DATA 624 - Homework 9

OMER OZEREN

Table of Contents

Exercises 8.1,8.2,8.3 & 8.7 from the K&J book.

library(knitr)

## Warning: package 'knitr' was built under R version 3.5.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.3

library(tidyr)

## Warning: package 'tidyr' was built under R version 3.5.3

library(AppliedPredictiveModeling)

## Warning: package 'AppliedPredictiveModeling' was built under R version 3.5.3

library(partykit)

## Loading required package: grid

## Loading required package: libcoin

## Loading required package: mvtnorm

data(solubility)  
library(mlbench)

## Warning: package 'mlbench' was built under R version 3.5.3

library(ggplot2)  
library(mice)

## Warning: package 'mice' was built under R version 3.5.3

##   
## Attaching package: 'mice'

## The following objects are masked from 'package:base':  
##   
## cbind, rbind

library(caret)

## Warning: package 'caret' was built under R version 3.5.3

## Loading required package: lattice

library(Cubist)

## Warning: package 'Cubist' was built under R version 3.5.3

library(gbm)

## Warning: package 'gbm' was built under R version 3.5.3

## Loaded gbm 2.1.5

library(ipred)

## Warning: package 'ipred' was built under R version 3.5.3

library(party)

## Loading required package: modeltools

## Warning: package 'modeltools' was built under R version 3.5.3

## Loading required package: stats4

## Loading required package: strucchange

## Warning: package 'strucchange' was built under R version 3.5.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.5.3

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

##   
## Attaching package: 'party'

## The following objects are masked from 'package:partykit':  
##   
## cforest, ctree, ctree\_control, edge\_simple, mob, mob\_control,  
## node\_barplot, node\_bivplot, node\_boxplot, node\_inner, node\_surv,  
## node\_terminal, varimp

library(partykit)  
library(randomForest)

## Warning: package 'randomForest' was built under R version 3.5.3

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':  
##   
## margin

library(rpart)

## Warning: package 'rpart' was built under R version 3.5.3

library(vip)

## Warning: package 'vip' was built under R version 3.5.3

##   
## Attaching package: 'vip'

## The following object is masked from 'package:utils':  
##   
## vi

## 8.1 Recreate the simulated data from Exercise 7.2

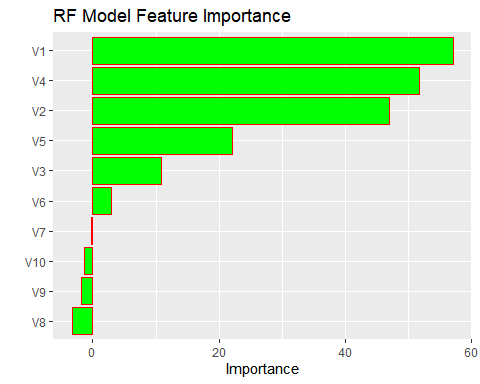
### Load data

library(mlbench)  
set.seed(200)  
simulated <- mlbench.friedman1(200, sd = 1)  
simulated <- cbind(simulated$x, simulated$y)  
simulated <- as.data.frame(simulated)  
colnames(simulated)[ncol(simulated)] <- "y"

### a) Fit a random forest model to all of the predictors, then estimate the variable importance scores: Did the random forest model significantly use the uninformative predictors ?

rf\_model <- randomForest(y ~ ., data = simulated,  
 importance = TRUE,  
 ntree = 1000)  
rf\_Imp1 <- rf\_model$importance   
vip(rf\_model, color = 'red', fill='green') +   
 ggtitle('RF Model Feature Importance')

## Warning in vip.default(rf\_model, color = "red", fill = "green"): Arguments  
## `width`, `alpha`, `color`, `fill`, `size`, and `shape` have all been deprecated  
## in favor of the new `mapping` and `aesthetics` arguments. They will be removed  
## in version 0.3.0.



The graph above shows feature importance that resulted from Random Forest. According to graph above, Tandom forest model didn’t use the uninformative predictors

### b) Now add an additional predictor that is highly correlated with one of the informative predictors. For example: