International Journal of Basic Sciences and Applied Computing (IJBSAC)
ISSN: 2394-367X (Online), Volume-9 Issue-7, March 2023

Validate Model Endorsed for Support Vector Machine Alignment with Kernel Function and Depth Concept to Get Superlative Accurateness

Muthukrishnan. R, Udaya Prakash. N

***Abstract:** A support vector machine (SVM) is authoritative tool for statistical learning model which is well proved based on the literature reviews which is rooted in finding the operational risk. The Key factor is kernel function and its parameters selection. Once the debate of finalizing the influence factor (i.e) kernel parameters and error penalty factors, we can able to find the new kernel function as a proposed model by bring together the kernel with robust depth procedures. Here the GSOsvm has turn out to be best kernel function with local features to a global representative for any type of dataset. As a final point, experiments are done for dataset with different groups that are formed to show the superior value based on its accuracy on prediction of this kind of model which proves the best validation. Though many research readings suggest the usage of Radial basis (RBF) kernels and polynomial kernels for the conventional techniques, it was found that the results produced by these models have unreliable values, because of the sensitive in the data. The new kernel GSOsvm has the good reliable results both for real and simulated data values precisely when the data contains extreme observations that violated assumptions.*

***Keywords:** Radial Basis Kernel Function, KSVM, Kernel with Weights, Projection Depth*

I. INTRODUCTION

Support Vector Machine (SVM) are a suitable technique for classifying the data, the perseverance of SVM learners are consider as a technique that was aimed for promptly attaining a tolerable result. A classification task usually involves response value i.e. the class labels and quite a few attributes i.e. the features or observed variables. The objective of SVM is to make a model that predicts best accuracy or the misclassification rate for response or target variable. The functioning of the model is explained as a reduction between the complexity of model and essential risk according to information of dataset. SVM has the hard margin classifier which was estimated and developed continuously since it was proposed [1]. Various efforts has apply to it in order to solve various real problems in improving its learning and inference, by enhancing the SVM algorithm [10] several kernel functions have also been continuously suggested and studied.

Manuscript received on 13 February 2023 | Revised Manuscript received on 20 February 2023 | Manuscript Accepted on 15 March 2023 | Manuscript published on 30 March 2023.
*Correspondence Author(s)
Dr. R. Muthukrishnan, Professor, Department of Statistics, Bharathiar University, Coimbatore. (Tamil Nadu), India. E-mail: muthukrishnan1970@gmail.com
N. Udaya Prakash*, Research Scholar, Department of Statistics, Bharathiar University, Coimbatore. (Tamil Nadu), India. E-mail: udpk.06@gmail.com, ORCID ID: https://orcid.org/0000-0003-1271-9437

The application of SVM [2] are in the field of pattern recognition (discriminated analysis and classification) also for regression (time series analysis). Most of the functional usage are widely in the fields of prediction and comprehensive assessment where the researchers has the assurance to achieve the highest accuracy to intend in solve challenging problems. The paper concise section 2 discuss the importance of SVM and its types - mcSVM, KSVM, section 3 gives a brief on Depth procedures and how it is connected with KSVM and its kernel functions which results us to find the proposed method, section 4 shows the compared efficiency results of proposed method GSOsvm that has been verified for real and simulated data values, section 5 express the inference and its further study in the framework of GSOsvm Classification model.

II. PROMINENCE AND CATEGORIES OF SVM

2.1 SVM takes the simplest way to separate two groups of data with a straight line for one dimension, when it comes for two dimension it takes flat plane. Similarly, the n-dimensional is intended to hyperplane. SVM handles this by using a kernel function [3] for the nonlinear region for which the it can separate the groups more efficiently, to map the data into a different space. 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 [7].

2.2 mcSVM - Multiclass support vector machines has made a significant impact by its influenced methods and its type of kernel and parameters. The methods used are (a) One-Against-All (OvA_ls, OvA_hinge) as 'n'-class problems (n >2), s-binary SVM classifiers are constructed, the i^{th} SVM samples class are considered as positive samples and all other remaining are taken as negative samples (b) All-Against-All (AvA_hinge, AvA_ls) the process for binary classifier for all the pair wise combination to the given s – classes are given n × (n -1) / 2 binary classifiers, the first procedural step for the classifier, is by taking the first class as positive and second class as negative. Furthermore, this method will give the competence of diverse methods in the multi class concepts when kernels are used for classification.

