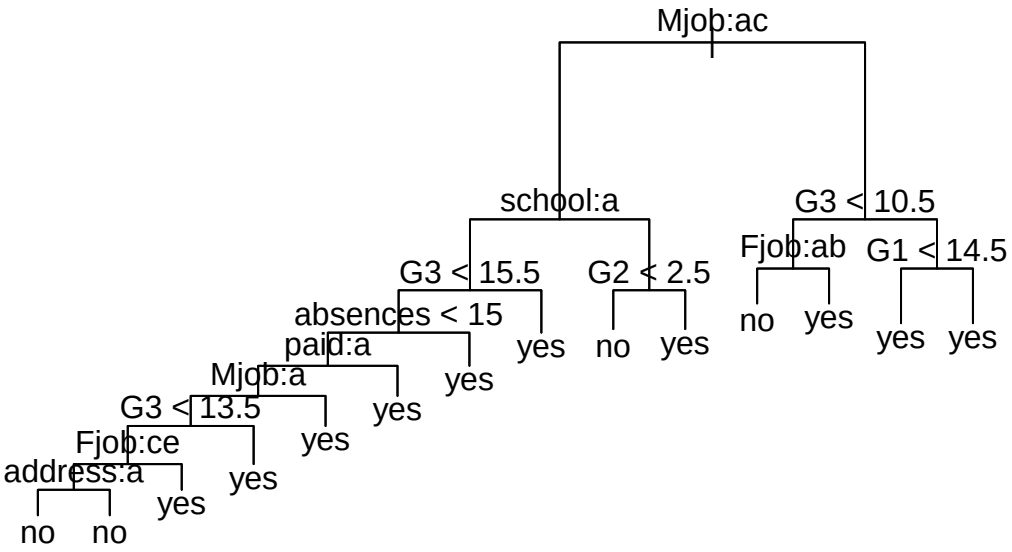
ML-2017-FALL-NOUL-PROJECT-RF

Noul Singla

November 14, 2017

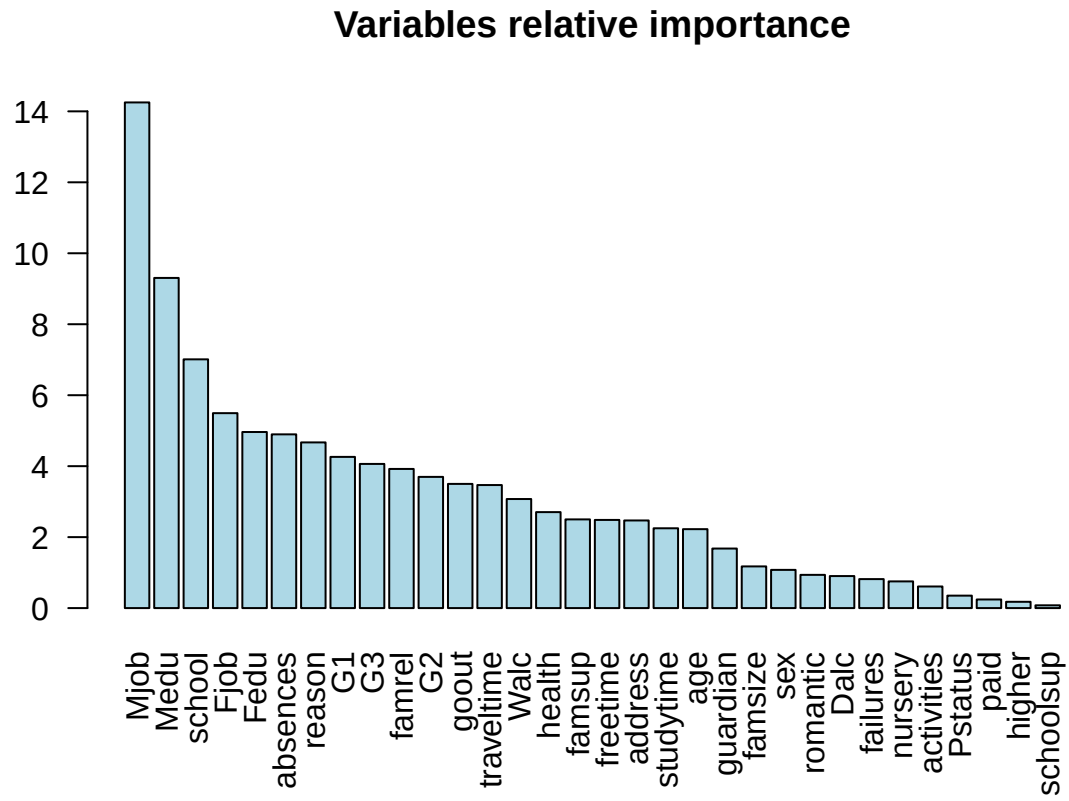
```
#implementation of a TREE

internet.tree1 <- tree(internet~.,Train.Student.Data)
plot(internet.tree1)
text(internet.tree1)
```




```
prd1 <- predict(internet.tree1,newdata =  Train.Student.Data,type="class")
table(prd1,Train.Student.Data$internet)
```

```
##
## prd1   no yes
##   no   25  10
##   yes  94 390
```

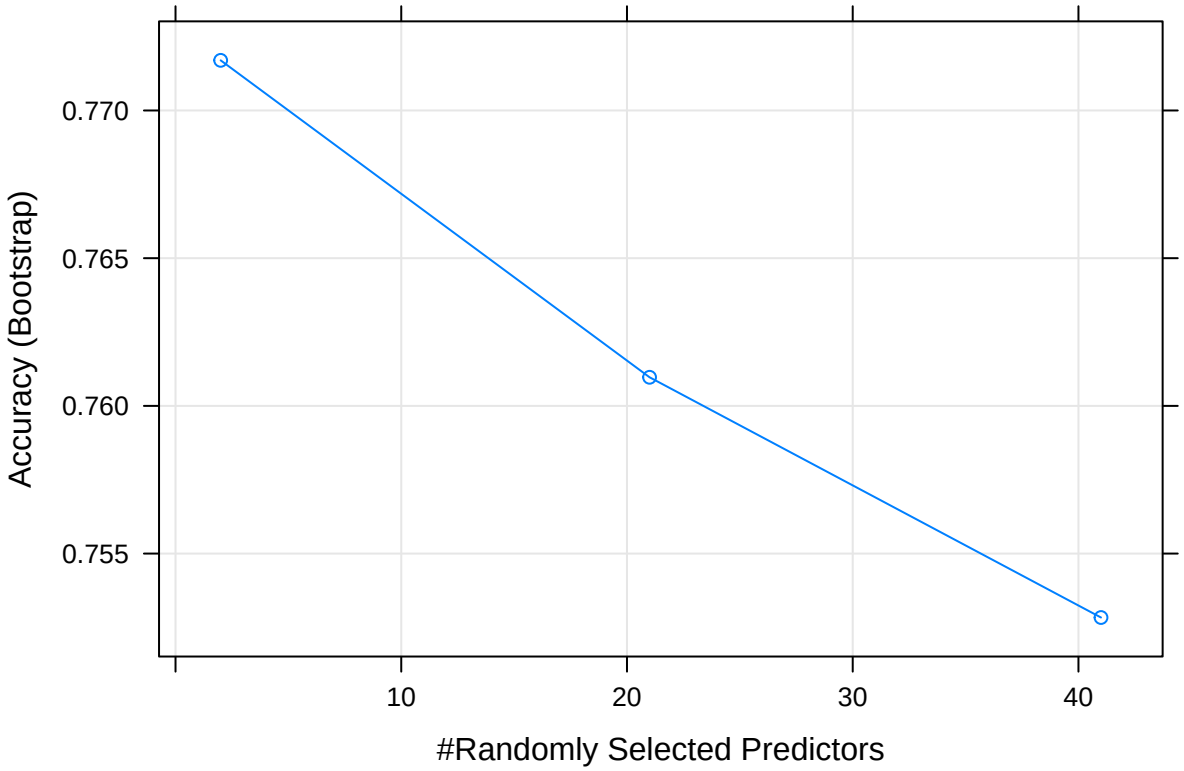
```
#implementation of a Bagging

#internet.bagging1 <- bagging(internet~.,data=Train.Student.Data,mfinal=50)
predict.train.bagging1 <- predict(internet.bagging1,newdata=Train.Student.Data)
importanceplot(internet.bagging1,las=2)
```




```
predict.train.bagging1$confusion
```

```
##                Observed Class
## Predicted Class  no yes
##             no   61   1
##             yes  58 399
```

```
#implementation of a Random Forest

#internet.rf <- train(internet~., data=Train.Student.Data,type="rf")


plot(internet.rf)
```




```
