Homework 5

1. Please find the attached divorce dataset and its description is available in the following link: https://archive.ics.uci.edu/ml/datasets/Divorce+Predictors+data+set#
   The dataset has one binary outcome variable (Class =1 for divorce, 0 for not divorce) and 54 predictor variables. The names of the predictor variables are available in the above link. The dataset has 170 observations.

   a. Please fit a classification tree and prune the tree using the cross-validation method. Provide the fitted tree diagram. Which predictors are significant for divorce?
   b. Please fit a random forest model and find the optimal number of trees using cross-validation method. Please provide the variable importance plot. Which predictors are significant?
   c. Please fit a boosting model and tune the parameters using cross validation method. For the optimal boosting model find the variables that are significant.
   d. Overlay the roc curves of the above three methods in a single plot. Identify, which method performs better and what is the best cut off point that we can consider for each of them.
   e. For each method and for best possible cut-off point, please compute the accuracy, sensitivity and specificity measures. What do they tell?