The Generalization Ability of SVM Classification
Based on Markov Sampling

B.Tech (IT) 6th semester, Indian Institute of Information Technology, Allahabad

Deepak Katre (IIT2018116), Anurag Makade (IIT2018121), Dhairya Patel (IIT2018122),
Akinchan Lunavat (IIT2018123), Kapil Gupta (IIT2018127)

Summary

Support vector machines (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. However, independence is a very restrictive concept. First, it is often an assumption, rather than a deduction on the basis of observations. Second, it is an all or nothing property, in the sense that two random variables are either independent or they are not—the definition does not permit an intermediate notion of being nearly independent. Therefore, relaxations of such i.i.d. assumptions have been considered for quite a while in both machine learning and statistics literature. There are many dependent sampling mechanisms (e.g., α-mixing, β-mixing and φ-mixing) studied in machine learning literature. In this paper, the authors focus only on an analysis in the case when the input samples are Markov chains. In real-world problems, Markov chain samples appear so often and naturally in applications, such as biological (DNA or protein) sequence analysis, content-based web search, marking prediction, and so on. Also, many empirical evidence show that learning algorithms very often perform well with Markov chain samples (e.g., biological sequence analysis, speech recognition). In this paper, they first establish two new concentration inequalities for uniformly ergodic Markov chains (u.e.M.c.), and then they establish the optimal learning rate of support vector machine classification (SVMC) for u.e.M.c samples. Inspired by the idea from Markov chain Monto Carlo (MCMC) methods, they introduced a new Markov sampling algorithm for SVMC to generate u.e.M.c. samples from a given dataset. The numerical studies of real-world datasets show that the SVMC based on Markov sampling not only has better learning performance, but also the classifiers are sparsity as the size of data is bigger with regard to the dimension of data.

Results

a) Markov Sampling for Pascal dataset

Dataset Link - http://host.robots.ox.ac.uk/pascal/VOC/voc2012/

Kernel	MS_SVM
Linear	0.20
RBF	0.29
Hellinger	0.18
X^2	0.22

b) Markov Sampling for Letters dataset

Dataset Link - https://archive.ics.uci.edu/ml/datasets/Letter+Recognition

Kernel	MS_SVM
Linear	0.85
RBF	0.91
Hellinger	0.78
X^2	0.96