# str(internet.rf)
#prd <- predict(internet.rf,newdata=Train.Student.Data)


tcontrol = trainControl(method="repeatedcv",number=5,repeats=5)
#internet.rf1 <- train(internet~., data=Train.Student.Data,metric="Accuracy",type="rf",ntree=10, trCont



tcontrol = trainControl(method="repeatedcv",number=5,repeats=5,search="random")
#internet.rf2 <- train(internet~., data=Train.Student.Data,metric="Accuracy",type="rf",ntree=10, trCont

tcontrol = trainControl(method="repeatedcv",number=5,repeats=5,search="grid")
#internet.rf3 <- train(internet~., data=Train.Student.Data,metric="Accuracy",type="rf",ntree=10, trCont



tcontrol = trainControl(method="repeatedcv",number=10,repeats=4)
#internet.rf11 <- train(internet~., data=Train.Student.Data,metric="Accuracy",type="rf",ntree=20, trCon


tcontrol = trainControl(method="repeatedcv",number=20,repeats=10,search="random")
#internet.rf22 <- train(internet~., data=Train.Student.Data,metric="Accuracy",type="rf",ntree=20, trCon


tcontrol = trainControl(method="repeatedcv",number=20,repeats=10,search="grid")
```

```
#internet.rf33 <- train(internet~., data=Train.Student.Data,metric="Accuracy",type="rf",ntree=20, trCon


prd.trn <- predict(internet.rf,Train.Student.Data)
rf.table <- table(prd.trn,Train.Student.Data$internet)
rf.table
```

```
##
## prd.trn  no yes
##     no  105   0
##     yes  14 400
```

```
prd.trn1 <- predict(internet.rf1,Train.Student.Data)
rf1.table <- table(prd.trn1,Train.Student.Data$internet)
rf1.table
```

```
##
## prd.trn1  no yes
##      no   71   0
##      yes  48 400
```

```
prd.trn2 <- predict(internet.rf2,Train.Student.Data)
rf2.table <- table(prd.trn2,Train.Student.Data$internet)
rf2.table
```

```
##
## prd.trn2  no yes
##      no  106   0
##      yes  13 400
```

```
prd.trn3 <- predict(internet.rf3,Train.Student.Data)
rf3.table <- table(prd.trn3,Train.Student.Data$internet)
rf3.table
```

```
##
## prd.trn3  no yes
##      no   75   0
##      yes  44 400
```

```
prd.trn11 <- predict(internet.rf11,Train.Student.Data)
rf11.table <- table(prd.trn11,Train.Student.Data$internet)
rf11.table
```

```
##
## prd.trn11  no yes
##       no   91   0
##       yes  28 400
```

