Data Mining Assignment-3
The Generalization Ability of SVM Classification Based on Markov Sampling

---

Bindu (IIT2018105)
Vikas(IIT2018110)
Ayushi Gupta(IIT2018118)
Sanjana(IIT2018120)
TS. Chinmay(IIT2018138)

I. Introduction:

The paper given to be reviewed studies the generalization ability of SVMC based on uniformly ergodic Markov chain (u.e.M.c.) samples.They have presented a new Markow Sampling Algorithm for SVMC to generate u.e.M.C samples and also have presented its numerical studies.

II. Algorithm:

SVM is one of the most popular algorithms for classification samples. It performs very efficiently in case the samples independently and identically distributed .Since most of the real world data are in the form of markov chains the paper given to be reviewed has studied the efficiency of svmc based on markov sampling and for this the have proposed a new markov sampling algorithm for SVMC.
The markov sampling algorithm presented is as follows :
Step 1: Let say the size of the training data be n and n%2 be the remainder we get by dividing n by 2. n- and n+ be the size of the training set labelled as -1 and +1 respectively. Let us withdraw a sample N1 (N1<=n) from the training set and design a SVMC model f from these sample. Now set n-=0 and n+=0.
Step 2: Again draw a random data sample from the dataset and name it z1. If n%2=0 and the label size of z1 is 1 then n+=n+1 else n-=n-+1.
Step 3: Again draw a random sample z* from the dataset.
Step 4: Find the ratio of probability measure of both samples z1 and z* P.
Step 5: On the off chance that P = 1, yt = −1 and y∗ = −1 acknowledge z∗ with likelihood $P = e^{-y^*f}/e^{-ytf}$. In the event that P = 1, yt = 1 and y∗ = 1 acknowledge z∗ with likelihood $P = e^{-y^*f}/e^{-ytf}$. On the off chance that P = 1 and yty∗ = −1 or P < 1,

acknowledge z* with likelihood P. On the off chance that there are k competitor tests z* can not be acknowledged ceaselessly, at that point set P = qP and with probability P acknowledge z*. Set zt+1 = z*, n+ = n+ +1 on the off chance that the mark of zt is +1, or set n− = n− + 1 if the mark of zt is −1 [if the acknowledged probability P (or then again P, P) is bigger than 1, acknowledge z* with likelihood 1].

Step 6: If n+<n/2 or n-<n/2 then go to step 3 else stop the process.

It has been found that SVM performs better with markov sampling as compared to random sampling.