Multi-class classification for large datasets with optimized SVM by non-linear kernel function

Lingam Sunitha¹ and M Bal Raju²

¹Department of CSE, Koneru Lakshmaiah Education Foundation, Deemed to be University, Hyderabad, Telangana-India, 500075. 
²Department of CSE, Swami Vivekananda Institute of Technology, Secunderabad, Telangana, India.

E-mail: sunithavvit@gmail.com

Abstract. Most important part of Support Vector Machines(SVM) are the kernels. Although there are several widely used kernel functions, a carefully designed kernel will help improve the accuracy of SVM. The proposed work aims to develop a new kernel function for a multi-class support vector machine, perform experiments on various data sets, and compare them with other classification methods. Directly multi-class classification with SVM is not possible. In this proposed work first designed a model for binary class then extended with the one-versus-all approach. Experimental results have proved the efficiency of the new kernel function. The proposed kernel reduces misclassification and time. Other classification methods observed better results for some data sets collected from the UCI repository.

Keywords: Support vector machines, Kernel Function, multi-class, Classification, Performance

1. Introduction

Support Vector Machines(SVM) are a prevalent tool in the recent literature of statistical learning and classification. SVM help to do classification and regression analysis. Many researchers and scientists are working on theory and applications like hand-written, text categorization, bioinformatics, and computer vision. Kernel functions help to regenerate feature spaces that an optimized hyperplane classifier is used to detect. The original idea of SVM is based on Mercer’s theorem [1], which defined several kernel function’s properties.

1.1 Support Vector Machines

History of SVM: In 1963, Vladimir Vapnik and Alexey Chervonenkis developed a new classification tool called SVM invented by Vapnik was a failure initially. It has been in the dark for nearly 27 years. Vapnik redefined his method and extended in 1990. SVM have become a famous tool for data scientists. Now it has become a popular machine learning algorithm and is highly used. In the year 1992 at the computational learning theory[2] conference. In his work, he proposed theoretical aspects and applications of maximization of margin. Duda and Hart [3] proposed a large margin hyper plane in 1973. The use of kernels was given by Wahba [4], Poggio and others.
Proposed work is divided into five sections. First section describes the SVM history. Second section is existing algorithms on multi-class SVM. Third section consists of proposed algorithm here derived a kernel function and objective function and extended to Multi class SVM by “one-versus-all” approach is explained. Finally in conclusion results are compared existing algorithms. [5]

2. Related Work
Initially, Support vector machines were actually for binary classification, then applied to multi-classes. A multi-class problem is viewed as the collection of binary class classification. Each one be individually solved by two class SVM. There is no surety to obtain optimal solution, even individually performed well. Genetic algorithm[6] is optimized by tuning the loss or penalty parameters. SVM also used for a single class, where all the examples are belonging to same class.

The classification problems with more than two classes is called Multi-class classification problem. Which is typically been solved by generating ‘K’ Binary classifier SVM and k-classes [7-8]. The K-class classification problem is to design a decision function considered. The support-vector network [9] is one new learning machine for binary classification. Input vectors cannot be linearly separated to transform the space of high dimensional functions. Due to special features of the decision boundary gives the more generation, which increases the machine learning ability. More generalized SVM networks uses the polynomial kernel. The classification of multi-class is based on support vector nominal type regression[10]. New SVM algorithm TSVM derived for multi classes [11-12]. The main aim of Twin SVM to detect two intersecting planes for each class. Every class is divided into a pair of small size quadratic programming sub problems. Yong Shi, Y. Tian proposed two types of multi class classification problems. Basic idea is "splitting and combine", this approach directly uses binary support vector classifier. Multi class classification can be solved in two ways that is "one verses rest" and "one verses one". C. Angulo, F. J. Ruiz published a paper on Tri-class [13]. Training data divided into two classes negative and positive used to detect faults[14] in sensors data.