2.3 KSVM The package kernlab has delivered the proposal of bringing KSVM from the SVM theories as it has diverse kernel function that are reprocessed by changing its set in the kernel parameters [13]. Ksvm uses John Platt's SMO algorithm for solving the SVM formulations with support class probabilities output and confidence intervals [14].

Classification are evolved here using c-SVC, nu-SVC. If we want to focus on regression that can be practice by applying the eps-svr, nu-svr formulation with native multi-class classification formulations and the bound constraint SVM formulation. Prediction is done by each of its classifier to produce robust results when used with SVMs [6]. Likewise, the ksvm has the proposal for providing the ability to produce class probabilities as output instead of class labels.

III. DEPTH PROCEDURE FORMATION

Depth procedure function can be used as the trailed idea in statistical inference for multivariate data. Two depth procedures (i.e) Projection Depth (PD) and Adjusted Projection Depth (APD) have been put forward to check its limitation, in order to find the best robust depth procedure. This paper explores how the types of support vector machine (ksvm) in combined with projection and adjusted projection depths has given rise to a model that classifies data based on irrespective of groups. Data depth was a function that quantifies the centrality of a point in a given data which is very near to the central or trimmed regions. As a follow up of that procedure the depth value are been generated irrespective for all observation of both real and simulated data. This section discusses about three depth algorithms that were used to obtain the depth values for the observation to the given datasets. The above conclusion is obtained based on [9] shows the comparison of various data depth algorithms were attempt based on two depth procedures (PD and APD) were compared in recent times which played a good role in the model of accuracy. The Algorithms for Fixed, Random and Gram-Schmidt Orthonormalization (GSO) for computational aspects of depth values in the framework of projection depth and adjusted depth procedure. The above three algorithms on depth procedure are been reviewed.

3.1 Projection Depth (PD)

This procedure has the first step to define the outlying of any point θ relative to the data set Y_m as explained in [15] the theoretical support is for the projection depth are as follows:

$$S\theta;Ym=\max v=1 v\tau\theta-medi v\tau yi MADi v\tau yi,$$

Where MAD are median absolute deviation for the univariate data set $\{u_1,\ldots,u_n\}$ its statistic $MAD_i\{u_i\} = med_i\left|u_i - med_j\{u_j\}\right|$.
Instead of the median and the MAD, also another pair (L, E) of a location and scatter estimate may be chosen. This leads to different notions of projection depth, all defined as

$$pdepth\theta;Yn=(1+S\theta;Yn)-1.$$

3.2 Adjusted Projection Depth (APD)

The univariate outlying function $s_1^* = (v^T y, v^T Y^n)$ are always symmetrical about $\mathrm{med}(v^T Y^n)$ with respect to y that results the projection depth to fail in capturing the real shape of the data, once the skewed data are exist [8]. This stimulates to extend the of an adjusted form of the above mentioned projection depth which are given as

$$\mathrm{APD}y,\ \mathrm{Y}n=1+\sup s1*v\mathrm{T}y,v\mathrm{TY}n-1$$

$$s_1^*(v^T y, v^T Y^n) = \begin{cases} \dfrac{v^T y - \mathrm{Med}(v^T Y^n)}{Q_1(v^T y) - \mathrm{Med}(v^T Y^n)}, & \mathrm{if} v^T y < \mathrm{med}(v^T Y^n) \\ \dfrac{v^T y - \mathrm{Med}(v^T Y^n)}{Q_3(v^T y) - \mathrm{Med}(v^T Y^n)}, & \mathrm{if} v^T y \geq \mathrm{med}(v^T Y^n) \end{cases}$$

3.3 KSVM with depths procedure

To develop a perfect model from one of the SVM, an approach has been made in blend with ksvm and the different projection depth like fixed, random and GSO method to acquire robust classification for multivariate dataset. A brief study on its algorithms have been done and these projection depth value are obtained for each observation both simulated and real data sets based on their respective procedures. It has been used as the weights that applied in KSVM, to analyses the accuracy performance of the dataset for classifying it.

