SVM-Boosting based on Markov resampling

Meghna IIT2018109
Riya Chaudhary IIT2018145
Vidhi Sah IIT2018169
Nandini Goyal IIT2018173
Chaitali Agrawal IIT2018504

Semester VI, Department of IT, Indian Institute of Information Technology, Allahabad, India.

Abstract: In this paper we introduce the idea of Markov resampling for Boosting methods. We first prove that Boosting algorithm with general convex loss function based on uniformly ergodic Markov chain (u.e.M.c.) examples are consistent and establish its fast convergence rate. We apply a Boosting algorithm based on Markov resampling to Support Vector Machine (SVM).

I. INTRODUCTION

With the advent of the high-tech era, the capacity of data is growing rapidly, and the value density of data is usually very low, which implies that there are many noise examples in big data. While the main idea of the AdaBoost algorithm is to adjust the weights of training examples so that the examples misclassified by the last classifier will be focused in the next train. Thus AdaBoost algorithm will be very time-consuming or hard to implement as the size of data is very big.

In this article we introduce the idea of Markov resampling for Boosting methods to sample a small amount of training examples from this given data and then these examples are used to train the base classifiers. The main idea of Markov resampling proposed in this paper is to generate uniformly ergodic Markov chain (u.e.M.c.) examples for many times. In order to systematically study a Boosting algorithm based on Markov resampling, we prove that Boosting algorithm with general convex loss function based on u.e.M.c. examples are consistent and establish its fast convergence rate.

To improve the proposed SVM-BM, we also introduce another new SVM-Boosting algorithm based on Markov resampling, the improved SVM-Boosting based on Markov resampling (ISVM-BM). Different from SVM-BM, the weights of base learners of ISVM-BM are calculated using the support vectors.

II. General Terms Used

1. SVM :

SVM is a binary classification model developed by Vapnik from Structural Risk minimization theory. It uses the technique known as The Kernel trick in which it transforms the data and accordingly finds the optimal boundary between the possible outputs.

Reasons for SVMs being so important are -

- When a dataset is considered with a large number of features and small sample size, SVMs are very powerful.
- Using SVM, both simple and highly complex classification models can also be learned.
- SVM is helpful in avoiding the overfitting of curves by utilizing advanced mathematical principles.

2. Markov sampling :

In statistics, Markov chain Monte Carlo (MCMC) methods comprise a class of algorithms for sampling from a probability distribution. By constructing a Markov chain that has the desired distribution as its equilibrium distribution, one can obtain a sample of the desired distribution by recording states from the chain. The more steps are included, the more closely the distribution of the sample matches the actual desired distribution. Various algorithms exist for constructing chains, including the Metropolis–Hastings algorithm.

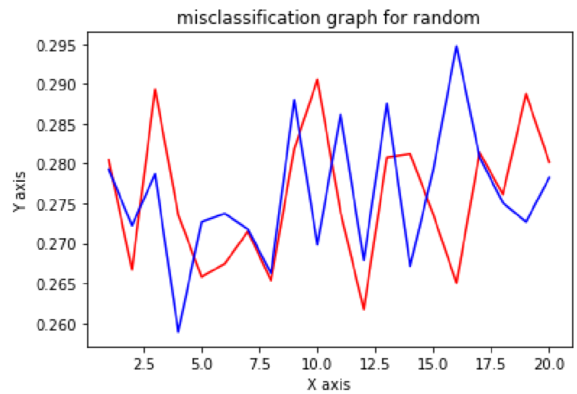
Markov Sampling is inspired from Markov Chain Monte Carlo methods. In statistics it comprises a class of algorithms for sampling from a probability distribution. If we construct a Markov chain with the desired distribution according to its equilibrium distribution, then we can obtain a sample of the desired distribution by recording states from the chain. The more steps are included, the more closely the distribution of the sample matches the actual desired distribution. Metropolis – Hastings algorithm is one of the various existing algorithms for construction of chains.

III. RESULTS AND OBSERVATIONS

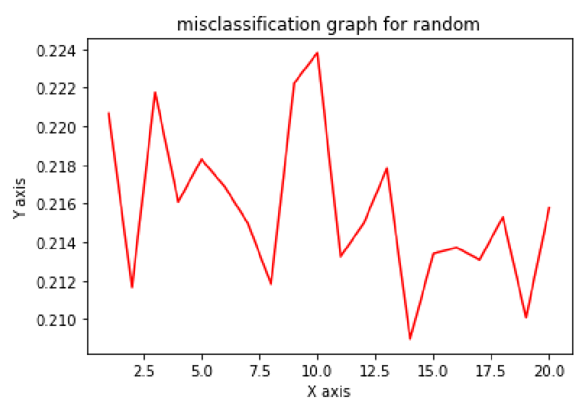
For Letter Dataset -

Kernel	**Accuracy**
Linear	0.81
RBF	0.82
X^2	0.87
Hellinger	0.77

Mean misclassification rate for random sample with std for letter - (27.576121193940306,0.8483597566911324)



mean misclassification rate for markov sample with std for letter - (27.604619769011553,0.8421835138019067)



For Pascal Dataset (worked only for 6 classes) -

Kernel	Accuracy
Linear	24.2738%
RBF	33.0705%
X^2	23.9834%
Hellinger	18.92116%

IV. CONCLUSION

To improve the learning execution of the traditional SVMC and the SVMC with Markov inspecting, this paper presents another SVMC calculation dependent on k-times Markov sampling(Algorithm 1) for the instance of adjusted preparing tests, and contrasted our calculation and the old style SVMC and the SVMC dependent on Markov examining. The exploratory outcomes demonstrated that the learning execution (the misclassification rates, the all out season of examining and preparing, and the quantities of help vector) of the SVMC with k-times (k = 1, 2) Markov testing is superior to that of the old style SVMC and the SVMC with Markov inspecting.