setwd("C:/Users/Matrix/Desktop/New folder")

> filenameo <- "heart.csv"

> heartds <- read.csv(filenameo, header=TRUE)

> dim(heartds)

[1] 289 14

> heartds$output<- as.factor(heartds$output)

> sapply(heartds, class)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa

"integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "numeric" "integer" "integer"

thall output

"integer" "factor"

> head(heartds)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa thall output

1 60 1 3 145 233 1 0 150 0 2.3 0 0 1 yes

2 35 1 2 130 250 0 1 187 0 3.5 0 0 2 yes

3 41 0 1 130 204 0 0 172 0 1.4 2 0 2 yes

4 55 1 1 120 236 0 1 178 0 0.8 2 0 2 yes

5 56 0 0 120 354 0 1 163 1 0.6 2 0 2 yes

6 55 1 0 140 192 0 1 148 0 0.4 1 0 1 yes

> summary(heartds)

age sex cp trtbps chol fbs restecg

Min. :29.00 Min. :0.0000 Min. :0.000 Min. : 94.0 Min. :126 Min. :0.0000 Min. :0.0000

1st Qu.:47.00 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:120.0 1st Qu.:212 1st Qu.:0.0000 1st Qu.:0.0000

Median :54.00 Median :1.0000 Median :1.000 Median :130.0 Median :243 Median :0.0000 Median :1.0000

Mean :54.01 Mean :0.6782 Mean :1.021 Mean :131.4 Mean :248 Mean :0.1453 Mean :0.5156

3rd Qu.:60.00 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:140.0 3rd Qu.:276 3rd Qu.:0.0000 3rd Qu.:1.0000

Max. :77.00 Max. :1.0000 Max. :3.000 Max. :200.0 Max. :564 Max. :1.0000 Max. :2.0000

thalachh exng oldpeak slp caa thall output

Min. : 71.0 Min. :0.0000 Min. :0.000 Min. :0.000 Min. :0.0000 Min. :0.000 no :124

1st Qu.:136.0 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 yes:165

Median :154.0 Median :0.0000 Median :0.600 Median :1.000 Median :0.0000 Median :2.000

Mean :150.2 Mean :0.3183 Mean :1.008 Mean :1.419 Mean :0.7128 Mean :2.315

3rd Qu.:168.0 3rd Qu.:1.0000 3rd Qu.:1.600 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000

Max. :202.0 Max. :1.0000 Max. :6.200 Max. :2.000 Max. :4.0000 Max. :3.000

> filenametr <- "train.csv"

> trainds <- read.csv(filenametr, header=TRUE)

> dim(trainds)

[1] 231 14

> trainds$output<- as.factor(trainds$output)

> sapply(trainds, class)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa

"integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "numeric" "integer" "integer"

thall output

"integer" "factor"

> head(trainds)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa thall output

1 60 1 3 145 233 1 0 150 0 2.3 0 0 1 yes

2 35 1 2 130 250 0 1 187 0 3.5 0 0 2 yes

3 41 0 1 130 204 0 0 172 0 1.4 2 0 2 yes

4 56 0 0 120 354 0 1 163 1 0.6 2 0 2 yes

5 55 1 0 140 192 0 1 148 0 0.4 1 0 1 yes

6 56 0 1 140 294 0 0 153 0 1.3 1 0 2 yes

> summary(trainds)

age sex cp trtbps chol fbs restecg thalachh

Min. :29.00 Min. :0.0000 Min. :0 Min. : 94.0 Min. :126.0 Min. :0.0000 Min. :0.0000 Min. : 71.0

1st Qu.:47.00 1st Qu.:0.0000 1st Qu.:0 1st Qu.:120.0 1st Qu.:212.5 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:133.0

Median :54.00 Median :1.0000 Median :1 Median :130.0 Median :245.0 Median :0.0000 Median :1.0000 Median :155.0

Mean :54.28 Mean :0.6623 Mean :1 Mean :131.9 Mean :249.7 Mean :0.1472 Mean :0.5368 Mean :150.1

3rd Qu.:61.00 3rd Qu.:1.0000 3rd Qu.:2 3rd Qu.:140.0 3rd Qu.:279.5 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:168.0

Max. :77.00 Max. :1.0000 Max. :3 Max. :200.0 Max. :564.0 Max. :1.0000 Max. :2.0000 Max. :202.0

exng oldpeak slp caa thall output

Min. :0.000 Min. :0.0000 Min. :0.000 Min. :0.0000 Min. :0.000 no : 99

1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 yes:132

Median :0.000 Median :0.6000 Median :1.000 Median :0.0000 Median :2.000

Mean :0.316 Mean :0.9874 Mean :1.429 Mean :0.7056 Mean :2.338

3rd Qu.:1.000 3rd Qu.:1.6000 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000

Max. :1.000 Max. :5.6000 Max. :2.000 Max. :4.0000 Max. :3.000

> filenamete <- "test.csv"

> testds <- read.csv(filenamete, header=TRUE)

> dim(testds)

[1] 58 14

> testds$output<- as.factor(testds$output)

