Project 2 Chandler Thompson

Chandler Thompson

2/28/2021

Library necessary packages

library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

library(mlbench)

Data

#load the mlbench package which has the BreastCancer data set require(mlbench)

# if you don’t have any required package, use the install.packages() command

# load the data set

data(BreastCancer)

Read in our data

data("BreastCancer")  
  
# some algorithms don't like missing values, so remove rows with missing values  
BreastCancer <- na.omit(BreastCancer)  
  
# remove the unique identifier, which is useless and would confuse the machine learning algorithms  
BreastCancer$Id <- NULL   
  
  
# partition the data set for 80% training and 20% evaluation (adapted from ?randomForest)  
set.seed(2)  
  
SampleIndex <- sample(2, nrow(BreastCancer), replace = TRUE, prob=c(0.8, 0.2))

Basic summary exploration of our data

BreastCancer <- cbind(BreastCancer[10],BreastCancer[1:9])  
  
str(BreastCancer)

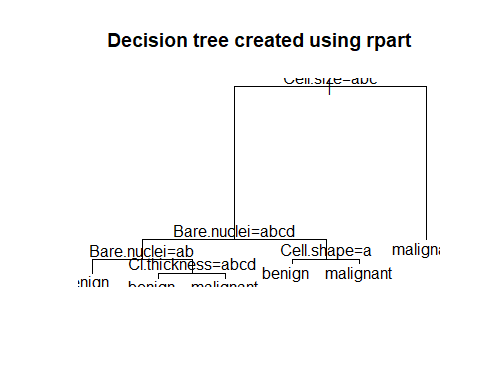
## 'data.frame': 683 obs. of 10 variables:  
## $ Class : Factor w/ 2 levels "benign","malignant": 1 1 1 1 1 2 1 1 1 1 ...  
## $ Cl.thickness : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 5 5 3 6 4 8 1 2 2 4 ...  
## $ Cell.size : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 1 4 1 8 1 10 1 1 1 2 ...  
## $ Cell.shape : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 1 4 1 8 1 10 1 2 1 1 ...  
## $ Marg.adhesion : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 1 5 1 1 3 8 1 1 1 1 ...  
## $ Epith.c.size : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 2 7 2 3 2 7 2 2 2 2 ...  
## $ Bare.nuclei : Factor w/ 10 levels "1","2","3","4",..: 1 10 2 4 1 10 10 1 1 1 ...  
## $ Bl.cromatin : Factor w/ 10 levels "1","2","3","4",..: 3 3 3 3 3 9 3 3 1 2 ...  
## $ Normal.nucleoli: Factor w/ 10 levels "1","2","3","4",..: 1 2 1 7 1 7 1 1 1 1 ...  
## $ Mitoses : Factor w/ 9 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 5 1 ...

summary(BreastCancer)

## Class Cl.thickness Cell.size Cell.shape Marg.adhesion  
## benign :444 1 :139 1 :373 1 :346 1 :393   
## malignant:239 5 :128 10 : 67 2 : 58 2 : 58   
## 3 :104 3 : 52 10 : 58 3 : 58   
## 4 : 79 2 : 45 3 : 53 10 : 55   
## 10 : 69 4 : 38 4 : 43 4 : 33   
## 2 : 50 5 : 30 5 : 32 8 : 25   
## (Other):114 (Other): 78 (Other): 93 (Other): 61   
## Epith.c.size Bare.nuclei Bl.cromatin Normal.nucleoli Mitoses   
## 2 :376 1 :402 3 :161 1 :432 1 :563   
## 3 : 71 10 :132 2 :160 10 : 60 2 : 35   
## 4 : 48 2 : 30 1 :150 3 : 42 3 : 33   
## 1 : 44 5 : 30 7 : 71 2 : 36 10 : 14   
## 6 : 40 3 : 28 4 : 39 8 : 23 4 : 12   
## 5 : 39 8 : 21 5 : 34 6 : 22 7 : 9   
## (Other): 65 (Other): 40 (Other): 68 (Other): 68 (Other): 17