3. Proposed algorithm

The objective function of the proposed multi-class SVM

$$\min_{w,b,\xi,\nu} \left\{ \frac{1}{2} w^T w - v \cdot \eta + \sum_{i=1}^{C} \lambda_i \right\}$$

ST: $y_i (w^T \phi(x_i, y_i) + b) \geq \eta - \lambda_i$

$\lambda_i \geq 0, i = 1 \ldots C, \eta \geq 0$

$\phi(x_i, y_i) = e^{x \log(\Sigma b_i^2)}; \text{if}(x_i > y_i)$

$= e^{x \log(\Sigma x_i^2)}; \text{if}(x_i < y_i)$

$= e^{x \log(\Sigma |x_i - y_i|^2)}; \text{if}(x_i == y_i)$

The decision boundary is given by the classifier

$$\text{sgn} \left( \sum_{i=1}^{C} y_i \phi(x_i, y_i) + b \right)$$

Algorithm multi-class SVM One-versus-all
1. Input $X=(x_1, x_2, x_i \ldots x_n}$ set of independent attributes
2. $y \in \{1, 2, \ldots N\}$, $Y$ is collection class variables
3. $N=$No. of classes
4. // Generating N Binary class models
5. for $j = 1$ to $N$
6. Index=compare($y$, class($j$))
7. svmmodel[$j$]=fitsvm($X$, index,classnames,’kernel’,’ExLog’)  
8. end
9. for each class $j$
10. Score=predict(SVMModels[$j$], x)
11. Score[$j$] = probability of +ve class
12. end
13. class of object=probability of max(score[$j$])

4. Experimental work

4.1 Data sets

4.1.1 Balance: All the data sets are collected from UCI [15]. Balance Data set consists of five attributes out of five four are weight of left ,right and distance of left , right and class attribute consists of L, B, R, i.e. Left, Right and Balance. It was generated with simulated model results on psychological experimental. Balance data set having 625 rows and four numeric attributes and one is class it consists 3 class labels.

4.1.2 Glass: Glass data set is a classification of glass types. This data set produced from experimental results of a criminal investigation. At the crime scene the glass left. Number of I rows 214, Number of Attributes: 10 and the class attribute, type of glass class attribute having five class labels.

4.1.3 Diabetes: Pima Indians Diabetes Data is consists of 768 instances and nine attributes and two classes. main motivation of this dataset is to identify the patient has diabetes or not, based on certain clinical test reports included in the dataset. Attributes are numerical and categorical. Especially all patients are female 21 and above years old of Pima Indian heritage.

4.1.4 heart: In heart data set "goal" class variable refers to the presence or absent i.e either 0 or 1. heart disease in the patient. is an integer values ranged from 0 to 4. This database contains 76 attributes, but all published experiments refer to using a subset of fourteen of them are used. data set having 303 tuples. Attributes are of different types i.e numeric,binary, categorical categorical. The person's age(numeric),sex(binary),chest pain type(categorical),resting blood type(numeric),serum cholesterol(numeric), fasting bloodsugar(numeric),fasting ECG result(categorical),maximum heart rate(numeric),number of major vessels(categorical), detected heart disease(binary), thalassemia(categorical), exercise(binary).

4.1.5 Spatial: This dataset consists of 89 attributes plus flag class and 208 instances.
4.2 Results

Table 1. Accuracy MC-SVM

| Dataset   | Accuracy |
|-----------|----------|
| balance   | 0.7049   |
| heart     | 0.9733   |
| diabetes  | 0.8680   |
| glass     | 0.6143   |
| spatial4  | 0.9269   |

Figure 1. Accuracy of MC-SVM with various datasets

From Table 1 and Figure 1, it is saying that for heart dataset proposed algorithm M-SVM is given highest performance 97.33%, and next spatial4 data set. For balance data 70.49% accuracy, diabetes 86.80%, glass 61.43% and spatial4 92.69%.

