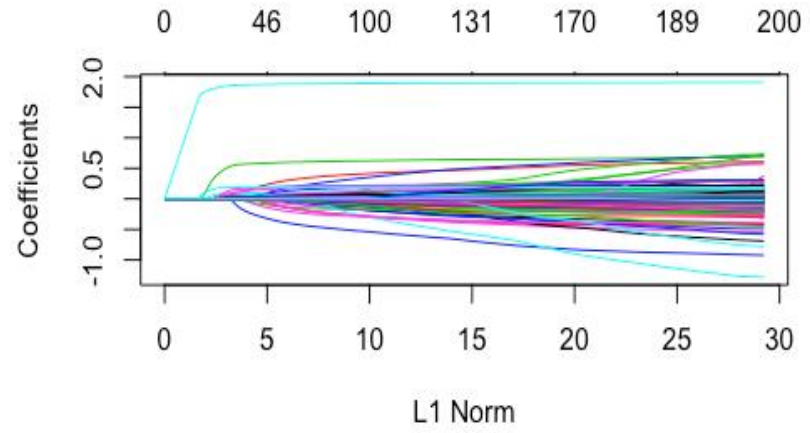
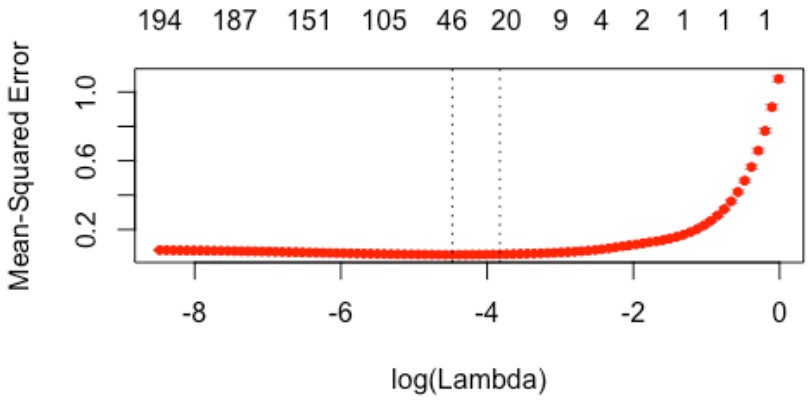
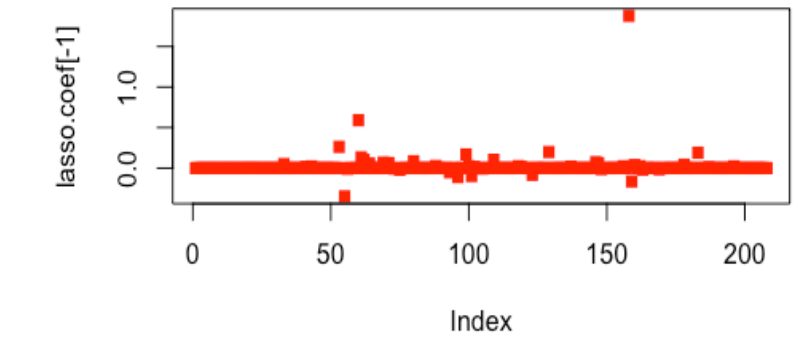
1. Lasso

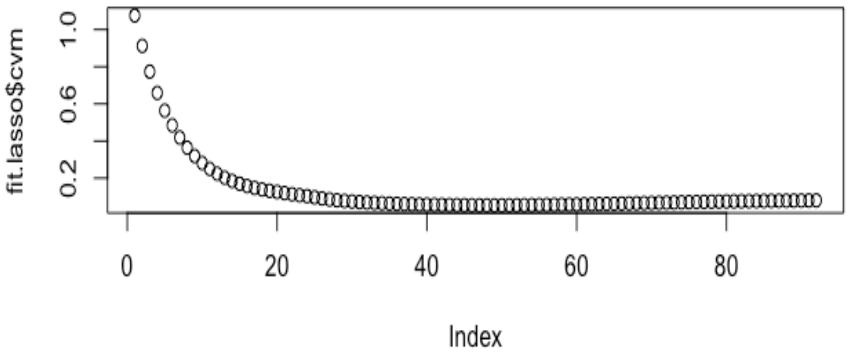
Lasso is also a regression analysis method that mot only performs regularization, but it also does variable selection. It estimates the coefficients based on the formula below:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$









