Lab 13

Exercise - BAGGING, BOOSTING, and RANDOM FORESTS

For this lab we will use the `Auto.rda` dataset. In addition we should use `library(randomForest)` and thus we have to install the `randomForest` package.

a) Display the output of the `randomForest` function when it is applied on the `Auto` dataset.

b) Measure the reduction of the node's impurity by using the function `importance`. What do you observe? Recall that node impurity is a measure of the homogeneity of the labels at the node. Two classical impurity measures for classification are Gini impurity and entropy and for regression is the variance. In addition, use `varImpPlot` to plot the results. What do you observe?

c) Apply cross-validation by using a training random subset of size 200 and calculate the mean-square error of prediction, estimated by the validation set cross-validation.

d) The next task is to search for the optimal solution. This implies that we have to estimate how many trees to grow? The default is 500, but we can decide based on the error.

e) Optimize both m and number of trees, by cross-validation.

f) Use the number of trees from previous part and construct an optimal random forest.