muthukrishnan, R and Mahesh, K., Evaluation of classical and robust Discriminant Methods under Apparent Error Rate, International Journal of Current Research, 2817-2820, 2013.

[15] Muthukrishnan, R and Udaya Prakash, N., " Performance of Classification Techniques along with Support Vector Machine", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-2, December 2019

[16] Niazmardi S, Safari A. A novel multiple kernel learning framework for multiple feature classification, IEEE Journal of selected Topics in Applied Earth Observations and Remote Sensing, Vol.10, no.8, pp.3734-3743, 2017.

[17] R Core Team R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. URL, http://www.r-project.org., 2019.

[18] Rousseeuw, P. J., Multivariate estimation with high breakdown point, Mathematical Statistics and

Applications, 8, 283-297.1985
 [19] Trung Le, Dat Tran, Wanli Ma, Thien Pham, Phuong Duong, and Minh Nguyen, "Robust Support Vector Machine", International Joint Conference on Neural Networks (IJCNN), Beijing, China, July 2014.

[20] Vapnik, V. N., The Nature of Statistical Learning Theory, Springer Press, 1995.
 [21] Weston, J., and Herbrich, R. Adaptive margin support vector machines. In Advances in Large Margin Classifiers. MIT Press.2000.

Appendix: A

Real Dataset			
Dataset		Anorexia (72) Observation (three groups)	Hemophilia (75) Observation (two groups)
Types			
SVM		0.694	0.893
mcSVM		0.819	0.867
mcSVM	OvA	0.972	0.893
	OvA_ls	0.986	0.867
	OvA_hinge	0.917	0.853
	AvA	0.833	0.867
	AvA_hinge	0.888	0.853
	AvA_ls	0.986	0.893
SVM Multiclass		0.861	0.893
SVM Multiclass	OvA	0.986	0.880
	OvA_ls	0.944	0.853
	OvA_hinge	0.792	0.893
	AvA	0.889	0.893
	AvA_hinge	0.875	0.907
	AvA_ls	0.972	0.867

Appendix: B

Simulation study-Case (i): Location Contamination

Types	Dataset	Simulated Dataset (three groups)					Simulated Dataset (two groups)				
	case I										
	cont	0.00	0.05	0.10	0.20	0.30	0.00	0.05	0.10	0.20	0.30
	SVM	0.933	0.926	0.893	0.940	0.913	0.900	0.860	0.850	0.840	0.820
	mcSVM	0.940	0.953	0.953	0.946	0.940	0.910	0.860	0.860	0.850	0.860
mcSVM	OvA	0.950	0.953	0.953	0.946	0.926	0.940	0.970	0.950	0.900	0.840
	OvA_ls	0.940	0.946	0.953	0.946	0.933	0.970	0.980	0.980	0.940	0.860
	OvA_hinge	0.946	0.960	0.953	0.946	0.940	0.980	0.990	0.960	0.920	0.830
	AvA	0.940	0.966	0.960	0.946	0.920	0.960	0.910	0.950	0.900	0.860
	AvA_hinge	0.933	0.960	0.953	0.946	0.933	0.850	0.860	0.970	0.880	0.840
	AvA_ls	0.940	0.946	0.953	0.946	0.933	0.910	0.990	0.960	0.910	0.840
	SVM Multiclass	0.950	0.966	0.960	0.953	0.953	0.920	0.910	0.880	0.870	0.870
SVM Multi class	OvA	0.940	0.953	0.946	0.946	0.946	0.980	0.940	0.960	0.870	0.850
	OvA_ls	0.933	0.933	0.953	0.946	0.933	0.970	0.970	0.960	0.900	0.870
	OvA_hinge	0.933	0.960	0.953	0.940	0.940	0.960	0.930	0.930	0.860	0.860
	AvA	0.940	0.966	0.953	0.946	0.940	0.940	0.970	0.930	0.870	0.860
	AvA_hinge	0.960	0.960	0.953	0.946	0.946	0.980	0.930	0.920	0.920	0.870
	AvA_ls	0.960	0.953	0.960	0.946	0.926	0.940	0.950	0.960	0.900	0.860