> sapply(testds, class)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa

"integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "numeric" "integer" "integer"

thall output

"integer" "factor"

> head(testds)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa thall output

1 55 1 1 120 236 0 1 178 0 0.8 2 0 2 yes

2 44 1 1 120 263 0 1 173 0 0.0 2 0 3 yes

3 48 0 2 130 275 0 1 139 0 0.2 2 0 2 yes

4 59 1 2 150 212 1 1 157 0 1.6 2 0 2 yes

5 51 1 2 110 175 0 1 123 0 0.6 2 0 2 yes

6 44 1 1 130 219 0 0 188 0 0.0 2 0 2 yes

> summary(testds)

age sex cp trtbps chol fbs restecg

Min. :34.00 Min. :0.0000 Min. :0.000 Min. :100.0 Min. :157.0 Min. :0.0000 Min. :0.000

1st Qu.:45.75 1st Qu.:0.2500 1st Qu.:0.000 1st Qu.:113.5 1st Qu.:209.0 1st Qu.:0.0000 1st Qu.:0.000

Median :55.00 Median :1.0000 Median :1.000 Median :125.5 Median :234.5 Median :0.0000 Median :0.000

Mean :52.95 Mean :0.7414 Mean :1.103 Mean :129.1 Mean :240.9 Mean :0.1379 Mean :0.431

3rd Qu.:59.00 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:138.0 3rd Qu.:264.5 3rd Qu.:0.0000 3rd Qu.:1.000

Max. :69.00 Max. :1.0000 Max. :3.000 Max. :178.0 Max. :409.0 Max. :1.0000 Max. :1.000

thalachh exng oldpeak slp caa thall output

Min. :103.0 Min. :0.0000 Min. :0.000 Min. :0.000 Min. :0.0000 Min. :0.000 no :25

1st Qu.:138.2 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 yes:33

Median :150.5 Median :0.0000 Median :0.700 Median :1.000 Median :0.0000 Median :2.000

Mean :150.6 Mean :0.3276 Mean :1.088 Mean :1.379 Mean :0.7414 Mean :2.224

3rd Qu.:164.5 3rd Qu.:1.0000 3rd Qu.:1.800 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000

Max. :188.0 Max. :1.0000 Max. :6.200 Max. :2.000 Max. :4.0000 Max. :3.000

>

> set.seed(100)

> library(randomForest)

randomForest 4.7-1.1

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

>

> rf <-randomForest(heartds$output~.,data=heartds, ntree=500)

> print(rf)

Call:

randomForest(formula = heartds$output ~ ., data = heartds, ntree = 500)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 3

OOB estimate of error rate: 17.3%

Confusion matrix:

no yes class.error

no 95 29 0.2338710

yes 21 144 0.1272727

>

>

> floor(sqrt(ncol(heartds) - 1))

[1] 3

> mtry <- tuneRF(heartds[-1],heartds$output, ntreeTry=500,stepFactor=1,improve=0.01, trace=TRUE, plot=TRUE)

mtry = 3 OOB error = 0%

Searching left ...

Searching right ...

> best.m <- mtry[mtry[, 2] == min(mtry[, 2]), 1]

> print(mtry)

mtry OOBError

3.OOB 3 0

>

> print(best.m)

[1] 3

> set.seed(100)

> rf <-randomForest(heartds$output~.,data=heartds, mtry=best.m, importance=TRUE,ntree=500)

> print(rf)

Call:

randomForest(formula = heartds$output ~ ., data = heartds, mtry = best.m, importance = TRUE, ntree = 500)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 3

OOB estimate of error rate: 16.61%

Confusion matrix:

no yes class.error

no 95 29 0.2338710

yes 19 146 0.1151515

>

>

> importance(rf)

no yes MeanDecreaseAccuracy MeanDecreaseGini

age 3.0332147 6.07481098 6.6139056 10.780584

sex 7.7347061 11.84416119 13.7804474 4.684511

cp 13.6402005 13.39843719 17.8730414 15.369008

trtbps 1.0128984 2.32477791 2.3608199 10.382716

chol 0.8144661 0.08720068 0.6714353 11.495593

fbs -1.4119054 2.70298539 1.1685071 1.480708

restecg 3.0626355 1.47159227 3.2691549 3.034816

thalachh 6.9705091 11.44954799 12.9060856 15.760843

exng 11.6738313 6.97114955 13.1882406 8.950842

oldpeak 14.8269689 10.31602538 17.5010945 16.784000

slp 8.9079202 2.53694272 8.2965137 6.955709

caa 19.3223328 20.88372330 25.9938211 17.991061

thall 14.4121200 18.15190839 21.2297917 16.021400

> varImpPlot(rf)

> pred1=predict(rf,type = "prob")

> library(ROCR)

> perf = prediction(pred1[,2], heartds$output)

> auc = performance(perf, "auc")

> auc

A performance instance

'Area under the ROC curve'

>

> pred3 = performance(perf, "tpr","fpr")

> plot(pred3,main="ROC Curve for Random Forest",col=2,lwd=2)

> abline(a=0,b=1,lwd=2,lty=2,col="gray")

> plot(rf)

>