The Generalization Ability of SVM Classification Based on Markov Sampling

*Ayush Raj(IIT2018188), Sanket Mundra (IIT2018189), Harsh Bajaj (IIT2018190), Raunak Rathour(IIT2018196), Utkarsh Priyam (IIT2018197), Prateek Mishra(IIT2018199)*

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

***Abstract: All the past works in finding the generalization ability of SVM classification algorithm generally assumes that the samples are iid in nature (independent and identically distributed). This paper proposes a new method which focuses on generalization ability of SVMC based on uniformly ergodic Markov chain (u.e.M.c) samples. The results show that the SVMCs based on Markov sampling have better generalization ability (as the number of training samples are larger).***

1. **INTRODUCTION**

The terms Support Vector Machine (SVM) and high dimensional data are very closely related in classification based problems [1]-[3]. Although SVMCs have better results, they also have a very clear theoretical justification considering the universal consistency as well as the learning rates if the dataset is independent and identically distributed in nature [4]-[6]. But may not always be true. This iid assumption cannot be strictly applied in a real world example. For example, the speech recognition system and market prediction which are inherently temporal in nature and hence cannot be proved independent and identically distributed i.e., i.i.d. in nature. Hence, many researches are ongoing in this field with some relaxation on the i.i.d. constraints. Some of them considered the Markov chain samples and inferred the learning rates for online learning algorithms whereas some paper considered the convergence rates of empirical processes for stationary mixing sequences. α-mixing, β-mixing and φ-mixing were also discussed in some of the papers.

The authors in [1] have discussed and analysed the case when the input was Markov samples. The main reason for selecting the Markov series is that they are frequently and naturally available such as web-based content, marking predictions, DNA analysis, protein analysis, etc. Thus, inspired from the Markov Chain Monte Carlo (MCMC) method, a sampling algorithm is discussed for sampling the dataset.

1. **ALGORITHM**
2. Let m be the size of training samples 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}i=1N1 from the dataset Dtr. Then we can obtain a preliminary learning model fo by SVMC and these samples. Set m+ = 0 and m- = 0.
3. Draw randomly a sample from Dtr and denote it as 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.
4. Draw randomly another sample from Dtr and denote it the candidate sample z\*.
5. Calculate the ratio P of e-l(f0,z) at the sample z\* and the sample zt, P = e−((f0,z∗)/e−(f0,zt))
6. 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].
7. If m+ < m/2 or m- < m/2 then return to Step C, else stop it.
8. **RESULTS AND OBSERVATIONS**

On implementing the above approach on the given dataset , this model was able to achieve approx. 84.2% to 88% accuracy on 25 iterations.

1. **CONCLUSION**

Thus, the generalization ability of SVMCs based on Markov Sampling has improved, as the number of training samples are very large.

1. **REFERENCES**

**[1]** Xu, Jie, et al. "The generalization ability of SVM classification based on Markov sampling." IEEE transactions on cybernetics 45.6 (2014): 1169-1179

**[2]** V. Vapnik, Statistical Learning Theory. New York, NY, USA: Wiley, 1998.

**[3]** Y. J. Tian, Z. Q. Qi, X. C. Ju, Y. Shi, and X. H. Liu, “Nonparallel support

vector machines for pattern classification,” IEEE Trans. Cybern., vol. 44,no. 7, pp. 1067–1079, Jul. 2014.

**[4]** Swapna Goutham, Soman KP and Vinayakumar R, Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals, January 2018 Procedia Computer Science 132:1253-1262, DOI: 10.1016/j.procs.2018.05.041

**[5]** I. Steinwart, “Consistency of support vector machines and other

regularized kernel classifiers,” IEEE Trans. Inf. Theory, vol. 51,

no. 1, pp. 128–142, Jan. 2005.

**[6]** S. Geman and D. Geman, “Stochastic relaxation, Gibbs distribution and

the Bayesian restoration of images,” IEEE Trans. Pattern Anal. Mach.

Intell., vol. 6, no. 6, pp. 721–741, Nov. 1984.