DMW C2 Assignment

Shubham Kumar IIT2018115
Raktim Bijoypuri IIT2018125
Aditya Kamble IIT2018126
Aakashdeep IIT2018128
Tejas Mane IIT2018135

VI Semester, Department of Information Technology,

Indian Institute of Information Technology, Allahabad, Prayagraj.

SVM-Boosting based on Markov resampling

Introduction:

In ensembled base learning, multiple weights are combined. Different base learners are also used which increases the overall accuracy. Bagging and boosting are major to sub types of this learning method. Boosting is to obtain base learners by adjusting the weights of training examples. One of the most famous types of boosting techniques is AdaBoost because it can converge to the Bayes risk as the size of training examples is big enough. Some data scientists studied the consistency of Boosting by minimizing the convex risk of classification error function. They showed that the Boosting algorithm is consistent under the condition of early-stopping strategies.

As the technology advances, the data generated is also increasing. Huge amount of data is generated each day. This big data also contains noise in it which is not a good sign for its processing. These noise examples not only lead to increased storage space, but also affect the accuracy of learning. The main idea of Markov resampling proposed in the paper is to generate uniformly ergodic Markov chains (u.e.M.c.) examples for many times. SVM-BM algorithm is time consuming as the size of the given training set is bigger. To improve the proposed SVM-BM, they also introduced another new SVM-Boosting algorithm based on Markov resampling, the improved SVM-Boosting based on Markov resampling (ISVM-BM). From this paper they showed that **"The Boosting algorithm with general convex loss function based on u.e.M.c. examples are proved to be consistent and its fast convergence rate is established."**

Uniformly Ergodic Markov Chains (u.e.M.c.) :

Let $\{Zt\}t \geq 1$ be a Markov chain, if there exist two constants $0 < \gamma 0 < \infty$ and $0 < \varphi < 1$ such that

d_{TV} $(Pm(\cdot|z), \pi(\cdot)) \leq \gamma 0 \varphi^m$, $\forall m \geq 1$, $m \in N$,

where $\pi(\cdot)$ is the stationary distribution of $\{Zt\}t \geq 1$. We say $\{Zt\}t \geq 1$ is uniformly ergodic.

Algorithms :

SVM with linear

kernel function is applied to Boosting algorithms to create two new Boosting algorithms:

1. SVM-Boosting based on Markov Resampling (SVM-BM) and
2. Improved SVM-Boosting based on Markov Resampling (ISVM-BM).

SVM-BM algorithm:

Draw randomly training set D_0 from D_{train} and obtain an initial classification function g0 by training SVM algorithm (8) with D0.

For $1 \leq t \leq T$, the transition probabilities $p^{i+1}{}_t$ used to generate the examples in Dt is based on the model $gt-1$,

If the loss $l(gt-1, zi)$ of the current examples zi is smaller, the acceptance probability p_t^{i+1} of the candidate examples z* will be smaller, which implies that generating the u.e.M.c examples in Dt will be time-consuming.

ISVM-BM algorithm:

For SVM algorithm with linear kernel function, the optimal base classification function gt can be expressed as

$\mathbf{g_t = \sum \omega_i y_i x_i{}' x + b, z_i = (x_i, y_i) \in D_t,}$

where x_i' is the transpose matrix of x_i. In (9), the vectors x_i that correspond to $\omega_i \neq 0$ are called to be support vector.

This implies that SVM has nice properties for compressing the training examples set in the form (9) of support vectors.

Result -

Kernel	Accuracy
linear	90.68
rbf	93.37
poly	93.04
hellinger	89.10
chi2_kernel	94.98