3.4 Fixed and Random Projection Depth

The algorithm for both fixed and random depth procedure are well explained in [16], if the vector are from fixed direction if it follows the Fixed Depth procedure the *m* directions that cut the upper half plane equally will choose the maximum. Similarly, for the Random Depth procedure it uses a random choice of *m* directions that chooses optimal direction for computing the projection depth. Exclusively when fixed and random depth procedure are applied for gaining the depth value, the effects falls in low efficiency in their accuracy. But GSO algorithms has shown to increase the efficiency and less time in computational complexity in getting the depth value for the observations, presented in [11].

3.5 GSO svm Adjusted Projection Depth

Gram-Schmidt Orthonormalization (GSOsvm) adjusted projection depth has been proposed to get the performance of a SVM model for best accuracy value. The performance of GSOsvm has been studied and applied in Ksvm with combination of Radial Basis kernel function(RBF). The general GSO algorithm projection depth method has been talk over based on computational and theoretically method by [11]. The effectiveness of GSO algorithm has been tested by its misclassification error in SVM models for classifying the real and stimulating data situation. This means GSO procedure based adjusted projection depth estimators performs well when compared with fixed and random procedures. It has to be noted that depth values are computed for all the observation irrespective of the given dataset for *n* number of classes / groups and also *n* dimensions of variables with projection and adjusted projection depth procedure.

IV. EXPERIMENTAL STUDY

In this section we see the performance of the proposed method Ksvm with GSO is been compared with several methods of SVM, mcSVM, Ksvm and Ksvm through Random and Fixed Depth procedure.

4.1 Real Data

The two real data sets has been considered for the study, to be precise anorexia and hemophilia data set. The Hemophilia data [4] has two stately variables, AHF activity and AHV antigen of 75 women, as the first group contains 30 observations in normal group, followed by second group has 45 observations that have its place to obligatory carrier. The second experiment, Anorexia data [5] contains two variables of three groups with 72 observations. As the data been described as the weight change data for young female 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).

It can be noticed that GSOsvm method shows the good accuracy level for real datasets, when it is compared to number of SVM models in grouping with kernel function through several depth techniques. The Table 1 shows in the appendix reveal the key contrast, as the KSVM positioned on kernel function grouping with depth values shows the highest accuracy for *Hemophilia* about *97.33%* followed by *Anorexia* with *71%*. Whereas the remaining models of SVM without the depth values shows lesser accuracy. From this result it clearly gives the declaration, that inclusion of kernel function with depth values has a worthy part in its performance for a model. It is because depth values are being obtained for all observation in getting the supreme accuracy.

4.2 Simulation study

The data are formed as a result of multivariate normal distribution, as this simulation study has been carried out precisely by location, scale and both the location and scale contaminations. The simulating the data for two and three groups of two and three dimensions across various level of contaminations such as 5%, 10%, 20%, and 30%.

For two groups, the mean vectors, (1,1) and (3,3), and the covariance matrix $1.5*I_2$ and $2*I_2$ 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 $3*I_2$ and $4*I_2$. The mean vectors, (-4,-4) and (-5,-5), and the covariance matrix $3*I_2$ and $2*I_2$ were considered for contamination of location and scale.

For three groups, the mean vectors (0,0,0), (3, 3,3) and (5,5,5) and the covariance matrix I_3, $2.5*I_3$, $3*I_3$ 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 $2.5*I_3$, $5*I_3$ and $6*I_3$. 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 $1.5*I_3$, $4*I_3$, $6*I_3$.

The obtained accuracy value of the classification of data under various levels of contaminations is summarizes and specified in the Table 2 for two groups, Table 3 shows the same for three groups. It is observed that GSOsvm method has good accuracy value in both simulated values at different contamination level. Especially when the contamination level increase the accuracy value is also increase when compared with random, fixed method with depth concept. Its been established that GSOsvm method which based on adjusted projection depth algorithm performs a fitting model in making accuracy or misclassification rate when it is compared with other methods of SVM with and without depth procedure.

