DMW Assignment-1

Submitted By - [Akhil Shukla, IIT2018112] [Akhil Singh, IIT2018198][Javed Ali, IIT2018501][Manan Bajaj, IIT2018502][Lokesh, IIT2018503]

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

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).

Markov Sampling Algorithm Implementation

We use Letter Dataset[2], it has 16 different features relating to alphabets A and B (for forming a binary classifier as given in paper) to be recognized. First we segment the dataset into a train and test set with 1088 samples for training and 467 for testing. We use k times-markov sampling (explained next) to choose samples from the training set that forms a markov chain.

K-times Markov Sampling Algorithm

1. Draw randomly N samples iid from ST . Train S_{iid} by SVMC and obtain a preliminary learning model f_0. Let i = 0.

We used N1 = 800, m = 1088

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

3: Draw randomly a sample z_t from ST , called it the current sample. Let $M_+ = M_+ + 1$ if the label of z_t is +1, or let $M_- = M_- + 1$ 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, $\alpha = e^{-l(f_i, z^*)}/e^{-l(f_i, z_t)}$.

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 α = 1 and $y_t\, y_* = -1$ or α < 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 $z_t+1 = z_*$, $M_+ = M_+ + 1$ if the label of z_t+1 is +1 and $M_+ < M/2$, or let $z_t+1 = z_*$, $M_- = M_-+1$ if the label of z_t+1 is −1 and $M_- < M/2$ (if the value α (or α_1, α_2) is bigger than 1, accept the candidate sample z_* with probability1).

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

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

Then we train the SVM Classifier with different kernels using the markov samples. The final classifier is tested against the test dataset and performance recorded.

Various SVM kernels (Hellinger, Intersection and Chi Squared) that are not directly implemented in SVC are implemented using custom kernels which follow the mentioned kennel functions-

kernel	$k(x, y)$
Hellinger's	$\sqrt{xy}$
χ^2	$2\frac{xy}{x+y}$
intersection	$\min\{x, y\}$

Figure1 : Simplified Kernel functions from [6]

Hellinger is a simple square root of dot product, Chi Squared is implemented using "AdditiveChi2Sampler" and "SGDClassifier" of sklearn. SGD classifier follows SGD training for SVM cited[3, 4]. Intersection kernel is implemented as a simple inner product of X and Y matrices cited[5, 7].

Observation
K = 20 Times Markov Sampling
(with q = 1.2 and k = 5 for updating P ratio when continuous reject)
N = 500 samples (Initial training sample for SVMC model inside sampling loop)

Accuracy on Linear Kernel SVM - 99.7858 %
Accuracy on RBF Kernel SVM - 100.0 %
Accuracy on Polynomial Kernel SVM - 99.7858 %
Accuracy on Hellinger Kernel SVM - 99.5717 %
Accuracy on Chi Squared Kernel SVM - 99.8041 %
Accuracy on Intersection Kernel SVM - 99.7858 %

Misclassification Rate on Linear Kernel SVM - 0.0021 %
Misclassification Rate on RBF Kernel SVM - 0.0 %
Misclassification Rate on Polynomial Kernel SVM - 0.0021 %
Misclassification Rateon Hellinger Kernel SVM - 0.0043 %
Misclassification Rate on Chi Squared Kernel SVM - 0.0019 %
Misclassification Rate on Intersection Kernel SVM - 0.0021 %

A1 SVM Misclassification						
Kernel	KPCA	SVDD	OCSVM	OCSSVM	OCSSVM with SMO	KT_SVM
Linear	0.02	0.09	0.01	0.07	0.04	0.0021
RBF	0.05	0.07	0.14	0.09	0.04	0
Intersection	0.18	0.01	0.04	0.26	0.22	0.0021
Hellinger	0.01	0.02	0.02	0.13	0.1	0.0043
Chi Squared	0.18	0	0.02	0.18	0.17	0.0019

The output by running Markov Sampling k times is much better than simply Running Markov Sampling once. The SVMC with Markov sampling introduced in Assignment3 is batch learning done twice on the m training samples whereas k time Markov Sampling is batch learning done k(k=20) times on N(=500) training samples and the total number of training samples is $(k + 1)N$.

References

[1] Zou, Bin, et al. "$ k $-Times Markov Sampling for SVMC." IEEE transactions on neural networks and learning systems 29.4 (2017): 1328-1341.

[2] Letter Dataset - https://archive.ics.uci.edu/ml/datasets/Letter+Recognition

[3] sklearn kernel approximations AdditiveChi2Sampler
https://scikit-learn.org/stable/modules/generated/sklearn.kernel_approximation.AdditiveChi2Sampler.html#sklearn.kernel_approximation.AdditiveChi2Sampler

[4] sklearn Linear Model SGD Classifier
https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html

[5] A. Barla, F. Odone and A. Verri, "Histogram intersection kernel for image classification," Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), Barcelona, Spain, 2003, pp. III-513. doi: 10.1109/ICIP.2003.1247294

[6] Vedaldi A, Zisserman A. Efficient additive kernels via explicit feature maps. IEEE Trans Pattern Anal Mach Intell. 2012 Mar;34(3):480-92. doi: 10.1109/TPAMI.2011.153. PMID: 21808094.

[7] Intersection Kernel Function
https://stats.stackexchange.com/questions/48506/what-function-could-be-a-kernel