

IE 425

Homework 2 (due June 13)

1. Consider the “ToyotaCorolla” dataset. Our goal is to predict the Price attribute.
 - a. Partition the data set into training and test sets randomly with 70% going into the training set by using a seed value of 425.
 - b. By using the 10-fold cross-validation approach with `repeatcv=5` and playing with the parameters “`mtry`” within the range [1, number of input attributes] and “`ntree=100,200,300,400,500`”, try to obtain the best result in terms of RMSE and report the RMSE in the test set using the `randomForest` package.
 - c. Comment on which input attributes are important in making predictions.
 - d. Compare the performance with that of linear regression.
2. Repeat Question 1b with Gradient Boosting Machines by playing with the following parameters:
 - Number of trees (`n.trees`) = 75,100,125,150
 - Max Tree Depth (`interaction.depth`) = 3,4,5
 - Min. Leaf Node Size (`n.minobsinnode`) = 5,10,15
 - Shrinkage (`shrinkage`)= 0.1,0.2,0.3