V. CONCLUSION

The proposed method GSOsvm method very adaptive and intensive for classification method as this model is been created based on RBF kernel function with depth values obtained individually for the observation irrespective of all groups. The result of the study also shows the effective performance in GSOsvm which has sensitive in data that gives worthy accuracy value through experimental study for real and simulation method. Further, the superiority of the GSOsvm proved over the usage by applying ksvm kernels with projection and adjusted projection depths resembling with Random, Fixed algorithms for PD and APD. Hence it can be concluded as a perfect kernel which plays a good functional for different kinds dataset. This will really help us to go for the further scope of study in building up the best model in finding the accurateness irrespective of attributes and groups that are focused on classification.

DECLARATION

Funding/ Grants/ Financial Support	No, we did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material/ Data Access Statement	Not relevant.
Authors Contributions	All authors have individual partnerships in this article. First Author: 60% (Review the article for theory concept) Second Author: 40% (application part and calculation)

REFERENCES

1. Boser, B. E., Guyon, I. M., & Vapnik, V. N. Training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, (1992) 144–152. [CrossRef]
2. Chang,C.C, Lin,C.J, LIBSVM - A Library for Support Vector Machines, http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html.
3. Guo, L, Sun, S, X. Duan, X, "Research for support vector machine and kernel function, Science Technology and Engineering", 8(2), (2008) 487-489
4. Habemma, J.D.F, Hermans, J. and van den Broek, K. Stepwise Discriminant Analysis Program Using Density Estimation in Proceedings in Computational statistics,(1974) 101–110
5. Hand, D. J., Daly, F., McConway, K., Lunn, D. and Ostrowski, E, A Handbook of Small Data Sets. Chapman & Hall, Data set, (1993) 285- 229 [CrossRef]

6. Hsu.C.W, Chang,C.C, Lin,C.J, A practical guide to support vector classification. Technical Report, Department of Computer Science, National Taiwan University,Taipei,(2003) 1-12
7. Huanjun Liu, Yaonan Wang, Xiaofen Lu, A Method to Choose Kernel Function and its Parameters for Support Vector Machines, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, (2005) 4277-4280 [CrossRef]
8. Hubert, M., Van der Veeken, S, Robust Classification for skewed data, Advances in Data Analysis and Classification, 4, (2010) 239–254 [CrossRef]
9. Muthukrishnan, R, Vadivel, M and Ramkumar,N, Gram-Schmidt Orthonormalization based Projection Depth, International Journal of Mathematics Trends and Technology, 52(6), (2017) 430- 434 [CrossRef]
10. 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, 9(2), (2019) 4366-4369 [CrossRef]
11. Muthukrishnan, R and Ramkumar, N, Computing adjusted Projection depth using GSO algorithm, AIP Conference Proceedings 2261, (2020) 030066-1- 030066-7 [CrossRef]
12. R Core Team, A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria,URL, http://www.r-project.org/index.html (2020)
13. Shengliang Lin, Zhi Liu, Parameters Selection Based on RBF Kernel Function. Jounal of Zhe Jiang University, 4(2), (2007) 121-145
14. Zhu, S., Zhang, R., Research for selection of kernel function used in support vector machine," Science Technology and Engineer, 8(16), (2008) 4513-4517
15. Zuo, Y, Projection-based depth functions and associated medians. The Annals of Statistics, 31, (2003) 1460-1490 [CrossRef]
16. Zuo, Y.J. and Lai, S.Y, Exact computation of bivariate projection depth and the Stahel-Donoho estimator. Computational Statistics and Data Analysis 55,(2011) 1173–1179. [CrossRef]

AUTHORS PROFILE

Dr. R. Muthukrishnan, was born in Tirunelveli. He holds B.Sc. (1991) in Computer Science from Madurai Kamaraj University, M.Sc. (1993) and Ph.D. (2000) in Statistics from Manonmaniam Sundaranar University. He is currently working as a Professor in the Department of Statistics, Bharathiar University, Coimbatore, Tamil Nadu, INDIA. He is a life member of various academic bodies such as ISPS, IBS, IISA and member in ISI and ISA. His main research work focuses on Robust Statistical Inference and Multivariate Analysis. He has more than 18 years of teaching and research experience. He has published 64 research papers in reputed national/international journals and guided 32 scholars for their research programs.