Recursive Partitioning

# create model using recursive partitioning on the training data set  
library(rpart)  
x.rp <- rpart(Class ~ ., data=BreastCancer[SampleIndex == 1,])  
# predict classes for the evaluation data set  
x.rp.pred <- predict(x.rp, type="class", newdata=BreastCancer[SampleIndex == 2,])  
# score the evaluation data set (extract the probabilities)  
x.rp.prob <- predict(x.rp, type="prob", newdata=BreastCancer[SampleIndex == 2,])  
  
#plot.new()  
  
plot(x.rp, main="Decision tree created using rpart") ; text(x.rp)



Conditional Inference Tree

library(party)

## Warning: package 'party' was built under R version 4.0.4

## Loading required package: grid

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Warning: package 'strucchange' was built under R version 4.0.4

## Loading required package: zoo

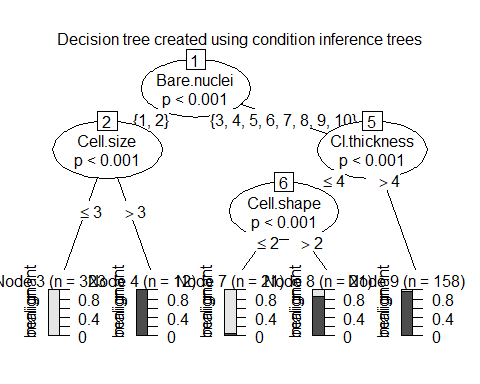
##   
## Attaching package: 'zoo'

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

## Loading required package: sandwich

## Warning: package 'sandwich' was built under R version 4.0.4

x.ct <- ctree(Class ~ ., data=BreastCancer[SampleIndex == 1,])  
x.ct.pred <- predict(x.ct, newdata=BreastCancer[SampleIndex == 2,])  
x.ct.prob <- 1- unlist(treeresponse(x.ct, BreastCancer[SampleIndex == 2,]), use.names=F)[seq(1,nrow(BreastCancer[SampleIndex == 2,])\*2,2)]  
  
plot(x.ct, main="Decision tree created using condition inference trees")



Random Forest

# create model using random forest and bagging ensemble using conditional inference trees  
x.cf <- cforest(Class ~ ., data=BreastCancer[SampleIndex == 1,], control = cforest\_unbiased(mtry = ncol(BreastCancer)-2))  
x.cf.pred <- predict(x.cf, newdata=BreastCancer[SampleIndex == 2,])  
x.cf.prob <- 1- unlist(treeresponse(x.cf, BreastCancer[SampleIndex == 2,]), use.names=F)[seq(1,nrow(BreastCancer[SampleIndex == 2,])\*2,2)]

Bagging Ensemble

# create model using bagging (bootstrap aggregating)  
library(ipred)

## Warning: package 'ipred' was built under R version 4.0.4

x.ip <- bagging(Class ~ ., data=BreastCancer[SampleIndex == 1,])  
x.ip.prob <- predict(x.ip, type="prob", newdata=BreastCancer[SampleIndex == 2,])

Support Vector Machine

# create model using svm (support vector machine)  
library(e1071)  
  
# svm requires tuning  
x.svm.tune <- tune(svm, Class~., data = BreastCancer[SampleIndex == 1,],  
 ranges = list(gamma = 2^(-8:1), cost = 2^(0:4)),  
 tunecontrol = tune.control(sampling = "fix"))  
  
# display the tuning results (in text format)  
x.svm.tune

##   
## Parameter tuning of 'svm':  
##   
## - sampling method: fixed training/validation set   
##   
## - best parameters:  
## gamma cost  
## 0.0625 1  
##   
## - best performance: 0.02234637

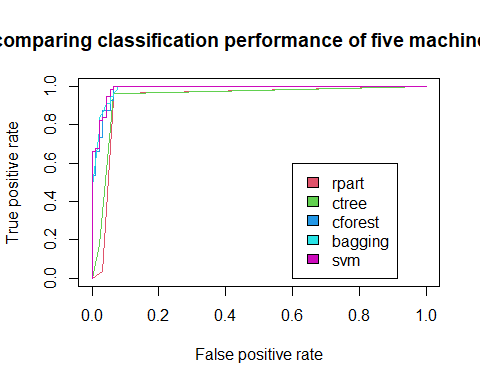
# If the tuning results are on the margin of the parameters (e.g., gamma = 2^-8),   
# then widen the parameters.  
# I manually copied the cost and gamma from console messages above to parameters below.  
x.svm <- svm(Class~., data = BreastCancer[SampleIndex == 1,], cost=4, gamma=0.0625, probability = TRUE)  
x.svm.prob <- predict(x.svm, type="prob", newdata=BreastCancer[SampleIndex == 2,], probability = TRUE)