```
prd.trn22 <- predict(internet.rf22,Train.Student.Data)
rf22.table <- table(prd.trn22,Train.Student.Data$internet)
rf22.table
```

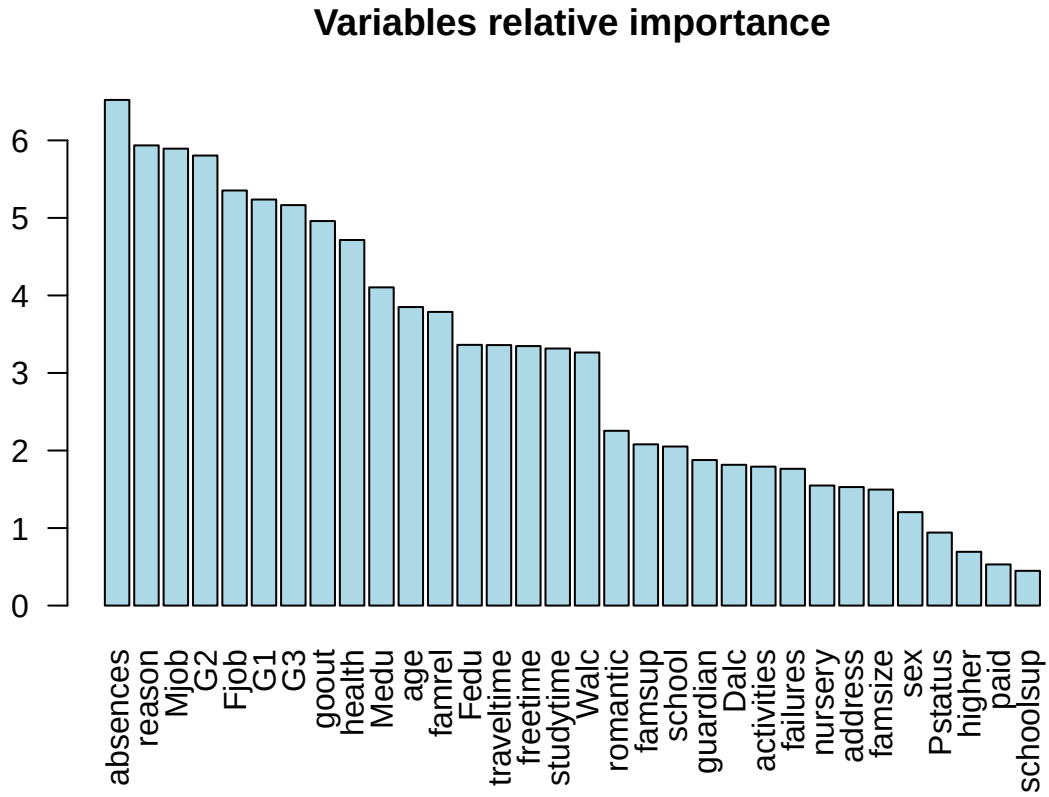
```
##
## prd.trn22  no yes
##       no  119   0
##       yes   0 400
```

```
prd.trn33 <- predict(internet.rf33,Train.Student.Data)
rf33.table <- table(prd.trn33,Train.Student.Data$internet)
```

```
rf33.table
```

```
##
## prd.trn33  no yes
##       no   79   0
##       yes  40 400
```

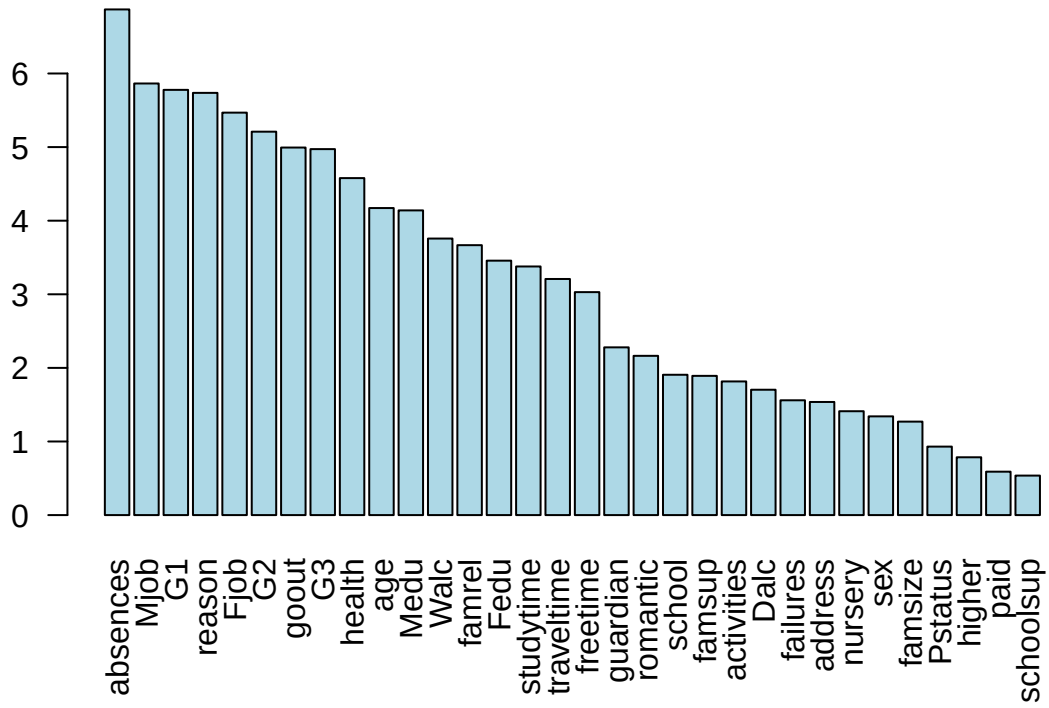
```
#implementation of a Adaboost


#internet.boost1 <- boosting(internet~.,data=Train.Student.Data,boos=TRUE,mfinal=100)
predict.train.boost1 <-predict(internet.boost1,newdata=Train.Student.Data)
importanceplot(internet.boost1,las=2)
```




```
predict.train.boost1$confusion
```

```
##                Observed Class
## Predicted Class  no yes
##             no  119   0
##             yes   0 400
```

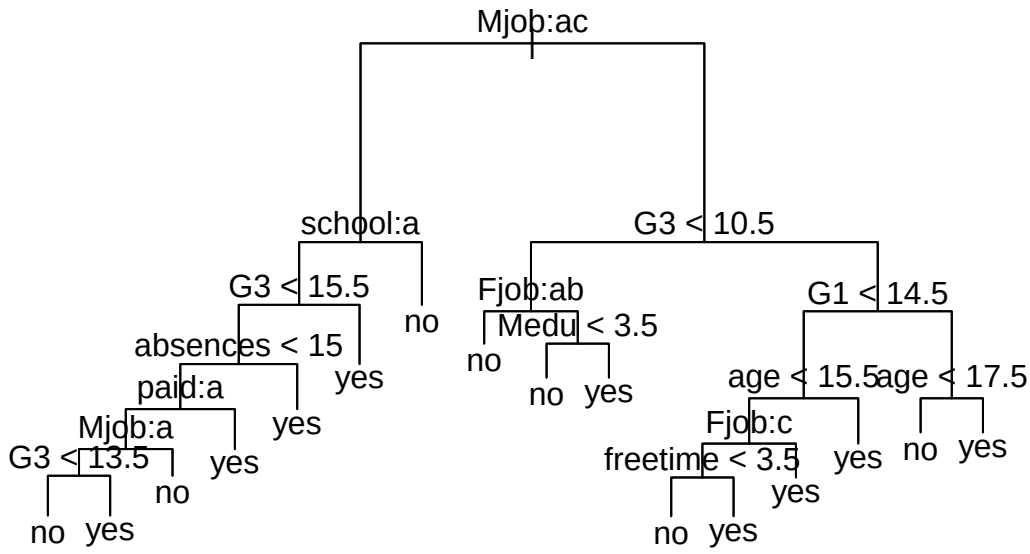
```
#internet.boost2 <- boosting(internet~.,data=Train.Student.Data,boos=TRUE,mfinal=500)