Simulation study-Case (ii): Scale Contamination

Types	Dataset	Simulated Dataset (three groups)					Simulated Dataset (two groups)				
	cont	0.00	0.05	0.10	0.20	0.30	0.00	0.05	0.10	0.20	0.30
	SVM	0.92	0.94	0.933	0.92	0.9	0.9	0.89	0.88	0.85	0.88
	mcSVM	0.946	0.946	0.946	0.94	0.913	0.92	0.9	0.89	0.84	0.87
mcSVM	OvA	0.933	0.966	0.96	0.96	0.9	0.89	0.88	0.87	0.82	0.91
	OvA_ls	0.933	0.96	0.946	0.92	0.913	0.900	0.920	0.870	0.870	0.850
	OvA_hinge	0.973	0.966	0.946	0.926	0.906	0.920	0.900	0.870	0.830	0.920
	AvA	0.966	0.966	0.953	0.92	0.906	0.880	0.910	0.870	0.830	0.850
	AvA_hinge	0.94	0.966	0.933	0.926	0.913	0.890	0.940	0.870	0.870	0.880
	AvA_ls	0.953	0.966	0.933	0.926	0.9	0.890	0.900	0.880	0.860	0.870
	SVM Multiclass	0.966	0.96	0.953	0.96	0.92	0.93	0.91	0.9	0.86	0.88
	OvA	0.906	0.960	0.953	0.960	0.906	0.920	0.900	0.870	0.860	0.830
	OvA_ls	0.906	0.953	0.953	0.940	0.886	0.900	0.890	0.890	0.830	0.820

SVM Multi class	OvA hinge	0.900	0.960	0.940	0.926	0.906	0.900	0.910	0.850	0.830	0.840
	AvA	0.900	0.953	0.946	0.920	0.906	0.890	0.900	0.880	0.860	0.900
	AvA hinge	0.913	0.966	0.946	0.920	0.926	0.910	0.900	0.880	0.830	0.840
	AvA ls	0.900	0.960	0.946	0.940	0.913	0.900	0.920	0.880	0.830	0.830

Simulation study-Case (iii): Location and Scale Contamination

Types	Dataset	Simulated Dataset (three groups)					Simulated Dataset (two groups)				
cont		0.00	0.05	0.10	0.20	0.30	0.00	0.05	0.10	0.20	0.30
SVM		0.973	0.926	0.926	0.926	0.906	0.960	0.960	0.890	0.900	0.900
mcSVM		0.993	0.953	0.953	0.933	0.933	0.940	0.900	0.920	0.880	0.850
mcSVM	OvA	0.993	0.96	0.96	0.946	0.946	0.980	0.950	0.940	0.860	0.890
	OvA ls	0.993	0.946	0.953	0.946	0.926	0.910	0.990	0.960	0.910	0.850
	OvA hinge	0.993	0.96	0.953	0.953	0.953	0.860	0.950	0.980	0.850	0.850
	AvA	0.993	0.946	0.946	0.94	0.953	0.980	0.890	0.920	0.870	0.860
	AvA hinge	0.993	0.953	0.953	0.96	0.933	0.960	0.980	0.960	0.910	0.870
	AvA ls	0.993	0.966	0.953	0.946	0.933	0.980	0.970	0.900	0.940	0.870
SVM Multiclass		0.993	0.960	0.960	0.946	0.940	0.940	0.980	0.960	0.890	0.870
SVM Multi class	OvA	0.993	0.946	0.946	0.946	0.946	0.980	0.950	0.980	0.860	0.870
	OvA ls	0.993	0.946	0.946	0.940	0.946	0.950	0.970	0.940	0.910	0.890
	OvA hinge	0.993	0.953	0.960	0.940	0.933	0.960	0.950	0.940	0.880	0.860
	AvA	0.993	0.960	0.960	0.946	0.933	0.970	0.960	0.940	0.910	0.850
	AvA hinge	0.993	0.946	0.953	0.946	0.926	0.940	0.880	0.920	0.850	0.850
	AvA ls	0.993	0.946	0.953	0.946	0.920	0.970	0.890	0.950	0.880	0.850