The Generalization Ability of SVM Classification Based on Markov Sampling

Harsh Goyal (IIT2018114)
Aaditya Gadhave (IIT2018144)
Sourabh Gupta (IIT2018149)
Meet Singh Gambhir (IIT2018158)
Tushar Atrey(IIT2018159)

Semester VI
Department of IT
Indian Institute of Information Technology, Allahabad, India.

Abstract: In this paper we have proposed the use of markov sampling for improving the accuracy of SVM based classifiers, in the particular case of a dataset with a high number of features.

I. INTRODUCTION

A suspicion has been made in the given algorithm that the information being taken care of to the SVC is Independent and Identically Distributed (I. I. D.).

It is likewise realized that this supposition that isn't in every case genuine uncommonly when we are discussing information which is intrinsically temporal in nature, For instance - market forecast, discourse acknowledgment, and so on.

The algorithm utilizes a sampling technique known as Markov Sampling which has been taken from Markov chain Monte Carlo (MCMC) strategies. This technique is utilized to replace the irregular examining which is by and large used to prepare the SVCs.

II. General Terms Used

1. SVM :

SVM is a binary order model created by Vapnik from Structural Risk minimization hypothesis. It utilizes the method known as The Kernel stunt in which it changes the information and appropriately tracks down the ideal limit between the potential yields.

Reasons for SVMs being so important are -

- At the point when a dataset is considered with countless features and little sample size, SVMs are extremely amazing.
- Utilizing SVM, both basic and exceptionally complex grouping models can likewise be learned.
- SVM is useful in staying away from the overfitting of bends by using progressed numerical standards.

2. Markov sampling :

In insights, Markov chain Monte Carlo (MCMC) techniques involve a class of algorithms for examining from a likelihood circulation. By developing a Markov chain that has the ideal circulation as its probability dissemination, one can acquire an example of the ideal dispersion by recording states from the chain.

The more steps are incorporated, the more intently the appropriation of the samples coordinates with the real wanted dispersion. Different algorithms exist for building chains, including the Metropolis–Hastings algorithm.

Markov Sampling is motivated from Markov Chain Monte Carlo strategies. In measurements it includes a class of algorithms for examining from a probability dissemination. On the off chance that we develop a Markov chain with the ideal conveyance as per its probability dissemination, at that point we can acquire an example of the ideal appropriation by recording states from the chain.

The more the means included, the more intently the conveyance of the example coordinates with the real wanted dispersion. Quite possibly the most existing algorithms for development of chains is Metropolis – Hastings algorithm.

III. ALGORITHM DESCRIPTION

1. Let the size of training samples be m and m%2 be the remainder of m divided by 2. m+ and m−denote the size of training samples which label are +1 and −1, respectively. Draw randomly N1(N1 ≤ m) training samples {zi} N1 i=1 from the dataset Dtr. Then we can obtain a preliminary learning model f0 by SVMC and these samples. Set m+ = 0 and m− = 0.

2. Draw randomly a sample from Dtr and denote it the current sample zt. If m%2 = 0, set m+ = m+ + 1 if the label of zt is +1, or set m− = m− + 1 if the label of zt is −1.

3. Draw randomly another sample from Dtr and denote it the candidate sample z*.

4. Calculate the ratio P of e−(f0,z) at the sample z* and the sample zt, P = e−(f0,z*)/e−(f0,zt)

5. If P = 1, yt = −1 and y* = −1 accept z*with probability P = e−y*f0 /e−ytf0 . If P = 1, yt = 1 and y* = 1 accept z* with probability P = e−y*f0 /e−ytf0 . If P = 1 and yty* = −1 or P < 1, accept z* with probability P. If there are k candidate samples z* can not be accepted continuously, then set P = qP and with probability P accept z*. Set zt+1 = z*, m+ = m+ +1 if the label of zt is +1, or set m− = m− + 1 if the label of zt is −1 [if the accepted probability P (or P, P) is larger than 1, accept z* with probability 1].

6. If m+ < m/2 or m− < m/2 then return to Step 3, else stop it.

IV. RESULTS AND OBSERVATIONS

Dataset Link :-
https://archive.ics.uci.edu/ml/datasets/Letter+Recognition

Kernel	Accuracy
Linear	**78.283%**
RBF	**82.933%**
X^2	**70.916%**
Optimal Hyperparameter (On RBF)	**82.416%**

V. CONCLUSION

In the step 2 of the algorithm it is taken into consideration whether the number of samples is even or odd.

It is usually done by making appropriate changes i.e incrementing the count of positive and negative classes.

In the proposed algorithm the constants are taken as k = 5 and q = 1.2 . To compute the transition probabilities P or P’ or P”, the model has to be initially trained by SVM classifier using a subset of training dataset.

The accuracy achieved using a non-linear kernel (approx 84 %) is much higher than that of a linear one (approx 79%). It can be concluded that the problem is highly non-linear in nature.

VI. REFERENCES

[1] Steinwart, “Consistency of support vector machines and other regularized kernel classifiers,” IEEE Trans. Inf. Theory, vol. 51, no. 1, pp. 128–142, Jan. 2005.

[2] Steinwart and A. Christmann, “Fast learning from non-i.i.d. observations,” in Proc. Adv. Neural Inf. Process. Syst., vol. 22. Vancouver, BC, Canada, Dec. 2009, pp. 1768–1776.

[3] T. Steinwart, D. Hush, and C. Scovel, “Learning from dependent observations,” J. Multivariate Anal., vol. 100, no. 1, pp. 175–194, Jan. 2009.

[4] S. Smale and D. X. Zhou, “Online learning with Markov sampling,” Anal. Appl., vol. 7, pp. 87–113, Jan. 2009.

[5] F. Cucker and S. Smale, “On the mathematical foundations of learning,” Bull. Amer. Math. Soc., vol. 39, no. 4, pp. 1–49, Jan. 2001.