CS 251/340 - Machine Learning

Spring 2019, AUA

Homework No. 05

Program files submission - Due date: 23:55, May 2, 2019

Go to https://rpubs.com/Hgoswami/368878 and by following the instructions there download german_credit data with the corresponding column names.

Assume that the output is german_credit$response with 2 classes "1" and "2". Assume that the positive class is response = "2".

Divide dataset into training (80%) and test sets (20%).

1. Apply classification trees (method = "rpart" in the Caret package) on the training set with 10-fold cross-validation. Show the confusion matrices and the ROC curves for the training and test sets. Draw the corresponding decision tree **(score = 30).**

2. Apply random forests (method = "rf") on the training set. Show the confusion matrices and the ROC curves for the training and test sets. **(score = 30).**

3. Apply boosting (method = "adaboost") of decision trees. Show the confusion matrices and the ROC curves for training and test sets **(score = 30).**

4. Compare the precisions on the test set and pick the best model for classification **(score = 10).**