MAT 443 – HW 9

Eric Agyemang

Summary Report for Generalized Boosted Models:

A guide to the gbm package

Boosting is a class of ensemble learning techniques for regression and classification problems. Also, undertake several algorithmic task with help of some functions and models. There are numerous packages used to build gradient boosting in R.gbm package is extremely powerful tool that execute boosting models mainly used the field of statistics. The idea of gradient boosting originated from several observations that the boosting can be interpreted as an optimization algorithm. The gmb package inherits stochastic gradient boosting strategy but a little change on the basic operation in their algorithm. Gradient Boosting Machine was a new work proposed as result of connection between boosting and optimization. The work of Friedman, Hastie and Tibsharirani contributed in building and setting up the foundation of a new age of boosting Algorithm. These researchers invented Gradient Boosting by invent Adaboost thus the first successful boosting algorithm, formulate Adaboost as gradient descent with a special and finally generalize Adaboost to Gradient Boosting in order to handle a variety of loss functions. However there numerous way to improve the basic structure of this s to develop the model and make suitable for robust regression with small absolute deviation and loss function.

The boosting methods can upgraded and improved by first using control of the learning rate thus by decreasing the learning rate to reduce the intensity of the basic framework. Secondly, by reducing the variance factor with the help of subsampling so we can estimate the regression. Again by decomposing the anova this mostly applicable to boosted trees. An extension was developed by Friedman, which is known to be relative influencer, which involves tree-based methods. Most users of gbm may have have the chances to change the several options to fit their preference pf modelling.Some of these options that can be tuned by user are loss functions, the relationship between shrinkage and number of iterations and finally estimating the optimal number of iterations. Most new users of the gbm package find difficulty with the options of n.trees and shrinkage. It is vital to also know the smaller values of shrinkage always output better prediction and accurate results. Even though gbm provide three ways for estimating, the number of iterations after it has been fit to the model functions like gbm.perf computes the estimate of iterations. This method mentioned is mostly not known by new users of gbm.

For each level of class has several methods for mathematically calculating the deviance, initial value, gradient, constants for prediction and terminal node estimates. The different distributions listed in the table below are offered by gbm.

|  |  |  |  |
| --- | --- | --- | --- |
| **Distributions** | **Deviance** | **Initial Value** | **Gradient** |
| **Gaussian** |  |  |  |
| **AdaBoost** |  |  |  |
| **Bernoulli** |  |  |  |
| **Laplace** |  |  |  |
| **Quantile regression** |  |  |  |
| **Cox proportional hazard** |  |  |  |
| **Poisson** |  |  |  |

**NB:** ( is hazard for risk, is an indicator, *i* is number of observation, denotes weights)