Plot ROC curves to compare our classifiers

##  
## plot ROC curves to compare the performance of the individual classifiers  
##  
  
# Output the plot to a PNG file for display on web. To draw to the screen,   
# comment this line out.  
# png(filename="roc\_curve\_5\_models.png", width=700, height=700)  
  
  
  
# load the ROCR package which draws the ROC curves  
library(ROCR)

## Warning: package 'ROCR' was built under R version 4.0.4

# create an ROCR prediction object from rpart() probabilities  
x.rp.prob.rocr <- prediction(x.rp.prob[,2], BreastCancer[SampleIndex == 2,'Class'])  
  
# prepare an ROCR performance object for ROC curve (tpr=true positive rate, fpr=false positive rate)  
x.rp.perf <- performance(x.rp.prob.rocr, "tpr","fpr")  
  
# plot it  
plot(x.rp.perf, col=2, main="ROC curves comparing classification performance of five machine learning models")  
  
# Draw a legend.  
legend(0.6, 0.6, c('rpart', 'ctree', 'cforest','bagging','svm'), 2:6)  
  
  
# ctree  
x.ct.prob.rocr <- prediction(x.ct.prob, BreastCancer[SampleIndex == 2,'Class'])  
x.ct.perf <- performance(x.ct.prob.rocr, "tpr","fpr")  
  
# add=TRUE draws on the existing chart   
plot(x.ct.perf, col=3, add=TRUE)  
  
  
# cforest  
x.cf.prob.rocr <- prediction(x.cf.prob, BreastCancer[SampleIndex == 2,'Class'])  
x.cf.perf <- performance(x.cf.prob.rocr, "tpr","fpr")  
  
plot(x.cf.perf, col=4, add=TRUE)  
  
# bagging  
x.ip.prob.rocr <- prediction(x.ip.prob[,2], BreastCancer[SampleIndex == 2,'Class'])  
x.ip.perf <- performance(x.ip.prob.rocr, "tpr","fpr")  
  
plot(x.ip.perf, col=5, add=TRUE)  
  
# svm  
x.svm.prob.rocr <- prediction(attr(x.svm.prob, "probabilities")[,2], BreastCancer[SampleIndex == 2,'Class'])  
x.svm.perf <- performance(x.svm.prob.rocr, "tpr","fpr")  
plot(x.svm.perf, col=6, add=TRUE)



# Close and save the PNG file.  
#dev.off()

Bagging and SVM appear to perform the best.

#### Compare classifiers from provided R file

SVM

library(e1071)  
mysvm <- svm(Class ~ ., BreastCancer)  
mysvm.pred <- predict(mysvm, BreastCancer)  
table(mysvm.pred,BreastCancer$Class)

##   
## mysvm.pred benign malignant  
## benign 431 8  
## malignant 13 231

length(mysvm.pred)

## [1] 683

length(BreastCancer$Class)

## [1] 683

Naive Bayes

library(klaR)

## Loading required package: MASS

mynb <- NaiveBayes(Class ~ ., BreastCancer)  
mynb.pred <- predict(mynb,BreastCancer)

