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 26 alphabets to be recognized.

First we segment the dataset into a train and test set with 14000 samples for training and 6000 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 Siid by SVMC and obtain a preliminary

learning model f0. Let i = 0.

2: Let N+ = 0, N− = 0, t = 1.

3: Draw randomly a sample zt from ST , called it the current sample. Let N+ = N+ + 1 if the label of zt is +1, or let N− = N− + 1 if the label of zt is −1.

4: Draw randomly another sample z∗ from ST , called it the candidate sample, and calculate the

ratio α, α = e−l( fi,z∗)/e−l( fi,zt).

5: If α ≥ 1, yt y∗ = 1 accept z∗ with probability α1 = e−y∗ fi /e−yt fi . If α = 1 and yt y∗ = −1

or α < 1, accept z∗ with probability α. If there are n2 candidate samples can not be accepted

continually, then set α2 = qα and accept z∗ with probability α2. If z∗ is not accepted, go to Step

4, else let zt+1 = z∗, N+= N+ + 1 if the label of zt+1 is +1 and N+ < N/2, or let zt+1 = z∗,

N− = N−+1 if the label of zt+1 is −1 and N− < N/2 (if the value α (or α1, α2) is bigger than 1,

accept the candidate sample z∗ with probability1).

6: If N+ +N− < N, 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 fi .

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

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.

***Observation***

Algorithm is run for k-Markov Sampling with k = 10 and q = 1.2

Each iteration in the k times sampling starts with an initial 2000 random samples from the 14000 training samples. Finally in the submitted run 18497 samples are used following the morkov sampling.

Accuracy on Linear Kernel SVM - 83.66 %

Accuracy on RBF Kernel SVM - 93.9 %

Accuracy on Polynomial Kernel SVM - 89.08 %

Misclassification Rate on Linear Kernel SVM - 16.34 %

Misclassification Rate on RBF Kernel SVM - 6.1 %

Misclassification Rate on Polynomial Kernel SVM - 10.92 %

***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