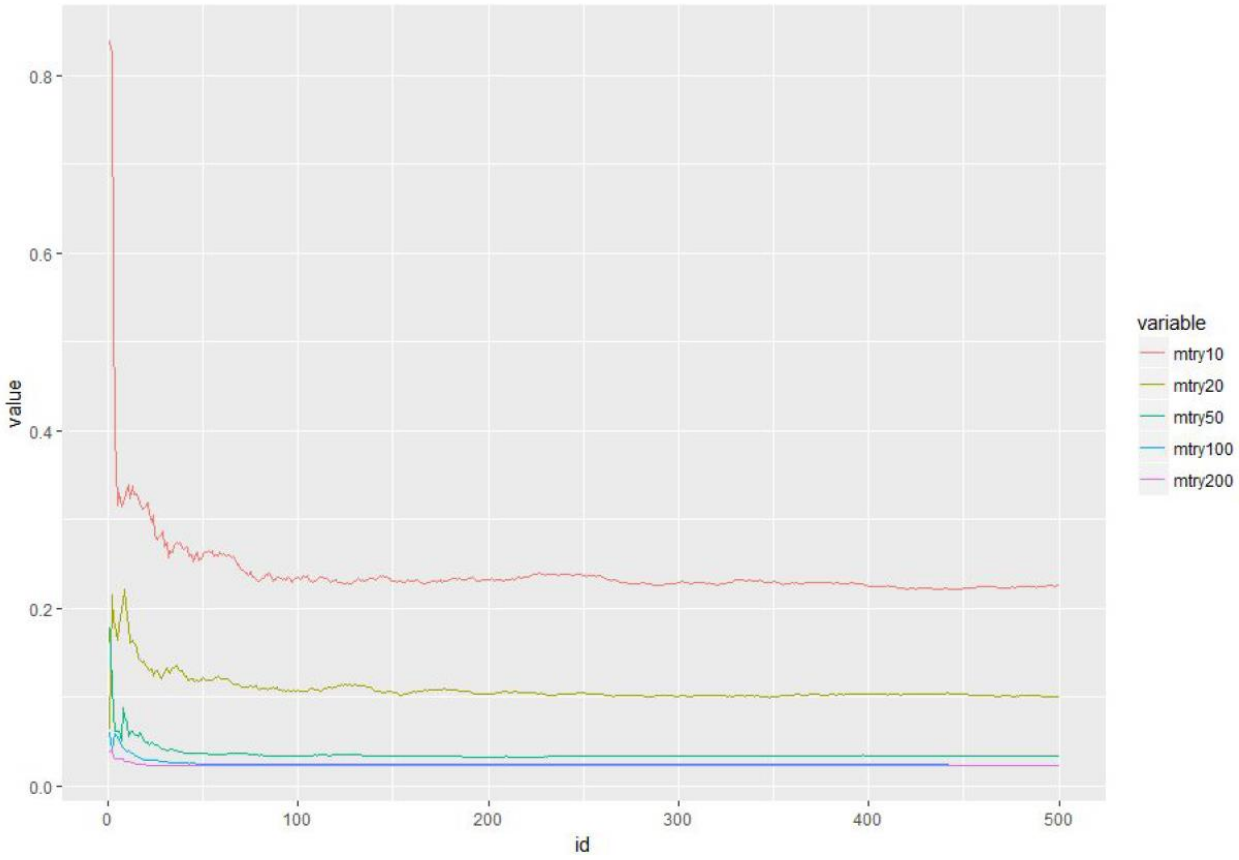
2. Random Forest

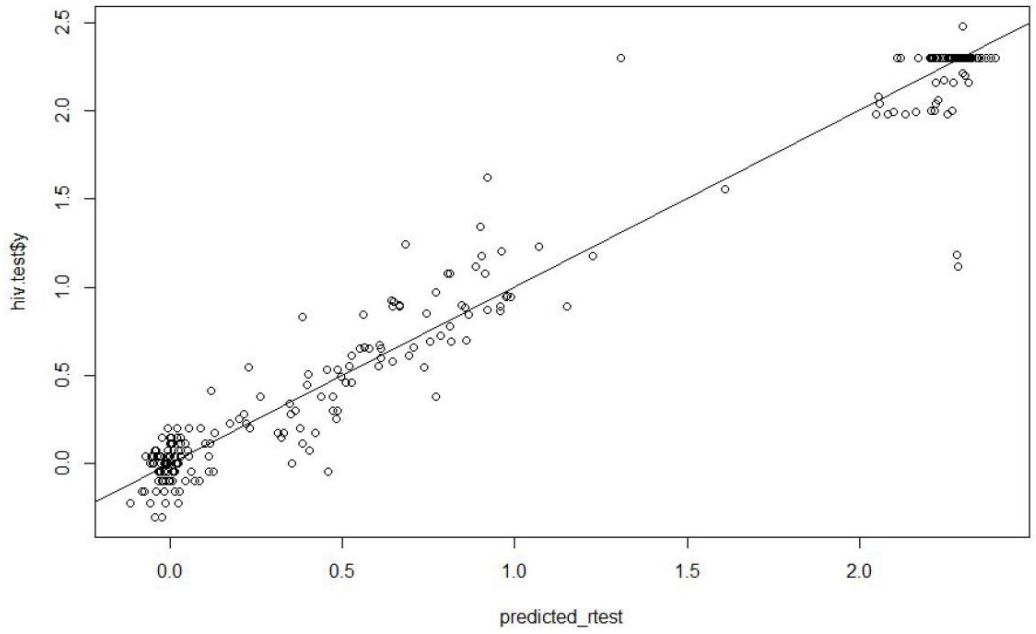
Random forests are an ensemble learning method for classification (and regression) that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees.

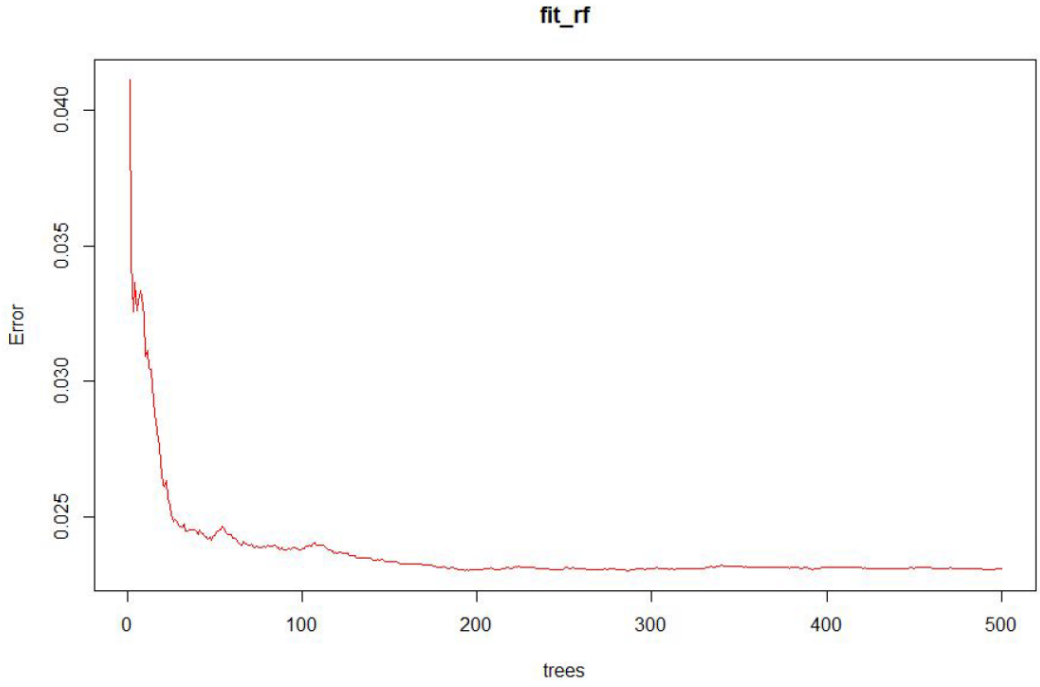
Bagging is a method of variance reduction for an estimated prediction function. It works especially well for high-variance, low-bias procedures. For regression, after getting bootstrap samples, fitting of the same regression tree to the bootstrap samples is needed. Afterwards, we get the average of the result.