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 2

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 4

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 6

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 13

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 15

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 16

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 21

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 25

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 32

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 38

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 39

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 40

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 42

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 45

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 48

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 49

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 50

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 51

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 52

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 54

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 55

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 56

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 58

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 59

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 62

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 64

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 67

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 70

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 73

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 83

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 84

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 85

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 86

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 97

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 98

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 99

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 100

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 102

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 103

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 104

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 106

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 108

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 109

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 110

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 115

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 116

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 123

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 125

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 127

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 140

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 143

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 145

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 146

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 148

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 155

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 156

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 162

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 168

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 170

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 172

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 181

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 191

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 195

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 200

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 201

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 210

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 213

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 218

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 222

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 225

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 226

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 227

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 230

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 231

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 232

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 241

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 245

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 252

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 254

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 257

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 261

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 263

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 266

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 271

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 274

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 276

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 280

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 281

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 284

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 286

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 288

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 291

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 293

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 297

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 301

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 305

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 307

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 308

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 313

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 317

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 321

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 331

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 335

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 336

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 339

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 340

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 343

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 345

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 346

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 348

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 368

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 372

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 387

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 398

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 400

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 401

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 402

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 404

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 413

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 420

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 422

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 439

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 441

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 443

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 451

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 452

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 453

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 468

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 474

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 475

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 480

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 492

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 505

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 516

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 535

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 554

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 557

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 560

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 574

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 576

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 577

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 590

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 591

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 606

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 611

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 618

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 621

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 633

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 642

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 643

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 653

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 655

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 681

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 682

## Warning in FUN(X[[i]], ...): Numerical 0 probability for all classes with  
## observation 683

table(mynb.pred$class,BreastCancer$Class)

##   
## benign malignant  
## benign 431 3  
## malignant 13 236

Neural Net

library(nnet)  
mynnet <- nnet(Class ~ ., BreastCancer, size=1)

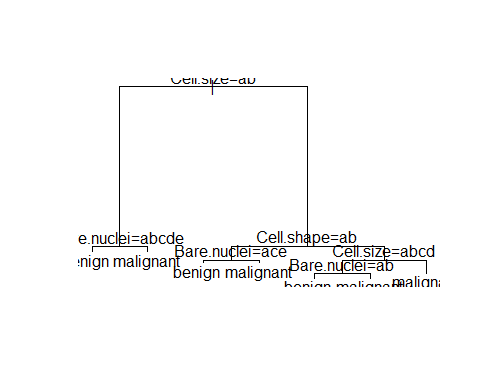
## # weights: 83  
## initial value 481.900570   
## iter 10 value 100.570572  
## iter 20 value 62.177408  
## iter 30 value 51.366553  
## iter 40 value 46.261994  
## iter 50 value 41.028270  
## iter 60 value 40.963399  
## iter 70 value 35.310481  
## iter 80 value 35.261793  
## iter 90 value 35.249971  
## iter 100 value 35.234052  
## final value 35.234052   
## stopped after 100 iterations

mynnet.pred <- predict(mynnet,BreastCancer,type="class")  
table(mynnet.pred,BreastCancer$Class)

##   
## mynnet.pred benign malignant  
## benign 438 1  
## malignant 6 238

Decision Tree

#Decision trees  
library(rpart)  
mytree <- rpart(Class ~ ., BreastCancer)  
plot(mytree); text(mytree) # in "iris\_tree.ps"



summary(mytree)