Table 2. Comparison with other existing algorithms

| Classification Models | Multi Class SVM(proposed) | Random Forest | Bayes Net |
|-----------------------|---------------------------|---------------|-----------|
| balance               | 0.7049                    | 0.8959        | 0.7711    |
| heart                 | 0.9733                    | 0.8733        | 0.8400    |
| diabetes              | 0.8680                    | 0.8320        | 0.8280    |
| glass                 | 0.6143                    | 0.8286        | 0.9000    |
| spatial4              | 0.9269                    | 0.8368        | 0.8951    |

Figure 2. Comparison of proposed with other algorithms

4.3 Observations

Random forest classification is well suited for a balanced dataset
Multi-class SVM with the new kernel, i.e. proposed algorithm, is suitable.
For the Diabetes data set, Multi-class SVM performance is better than other algorithms.
Bayes Net is suitable for glass data set. Proposed algorithm Multi-class SVM is better than other classification and it is well suited and given the highest 97% accuracy.

5. Conclusion
The enhancements noted with the novel kernel. It is designed and compared to existing classification algorithms. But these best results are not universal or fixed. The solution is achieved as a set of supported carriers may be rare. Weight vector minimization can be used as a criterion for regression problems with a modified loss function. The future guidance includes a technique for selecting the kernel function and additional capability control. Kernel development with invariance. Training is a relatively easy process. No local optimal, unlike in neural networks. It adapts relatively well to extensive data. Complexity and classifier error are inverse proportional

References
[1] K. Verma, S. Bhardwaj, R. Arya, M.S.U. Islam, M. Bhushan, A. Kumar, and P. Samant, “Latest tools for data mining and machine learning”, “International Journal of Innovative Technology and Exploring Engineering (IJITEE) “, vol. 8, no. 9S, pp. 18-23, ISSN: 2278-3075, 2019. Available:
[2] B.E.Boser, I.M.Guyon, and V.N.Vapnik. “A training algorithm for optimal margin classifiers”. Proceedings of the 5th Annual ACM Workshop on Computational Learning, pages 144–152, 1992.
[3] R.O.Duda and P.E. Hart. Pattern Classification and Scene Analysis. Wiley, 1973.
[4] G.Wahba,” Spline models for observational data. CBMS-NSF Regional conference series 
In Applied Mathematics", 59, 1990.
[5] A. Nalavade, A. Bai, M. Bhushan, “Deep Learning Techniques and Models for improving Machine Reading Comprehension System”, IJAST, vol. 29, no. 04, pp. 9692–9710, Oct. 2020
[6] Peng Xu, A. K. Chan. Support vector machines for multi-class signal classification with unbalanced samples ,IEEEExp,ore-2003.
[7] J. Weston, C. Watkins, Computer Science, Support vector machines for multi-class pattern recognition, ESANN,1999
[8] CecilioAnguloXavierParraAndreuCatalà , K-SVCR. A support vector machine for multi-class classification, Neurocomputing, Volume 55, Issues 1–2, September 2003, Pages 57-77
[9]. Corinna Cortes, V. Vapnik, support-Vector Networks, Computer Science, Machine Learning, Proceedings of the International Joint Conference on Neural Networks, 2003
[10] Shi Y., Tian Y., Kou G., Peng Y., Li J. Support Vector Machines for Classification Problems.. Advanced Information and Knowledge Processing. Springer, London.,2011
[11] Yitian Xu, R. Guo, Laisheng Wang, Twin Multi-Class Classification Support Vector Machine, 2012.Cognitive Computation,
[12] C. Angulo, F. J. Ruiz, Multi-Classification by Using Tri-Class SVM, Neural Processing Letters, February 2006, 23(1):89-101.
[13] Zeqian Xu, Tongling Lv, , A regression-type support vector machine for the k-class problem, 2018, Computer Science Neuro computing
[14] S. Jan, Young-Doo Lee, 2017.Sensor Fault Classification Based on Support Vector Machine and Statistical Time-Domain Features, Computer Science.
[15] Bache and M. Lichman. UCI machine learning repository, 2013.

5