N. Udaya Prakash, Research scholar, Department of Statistics, Bharathiar University, his research interest conventional procedures in Statistical Inference associated with its alternatives in the context of robust methods predominantly in getting robust Inference techniques in SVM procedures using R software. His keen interest is towards Data Analytics with specific attention in the area of Predictive Modeling using Statistical software / applications.

Appendix

Table 1: Classification Accuracy Based on Kernel Functions Through Depth

				KSVM (with kernels & depths)		
Dataset	**SVM**	**mcSVM**	**KSVM**	**Random PD**	**Fixed APD**	**GSOsvm APD**
Hemophilia	0.8933	0.8670	0.8894	0.9600	0.9600	**0.9733**
Anorexia	0.6944	0.8190	0.6944	0.7069	0.7022	**0.7100**

Table 2: Classification Accuracy Based on Kernels with Depths for Two Groups Various Levels of Contaminations

	Two Groups					
				KSVM (with weights)		
Error	**SVM**	**mcSVM**	**ksvm**	**Random PD**	**Fixed APD**	**GSOsvm APD**
Location Contamination						
0.00	0.9600	0.9400	0.8583	0.9475	0.9502	**0.9498**
0.05	0.9600	0.9000	0.8577	0.9431	0.9423	**0.9428**
0.10	0.8900	0.9200	0.8774	0.9452	0.9451	**0.9455**
0.20	0.9000	0.8800	0.8849	0.9365	0.9351	**0.9369**
0.30	0.9000	0.8500	0.8516	0.9327	0.9324	**0.9336**
Scale Contamination						
0.00	0.9000	0.9200	0.9121	0.9194	0.9182	**0.9211**
0.05	0.8900	0.9000	0.9029	0.9155	0.9142	**0.9147**
0.10	0.8800	0.8900	0.8866	0.9266	0.9274	**0.9240**
0.20	0.8500	0.8400	0.8495	0.9072	0.9088	**0.9164**
0.30	0.8800	0.8700	0.8294	0.9329	0.9388	**0.9390**
Location & Scale Contamination						
0.00	0.9000	0.9100	0.9119	0.9185	0.9182	**0.9182**
0.05	0.8600	0.8600	0.9028	0.9149	0.9172	**0.9162**
0.10	0.8500	0.8600	0.8827	0.9193	0.9152	**0.9171**
0.20	0.8400	0.8500	0.8585	0.9101	0.9107	**0.9129**
0.30	0.8200	0.8600	0.8681	0.9448	0.9454	**0.9460**

Retrieval Number: 100.1/ijbsac.G0486039723
DOI:10.35940/ijbsac.G0486.039723
Journal Website: www.ijbsac.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 3: Classification Accuracy Based on Kernels with Depths for Three Groups Various Levels of Contaminations

	Three Groups					
				KSVM (with weights)		
Error	SVM	mcSVM	Ksvm	Random PD	Fixed APD	**GSOsvm APD**
Location Contamination						
0.00	0.9330	0.9400	0.9524	0.9676	0.9679	**0.9678**
0.05	0.9260	0.9530	0.9638	0.9652	0.9653	**0.9650**
0.10	0.8930	0.9530	0.9544	0.9576	0.9576	**0.9560**
0.20	0.9400	0.9460	0.9466	0.9468	0.9470	**0.9470**
0.30	0.9130	0.9400	0.9312	0.9340	0.9342	**0.9347**
Scale Contamination						
0.00	0.9200	0.9460	0.9448	0.9690	0.9682	**0.9686**
0.05	0.9400	0.9460	0.9621	0.9661	0.9659	**0.9655**
0.10	0.9330	0.9460	0.9501	0.9532	0.9543	**0.9527**
0.20	0.9200	0.9400	0.9475	0.9505	0.9512	**0.9513**
0.30	0.9000	0.9130	0.9290	0.9336	0.9336	**0.9336**
Location & Scale Contamination						
0.00	0.973	0.9930	0.9443	0.9704	0.9673	**0.9692**
0.05	0.926	0.9530	0.9598	0.9666	0.9663	**0.9668**
0.10	0.926	0.9530	0.9473	0.9664	0.9664	**0.9662**
0.20	0.926	0.9330	0.9270	0.9570	0.9572	**0.9576**
0.30	0.906	0.9330	0.9452	0.9448	0.9458	**0.9459**