## Call:  
## rpart(formula = Class ~ ., data = BreastCancer)  
## n= 683   
##   
## CP nsplit rel error xerror xstd  
## 1 0.79079498 0 1.00000000 1.0000000 0.05215335  
## 2 0.05439331 1 0.20920502 0.2175732 0.02900067  
## 3 0.02510460 2 0.15481172 0.1757322 0.02626911  
## 4 0.01255230 3 0.12970711 0.1631799 0.02537272  
## 5 0.01000000 6 0.09205021 0.1631799 0.02537272  
##   
## Variable importance  
## Cell.size Cell.shape Bare.nuclei Epith.c.size Bl.cromatin   
## 21 18 16 15 14   
## Normal.nucleoli Cl.thickness   
## 14 1   
##   
## Node number 1: 683 observations, complexity param=0.790795  
## predicted class=benign expected loss=0.3499268 P(node) =1  
## class counts: 444 239  
## probabilities: 0.650 0.350   
## left son=2 (418 obs) right son=3 (265 obs)  
## Primary splits:  
## Cell.size splits as LLRRRRRRRR, improve=222.3221, (0 missing)  
## Cell.shape splits as LLLRRRRRRR, improve=216.4111, (0 missing)  
## Bare.nuclei splits as LLRRRRRRRR, improve=203.7284, (0 missing)  
## Bl.cromatin splits as LLLRRRRRRR, improve=196.3903, (0 missing)  
## Epith.c.size splits as LLRRRRRRRR, improve=193.1310, (0 missing)  
## Surrogate splits:  
## Cell.shape splits as LLLRRRRRRR, agree=0.917, adj=0.785, (0 split)  
## Epith.c.size splits as LLRRRRRRRR, agree=0.900, adj=0.743, (0 split)  
## Bare.nuclei splits as LLRRRRRRRR, agree=0.880, adj=0.691, (0 split)  
## Normal.nucleoli splits as LLRRRRRRRR, agree=0.877, adj=0.683, (0 split)  
## Bl.cromatin splits as LLLRRRRRRR, agree=0.876, adj=0.679, (0 split)  
##   
## Node number 2: 418 observations, complexity param=0.0251046  
## predicted class=benign expected loss=0.02870813 P(node) =0.6120059  
## class counts: 406 12  
## probabilities: 0.971 0.029   
## left son=4 (410 obs) right son=5 (8 obs)  
## Primary splits:  
## Bare.nuclei splits as LLLLLRRR-R, improve=11.68296, (0 missing)  
## Normal.nucleoli splits as LLLR-RRL-R, improve=11.68296, (0 missing)  
## Cl.thickness splits as LLLLLLRRRR, improve=10.32214, (0 missing)  
## Bl.cromatin splits as LLLLR-R---, improve= 8.53307, (0 missing)  
## Epith.c.size splits as LLLRRRRRRR, improve= 4.63208, (0 missing)  
## Surrogate splits:  
## Cl.thickness splits as LLLLLLLLRR, agree=0.988, adj=0.375, (0 split)  
## Normal.nucleoli splits as LLLR-RRL-L, agree=0.988, adj=0.375, (0 split)  
## Mitoses splits as LLRLL-LL-, agree=0.983, adj=0.125, (0 split)  
##   
## Node number 3: 265 observations, complexity param=0.05439331  
## predicted class=malignant expected loss=0.1433962 P(node) =0.3879941  
## class counts: 38 227  
## probabilities: 0.143 0.857   
## left son=6 (23 obs) right son=7 (242 obs)  
## Primary splits:  
## Cell.shape splits as LLRRRRRRRR, improve=20.58158, (0 missing)  
## Cell.size splits as LLLRRRRRRR, improve=18.27650, (0 missing)  
## Bare.nuclei splits as LRRRRRRRRR, improve=16.81493, (0 missing)  
## Bl.cromatin splits as LLRRRRRRRR, improve=13.91034, (0 missing)  
## Marg.adhesion splits as LLRRRRRRRR, improve=11.17148, (0 missing)  
## Surrogate splits:  
## Bl.cromatin splits as LRRRRRRRRR, agree=0.932, adj=0.217, (0 split)  
##   
## Node number 4: 410 observations  
## predicted class=benign expected loss=0.01219512 P(node) =0.6002928  
## class counts: 405 5  
## probabilities: 0.988 0.012   
##   
## Node number 5: 8 observations  
## predicted class=malignant expected loss=0.125 P(node) =0.01171303  
## class counts: 1 7  
## probabilities: 0.125 0.875   
##   
## Node