The Generalization Ability of SVM Classification Based on Markov Sampling

IIT2018103,IIT2018104,IIT2018124,IIT2018137,IIT2018139

**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.**

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 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 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.

**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−(f0,z∗)/e−(f0,zt).

**Step 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].

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

Experiment Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Kernel** | KPCA | SVDD | OCSVM | OCSSVM | OCSSM with SMO | **FS\_**  **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 |  |
| X2 | 0.18 | 0.0 | 0.02 | 0.18 | 0.17 |  |
| Polynomial |  |  |  |  |  | 0.72 |
| Sigmoid |  |  |  |  |  |  |