predict.train.boost2 <- predict(internet.boost2,newdata=Train.Student.Data)
importanceplot(internet.boost2,las=2)
```




```
predict.train.boost2$confusion
```

```
##                 Observed Class
## Predicted Class  no yes
##             no  119   0
##             yes   0 400
```

```
# TREE by training on increased dat for favourable set
train_data_no <- Train.Student.Data[Train.Student.Data$internet=="no",]
Train.Student.Data.biased <- rbind(Train.Student.Data,train_data_no,train_data_no,train_data_no,train_da
internet.tree2 <- tree(internet~.,Train.Student.Data.biased)
plot(internet.tree2)
text(internet.tree2)
```




```
prd1 <- predict(internet.tree2,newdata =  Train.Student.Data.biased,type="class")
table(prd1,Train.Student.Data.biased$internet)
```

```
##
## prd1   no yes
##   no  595 212
##   yes   0 188
```

```
#varying threshold

#threshold1 <- 0.4
#var.prd1.tree1 <- predict(internet.tree1,newdata =  Train.Student.Data)
#table(ifelse(var.prd1.tree1[,1]>threshold1,"no","yes"),Train.Student.Data$internet)

#var.prd1.tree2 <- predict(internet.tree2,newdata =  Train.Student.Data)
#table(ifelse(var.prd1.tree2[,1]>threshold1,"no","yes"),Train.Student.Data$internet)


#TEST DATA PREDICTIONS And ERROR CHECK

#TREE
prd.tst1 <- predict(internet.tree1,Test.Student.Data,type="class")
table(prd.tst1,Test.Student.Data$internet)
```

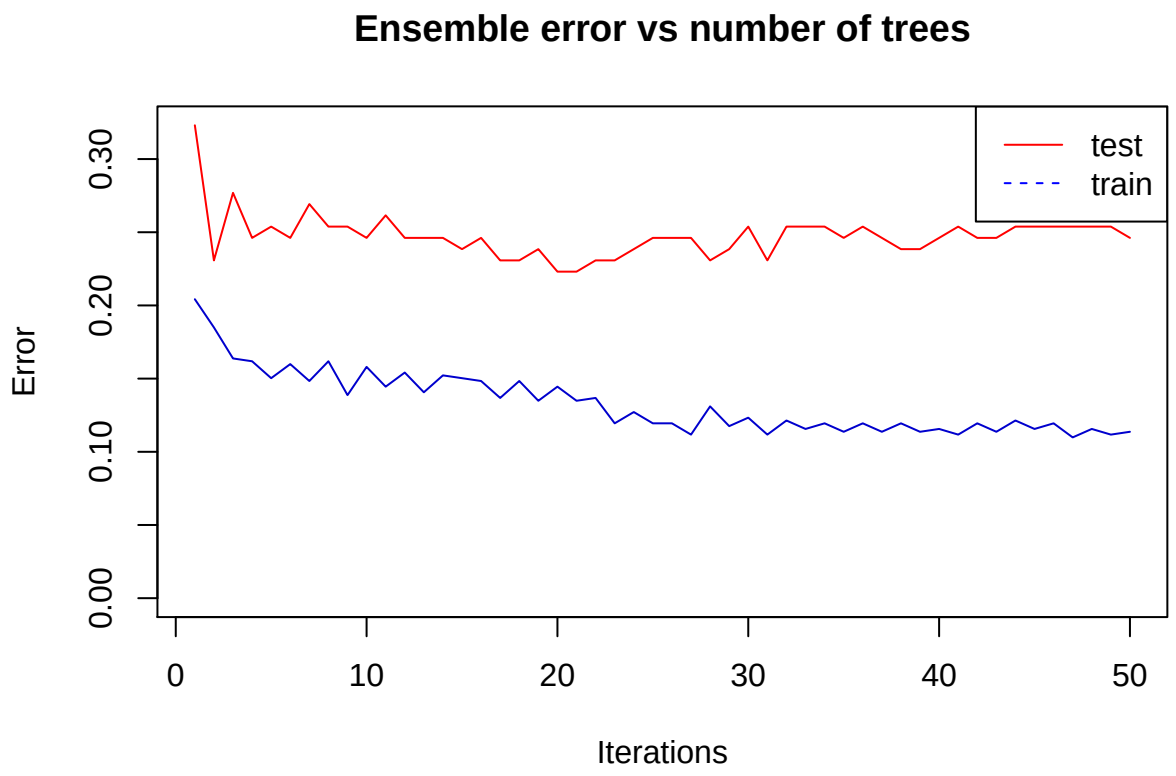
```
##
## prd.tst1 no yes
##      no   2   6
```

```
##      yes 30  92
```

```
#Baggin Best
predict.test.bagging1 <- predict(internet.bagging1,newdata=Test.Student.Data)
predict.test.bagging1$confusion
```

```
##                Observed Class
## Predicted Class no yes
##             no   3   3
##             yes 29  95
```

```
error1 <- errorevol(internet.bagging1,newdata=Train.Student.Data)
error2 <- errorevol(internet.bagging1,newdata=Test.Student.Data)
plot.errorevol(error2,error1)
```




```
#Random Forest Best

prd.tst22 <- predict(internet.rf22,Test.Student.Data)
rf22.test.table <- table(prd.tst22,Test.Student.Data$internet)
rf22.test.table
```

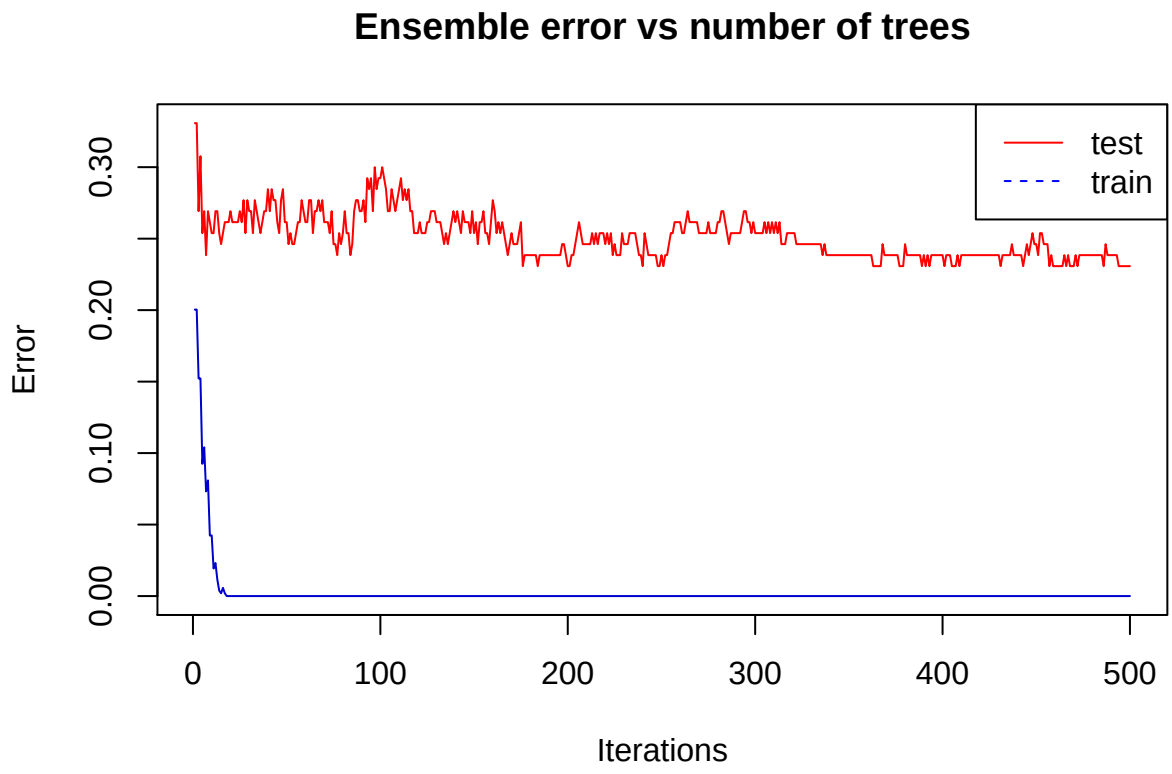
```