number 6: 23 observations, complexity param=0.0125523  
## predicted class=benign expected loss=0.2173913 P(node) =0.03367496  
## class counts: 18 5  
## probabilities: 0.783 0.217   
## left son=12 (16 obs) right son=13 (7 obs)  
## Primary splits:  
## Bare.nuclei splits as LRLRL----R, improve=4.968944, (0 missing)  
## Bl.cromatin splits as LLLRR-RR--, improve=4.968944, (0 missing)  
## Cl.thickness splits as LLLLRRRRRR, improve=3.381643, (0 missing)  
## Epith.c.size splits as LLRRRRRRRR, improve=1.992754, (0 missing)  
## Cell.shape splits as LRRRRRRRRR, improve=1.397516, (0 missing)  
## Surrogate splits:  
## Bl.cromatin splits as LLLRR-RR--, agree=0.913, adj=0.714, (0 split)  
## Cl.thickness splits as LLLLLRRRRR, agree=0.870, adj=0.571, (0 split)  
## Mitoses splits as LRLR----R, agree=0.870, adj=0.571, (0 split)  
## Marg.adhesion splits as LLLLLLLRRR, agree=0.826, adj=0.429, (0 split)  
## Normal.nucleoli splits as LLRRLL-L--, agree=0.826, adj=0.429, (0 split)  
##   
## Node number 7: 242 observations, complexity param=0.0125523  
## predicted class=malignant expected loss=0.08264463 P(node) =0.3543192  
## class counts: 20 222  
## probabilities: 0.083 0.917   
## left son=14 (68 obs) right son=15 (174 obs)  
## Primary splits:  
## Cell.size splits as LLLLRRRRRR, improve=5.297663, (0 missing)  
## Bare.nuclei splits as LLRRRRRRRR, improve=4.093695, (0 missing)  
## Cell.shape splits as LLLLRRRRRR, improve=2.958548, (0 missing)  
## Bl.cromatin splits as LLLLRLRRRR, improve=2.838336, (0 missing)  
## Marg.adhesion splits as LLLLLRRRRR, improve=2.754821, (0 missing)  
## Surrogate splits:  
## Cell.shape splits as LLLLRRRRRR, agree=0.789, adj=0.250, (0 split)  
## Epith.c.size splits as LLRRRRRRRR, agree=0.777, adj=0.206, (0 split)  
## Marg.adhesion splits as LRRRRRRRRR, agree=0.744, adj=0.088, (0 split)  
## Bl.cromatin splits as LLRRRRRRRR, agree=0.736, adj=0.059, (0 split)  
## Bare.nuclei splits as RRRRRRLRRR, agree=0.723, adj=0.015, (0 split)  
##   
## Node number 12: 16 observations  
## predicted class=benign expected loss=0 P(node) =0.02342606  
## class counts: 16 0  
## probabilities: 1.000 0.000   
##   
## Node number 13: 7 observations  
## predicted class=malignant expected loss=0.2857143 P(node) =0.0102489  
## class counts: 2 5  
## probabilities: 0.286 0.714   
##   
## Node number 14: 68 observations, complexity param=0.0125523  
## predicted class=malignant expected loss=0.25 P(node) =0.09956076  
## class counts: 17 51  
## probabilities: 0.250 0.750   
## left son=28 (14 obs) right son=29 (54 obs)  
## Primary splits:  
## Bare.nuclei splits as LLRRR-RRRR, improve=7.600529, (0 missing)  
## Cl.thickness splits as LLLLLLRRRR, improve=3.558824, (0 missing)  
## Normal.nucleoli splits as LLRRRLLLRR, improve=2.951389, (0 missing)  
## Marg.adhesion splits as LLLLLRRRRR, improve=2.615385, (0 missing)  
## Bl.cromatin splits as LLLLRLLR-R, improve=1.640351, (0 missing)  
##   
## Node number 15: 174 observations  
## predicted class=malignant expected loss=0.01724138 P(node) =0.2547584  
## class counts: 3 171  
## probabilities: 0.017 0.983   
##   
## Node number 28: 14 observations  
## predicted class=benign expected loss=0.2857143 P(node) =0.0204978  
## class counts: 10 4  
## probabilities: 0.714 0.286   
##   
## Node number 29: 54 observations  
## predicted class=malignant expected loss=0.1296296 P(node) =0.07906296  
## class counts: 7 47  
## probabilities: 0.130 0.870

