INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, ALLAHABAD

VI Semester B.Tech in Information Technology

Report - Group Assignment 3

Data Mining and Warehousing

k-Times Markov Sampling for SVMC

By :-

Palak Mittal (IIT2018117)

Nehal Singh (IIT2018119)

Abhishek Kumar Gupta (IIT2018187)

Puja Kumari (IIT2018191)

Prabha Kumari (IIT2018195)

QUESTION:-

You have to understand the algorithm proposed in the paper "k -Times Markov Sampling for SVMC ".
Run the algorithm on the shared given two datasets and show the accuracy in terms of the attached image table: (make one more column in the last name KT_SVM with the new algorithm and give the result).
Try to run the algorithm using different kernels as the attached image table.
1) http://host.robots.ox.ac.uk/pascal/VOC/voc2012/ - pascal
2) https://archive.ics.uci.edu/ml/datasets/Letter+Recognition - letter

Kernel	KPCA	SVDD	OCSVM	OCSSVM	OCSSVM with SMO
Linear	0.02	0.09	0.01	0.07	0.04
RBF	0.05	0.07	0.14	0.09	0.04
Intersection	0.18	0.01	0.04	0.26	0.22
Hellinger	0.01	0.02	0.02	0.13	0.10
χ^2	0.18	0.0	0.02	0.18	0.17

SOLUTION

1. INTRODUCTION

SMVC is one of the best classification methods used for classification as it takes all the edges and boundaries in account and then classifies the data in different classes. If training samples are identically distributed and unique, SVMC has good consistency rates. It is theoretically very well validated for the independent samples, but it is very important for us to study its practical implementations and the extent of accuracy and limitations it holds practically. Therefore, we will try to implement it in the Markov chain.

2. PROPOSED PROBLEM

How to reduce the training and sampling time of SVMC with Markov sampling introduced, at the same time keeping its smaller classification rates?
So, to resolve it and get a solution and conclusion we will compare the results of classical SMVC with SMVC with Markov sampling introduced.

3. PRELIMINARIES

1. SVM CLASSIFICATION ALGORITHM
2. UNIFORMLY EGRODIV MARKOV CHAINS
3. Generalization Ability of SVMC With u.e.M.c. Samples

4. ALGORITHM FOR SMVC WITH MARKOV

Initialisation of variables:

- S(t) is the sample set
- m is the size of sample set
- n is size of iid
- S(iid) is the training subset
- S(mar) training subset
- N(+), N(-) is the size of samples for training which are marked as +1 and -1
- α = e−(fi,z*)/e−(fi,zt)

Pseudo code:

1: Draw randomly N samples $S_{iid} := \{z_j\}_{j=1}^{N}$ from S_T. Train S_{iid} by SVMC and obtain a preliminary learning model f_0. Let $i = 0$.

2: Let $N_+ = 0$, $N_- = 0$, $t = 1$.

3: Draw randomly a sample z_t from S_T, called it the current sample. Let $N_+ = N_+ + 1$ if the label of z_t is $+1$, or let $N_- = N_- + 1$ if the label of z_t is -1.

4: Draw randomly another sample z_* from S_T, called it the candidate sample, and calculate the ratio α, $\alpha = e^{-\ell(f_i,z_*)}/e^{-\ell(f_i,z_t)}$.

5: If $\alpha \geq 1$, $y_t y_* = 1$ accept z_* with probability $\alpha_1 = e^{-y_* f_i}/e^{-y_t f_i}$. If $\alpha = 1$ and $y_t y_* = -1$ or $\alpha < 1$, accept z_* with probability α. If there are n_2 candidate samples can not be accepted continually, then set $\alpha_2 = q\alpha$ and accept z_* with probability α_2. If z_* is not accepted, go to Step 4, else let $z_{t+1} = z_*$, $N_+ = N_+ + 1$ if the label of z_{t+1} is $+1$ and $N_+ < N/2$, or let $z_{t+1} = z_*$, $N_- = N_- + 1$ if the label of z_{t+1} is -1 and $N_- < N/2$ (if the value α (or α_1, α_2) is bigger than 1, accept the candidate sample z_* with probability 1).

6: If $N_+ + N_- < N$, return to Step 4, else we obtain N Markov chain samples S_{Mar}. Let $i = i + 1$. Train S_{Mar} by SVMC and obtain a learning model f_i.

7: If $i < k$, go to Step 2, else output $sign(f_k)$.

Implementation :

Results

DATASET 1:

Accuracy of "**k -Times Markov Sampling for SVMC**" with different Kernels for LETTER-RECOGNITION dataset.

Kernel	KPCA	SVDD	OCSVM	OCSSVM	OCSSVM with SMO	KT_SVM
Linear	0.02	0.09	0.01	0.07	0.04	0.8121
RBF	0.05	0.07	0.14	0.09	0.04	0.869
Intersection	0.18	0.01	0.04	0.26	0.22	0.0192
Hellinger	0.01	0.02	0.02	0.13	0.10	0.7092
chi_square	0.18	0.0	0.02	0.18	0.17	0.8495

DATASET 2

Accuracy of "**k -Times Markov Sampling for SVMC**" with different Kernels for PASCAL dataset.

Kernel	KPCA	SVDD	OCSVM	OCSSVM	OCSSVM with SMO	KT_SVM
Linear	0.02	0.09	0.01	0.07	0.04	0.2160
RBF	0.05	0.07	0.14	0.09	0.04	0.3092
Intersection	0.18	0.01	0.04	0.26	0.22	TOO MUCH TIME
Hellinger	0.01	0.02	0.02	0.13	0.10	0.18
chi_square	0.18	0.0	0.02	0.18	0.17	0.2336

CONCLUSION:

An improved SMVC is introduced in this paper that is SMVC with k times Markov sampling. This is to improve the results that originated from the classical SMVC and the SMVC with Markov sampling. SMVC is known to give good results theoretically and therefore it was very necessary to use and develop the algorithm to give the practice results that can be imposed on the real life problem.

Along the line of this paper, several open problems deserves further research, for example, studying the performance of SVM for regression based on k-times Markov sampling and establishing the bounds on the support vector numbers for the SVMC with k-times Markov sampling. All these problems are under our current investigation.