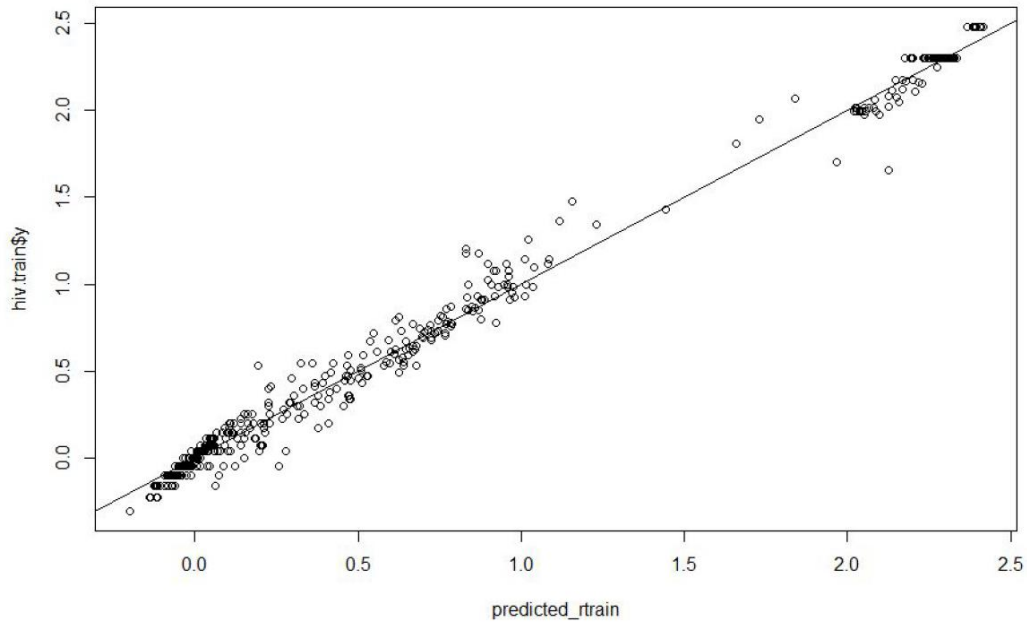
Random forests is a modification of bagging. It Builds a large collection of de-correlated trees, and gets their average. Random forests are more popular, because it is easier to train and tune in comparison to boosting.

In this case, the results of performing random forest are as follows:


value
id
variable
mtry10
mtry20
mtry50
mtry100
mtry200
0.0
0.2
0.4
0.6
0.8
0
100
200
300
400
500


hiv.test$y
predicted_rtest
0.0
0.5
1.0
1.5
2.0
2.5


fit_rf
Error
0.025
0.030
0.035
0.040
0
100
200
300
400
500
trees


hiv.train$y
0.0
0.5
1.0
1.5
2.0
2.5
predicted_rtrain