##
## prd.tst22 no yes
##       no   5   7
##       yes 27  91
```

```
# Boosting Best

predict.test.boost2 <- predict(internet.boost2,newdata=Test.Student.Data)
```

```
predict.test.boost2$confusion
```

```
##                 Observed Class
## Predicted Class no yes
##             no   8   9
##             yes 24  89
```

```
error21 <- errorevol(internet.boost1,newdata=Train.Student.Data)
error22 <- errorevol(internet.boost1,newdata=Test.Student.Data)
plot.errorevol(error22,error21)
```




```
#Biased tree
prd.tree2 <- predict(internet.tree2,Test.Student.Data,type="class")
table(prd.tree2,Test.Student.Data$internet)
```

```
##
## prd.tree2 no yes
##       no  24  65
##       yes  8  33
```

```
#Varying threshold for cutoff

#TREE
#threshold1 <- 0.4
#var.prd1.tree1 <- predict(internet.tree1,newdata =  Train.Student.Data)
#table(ifelse(var.prd1.tree1[,1]>threshold1,"no","yes"),Train.Student.Data$internet)
```

```
#var.prd1.tree2 <- predict(internet.tree2,newdata =  Train.Student.Data)
#table(ifelse(var.prd1.tree2[,1]>threshold1,"no","yes"),Train.Student.Data$internet)
```