mytree.pred <- predict(mytree,BreastCancer,type="class")  
table(mytree.pred,BreastCancer$Class)

##   
## mytree.pred benign malignant  
## benign 431 9  
## malignant 13 230

Leave 1 Out Cross Validation

# Leave-1-Out Cross Validation (LOOCV)  
ans <- numeric(length(BreastCancer[,1]))  
for (i in 1:length(BreastCancer[,1])) {  
 mytree <- rpart(Class ~ ., BreastCancer[-i,])  
 mytree.pred <- predict(mytree,BreastCancer[i,],type="class")  
 ans[i] <- mytree.pred  
}  
ans <- factor(ans,labels=levels(BreastCancer$Class))  
table(ans,BreastCancer$Class)

##   
## ans benign malignant  
## benign 430 20  
## malignant 14 219

# The same as above in this case

Quadratic Discriminant Analysis

#Quadratic Discriminant Analysis  
library(MASS)  
  
BreastCancerInts <- BreastCancer  
  
BreastCancerInts$Cl.thickness <- as.integer(BreastCancerInts$Cl.thickness)  
BreastCancerInts$Cl.thickness <- as.integer(BreastCancerInts$Cl.thickness)  
BreastCancerInts$Cell.size <- as.integer(BreastCancerInts$Cell.size)  
BreastCancerInts$Cell.shape <- as.integer(BreastCancerInts$Cell.shape)  
BreastCancerInts$Marg.adhesion <- as.integer(BreastCancerInts$Marg.adhesion)  
BreastCancerInts$Epith.c.size <- as.integer(BreastCancerInts$Epith.c.size)  
BreastCancerInts$Bare.nuclei <- as.integer(BreastCancerInts$Bare.nuclei)  
BreastCancerInts$Bl.cromatin <- as.integer(BreastCancerInts$Bl.cromatin)  
BreastCancerInts$Normal.nucleoli <- as.integer(BreastCancerInts$Normal.nucleoli)  
BreastCancerInts$Mitoses <- as.integer(BreastCancerInts$Mitoses)  
  
myqda <- qda(Class ~ ., BreastCancerInts)  
myqda.pred <- predict(myqda, BreastCancerInts)  
table(myqda.pred$class,BreastCancerInts$Class)

##   
## benign malignant  
## benign 422 6  
## malignant 22 233

Regularized Discriminant Analysis

#Regularised Discriminant Analysis  
library(klaR)  
myrda <- rda(Class ~ ., BreastCancer)  
myrda.pred <- predict(myrda, BreastCancer)  
table(myrda.pred$class,BreastCancer$Class)

##   
## benign malignant  
## benign 433 3  
## malignant 11 236

Random Forests

#Random Forests  
library(randomForest)

## 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

myrf <- randomForest(Class ~ .,BreastCancer)  
myrf.pred <- predict(myrf, BreastCancer)  
table(myrf.pred, BreastCancer$Class)

##   
## myrf.pred benign malignant  
## benign 444 0  
## malignant 0 239

Ensemble

Now we will combine our different predictions and use a ‘majority vote’ between the predictions to assign our final prediction.

combine.classes<-data.frame(myrf.pred, myrda.pred$class,myqda.pred,  
mytree.pred,mynnet.pred,mysvm.pred, mynb.pred$class)  
  
combine.classes$myrf.pred<-ifelse(combine.classes$myrf.pred=="benign", 0, 1)  
combine.classes[,2]<-ifelse(combine.classes[,2]=="benign", 0, 1)  
combine.classes[,3]<-ifelse(combine.classes[,3]=="benign", 0, 1)  
combine.classes[,4]<-ifelse(combine.classes[,4]=="benign", 0, 1)  
combine.classes[,5]<-ifelse(combine.classes[,5]=="benign", 0, 1)  
combine.classes[,6]<-ifelse(combine.classes[,6]=="benign", 0, 1)  
combine.classes[,7]<-ifelse(combine.classes[,7]=="benign", 0, 1)  
combine.classes[,8]<-ifelse(combine.classes[,8]=="benign", 0, 1)  
combine.classes[,9]<-ifelse(combine.classes[,9]=="benign", 0, 1)  
  
combine.classes$MajorityVote <- rowSums(combine.classes)  
  
combine.classes$MajorityVote <-ifelse(combine.classes$MajorityVote >= 5, "malignant", "benign")  
  
table(combine.classes$MajorityVote, BreastCancer$Class)

##   
## benign malignant  
## benign 431 1  
## malignant 13 238

confusionMatrix(as.factor(combine.classes$MajorityVote), BreastCancer$Class)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction benign malignant  
## benign 431 1  
## malignant 13 238  
##   
## Accuracy : 0.9795   
## 95% CI : (0.9658, 0.9887)  
## No Information Rate : 0.6501   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.9555   
##   
## Mcnemar's Test P-Value : 0.003283   
##   
## Sensitivity : 0.9707   
## Specificity : 0.9958   
## Pos Pred Value : 0.9977   
## Neg Pred Value : 0.9482   
## Prevalence : 0.6501   
## Detection Rate : 0.6310   
## Detection Prevalence : 0.6325   
## Balanced Accuracy : 0.9833   
##   
## 'Positive' Class : benign   
##

We can see that with our ensemble method we achieve a 97.8% accuracy with similarly high sensitivity and specificity.