The Generalization Ability of SVM Classification Based on Markov Sampling

M.Rahul Reddy (IIT2018103), N.Lokesh Naik (IIT2018104), A.Prathyush (IIT2018124), M.Maheswarnath (IIT2018137), G.Sai Pavan (IIT2018139)

6th Semester, B.Tech, Department of Information Technology, IIIT Allahabad

Abstract:In this paper, we will discuss the generalization ability of SVMC based on uniformly ergodic Markov chain (u.e.M.c.) samples and obtain the optimal learning rate of SVMC for u.e.M.c. samples. We also introduce a new Markov sampling algorithm for SVMC to generate u.e.M.c. samples and present the numerical studies on the learning performance of SVMC based on Markov sampling.

I. Introduction

Support vector machine (SVM) is one of the most widely used machine learning algorithms for classification problems, in particular for classifying high-dimensional data. Besides their good performance in practical applications, they also enjoy a good theoretical justification in terms of both universal consistency and learning rates, if the training samples come from an independent and identically distributed (i.i.d.) process.Markov sampling has three advantages at the same time compared with the classical SVMC and the SVMC with Markov sampling:

1. The misclassification rates are smaller.
2. The total time of sampling and training is less.
3. The obtained classifiers are more sparse.

II. Markov Sampling Algorithm

For a given original training sample set Dtr, the new Markov sampling algorithm for SVMC is stated as follows.

- **Step 1:** 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 $N_1(N_1 \leq m)$ training samples $\{z_i\}$ N_1 i=1 from the dataset Dtr. Then we can obtain a preliminary learning model f_0 by SVMC and these samples.Set m+ = 0 and m− = 0.
- **Step 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.

- **Step 3:** Draw randomly another sample from Dtr and denote it the candidate sample z_*.

- **Step 4:** Calculate the ratio P of e−(f0,z) at the sample z_* and the sample zt, P = e−(f_0,z_*)/e−(f_0,zt).
- **Step 5:** If P = 1, y_t = −1 and y_* = −1 accept z_* with probability P = e−y_*f_0 /e−ytf$_0$. If P = 1, y_t= 1 and y_* = 1 accept z_* with probability P = e−y_*f_0 /e−y_tf_0 . If P = 1 and y_ty_* = −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 z_t+1 = z_*, m+ = m+ +1 if the label of z_t is +1, or set m− = m− + 1 if the label of z_t is −1 [if the accepted probability P (or P, P) is larger than 1, accept z* with probability 1].
- **Step 6:** If m+ < m/2 or m− < m/2 then return to Step 3, else stop it.

III. Experiment Results:

Kernel	KPCA	SVDD	OCSVM	OCSSVM	OCSSM with SMO	**MS SVM**
Linear	0.02	0.09	0.01	0.07	0.04	0.76
RBF	0.05	0.07	0.14	0.09	0.04	0.82
Intersection	0.18	0..01	0.04	0.26	0.22	
Hellinger	0.01	0.02	0.02	0.13	0.10	
X^2	0.18	0.0	0.02	0.18	0.17	
Polynomial						0.72
Sigmoid						