library(mlbench)

## Warning: package 'mlbench' was built under R version 4.2.3

library(e1071)

## Warning: package 'e1071' was built under R version 4.2.2

library(klaR)

## Warning: package 'klaR' was built under R version 4.2.2

## Loading required package: MASS

library(nnet)  
library(MASS)  
library(rpart)  
library(randomForest)

## Warning: package 'randomForest' was built under R version 4.2.2

## randomForest 4.7-1.1

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

library(caret)

## Warning: package 'caret' was built under R version 4.2.2

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.2.2

##   
## Attaching package: 'ggplot2'

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

## Loading required package: lattice

#Load the Data

library(mlbench)  
data(BreastCancer)  
BreastCancer <- na.omit(BreastCancer)   
  
BreastCancer$Id <- NULL   
  
df2 <- data.frame(sapply(BreastCancer[1:9], function(x) as.numeric(as.character(x))))  
z <- scale(df2[,1:9],center=TRUE,scale=TRUE)  
  
  
set.seed(2)  
ind <- createDataPartition(BreastCancer$Class, p = 0.6, list = FALSE)  
breastCance.train <- BreastCancer[ind,]  
breastCance.test <- BreastCancer[-ind,]

#SVM

mysvm <- svm(Class ~ ., breastCance.train)  
mysvm.pred <- predict(mysvm, breastCance.test)  
table(mysvm.pred,breastCance.test$Class)

##   
## mysvm.pred benign malignant  
## benign 174 7  
## malignant 3 88

#Naive Bayes

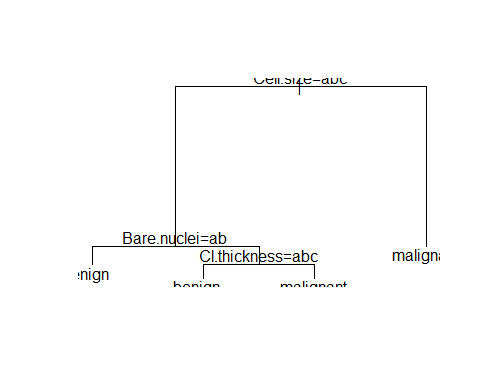
library(klaR)  
mynb <- NaiveBayes(Class ~ ., breastCance.train, usekernel = TRUE)  
mynb.pred <- predict(mynb,breastCance.test)

#head(mynb.pred$class)  
table(mynb.pred$class,breastCance.test$Class)

##   
## benign malignant  
## benign 174 2  
## malignant 3 93

#Decision trees

mytree <- rpart(Class ~ ., breastCance.train)  
plot(mytree); text(mytree)



summary(mytree)

## Call:  
## rpart(formula = Class ~ ., data = breastCance.train)  
## n= 411   
##   
## CP nsplit rel error xerror xstd  
## 1 0.80555556 0 1.00000000 1.0000000 0.06716662  
## 2 0.05555556 1 0.19444444 0.2638889 0.04078146  
## 3 0.01000000 3 0.08333333 0.1875000 0.03487901  
##   
## Variable importance  
## Cell.size Cell.shape Bare.nuclei Bl.cromatin Normal.nucleoli   
## 20 17 17 15 14   
## Marg.adhesion Cl.thickness Epith.c.size Mitoses   
## 13 2 1 1   
##   
## Node number 1: 411 observations, complexity param=0.8055556  
## predicted class=benign expected loss=0.350365 P(node) =1  
## class counts: 267 144  
## probabilities: 0.650 0.350   
## left son=2 (277 obs) right son=3 (134 obs)  
## Primary splits:  
## Cell.size splits as LLLRRRRRRR, improve=134.9103, (0 missing)  
## Bare.nuclei splits as LLRRRRRRRR, improve=131.2616, (0 missing)  
## Cell.shape splits as LLRRRRRRRR, improve=129.3126, (0 missing)  
## Bl.cromatin splits as LLLRRRRRRR, improve=119.7576, (0 missing)  
## Normal.nucleoli splits as LLRRRRRRRR, improve=118.1358, (0 missing)  
## Surrogate splits:  
## Cell.shape splits as LLLRRRRRRR, agree=0.942, adj=0.821, (0 split)  
## Bare.nuclei splits as LLLRRRRRRR, agree=0.905, adj=0.709, (0 split)  
## Bl.cromatin splits as LLLRRRRRRR, agree=0.895, adj=0.679, (0 split)  
## Normal.nucleoli splits as LLRRRRRRRR, agree=0.888, adj=0.657, (0 split)  
## Marg.adhesion splits as LLLRRRRRRR, agree=0.881, adj=0.634, (0 split)  
##   
## Node number 2: 277 observations, complexity param=0.05555556  
## predicted class=benign expected loss=0.06859206 P(node) =0.6739659  
## class counts: 258 19  
## probabilities: 0.931 0.069   
## left son=4 (247 obs) right son=5 (30 obs)  
## Primary splits:  
## Bare.nuclei splits as LLRRRR-RRR, improve=19.00160, (0 missing)  
## Normal.nucleoli splits as LLRRLR-LRR, improve=18.21888, (0 missing)  
## Cl.thickness splits as LLLLLLRRRR, improve=15.16591, (0 missing)  
## Cell.shape splits as LLRRRRRRRR, improve=11.30634, (0 missing)  
## Epith.c.size splits as LLLRRRRRRR, improve=10.62647, (0 missing)  
## Surrogate splits:  
## Normal.nucleoli splits as LLRRRL-LLR, agree=0.924, adj=0.300, (0 split)  
## Cl.thickness splits as LLLLLLRRRR, agree=0.921, adj=0.267, (0 split)  
## Bl.cromatin splits as LLLLRLLR--, agree=0.921, adj=0.267, (0 split)  
## Mitoses splits as LLRRL-RR-, agree=0.917, adj=0.233, (0 split)  
## Marg.adhesion splits as LLLRRRRRRR, agree=0.910, adj=0.167, (0 split)  
##   
## Node number 3: 134 observations  
## predicted class=malignant expected loss=0.06716418 P(node) =0.3260341  
## class counts: 9 125  
## probabilities: 0.067 0.933   
##   
## Node number 4: 247 observations  
## predicted class=benign expected loss=0.004048583 P(node) =0.6009732  
## class counts: 246 1  
## probabilities: 0.996 0.004   
##   
## Node number 5: 30 observations, complexity param=0.05555556  
## predicted class=malignant expected loss=0.4 P(node) =0.0729927  
## class counts: 12 18  
## probabilities: 0.400 0.600   
## left son=10 (12 obs) right son=11 (18 obs)  
## Primary splits:  
## Cl.thickness splits as LLLRRRRRRR, improve=10.677780, (0 missing)  
## Normal.nucleoli splits as LRRRLR--RR, improve= 8.400000, (0 missing)  
## Cell.shape splits as LLRRRRRRRR, improve= 7.810714, (0 missing)  
## Cell.size splits as LRRRRRRRRR, improve= 6.074641, (0 missing)  
## Bl.cromatin splits as LLRRR-RR--, improve= 3.816149, (0 missing)  
## Surrogate splits:  
## Cell.shape splits as LRRRRRRRRR, agree=0.867, adj=0.667, (0 split)  
## Cell.size splits as LRRRRRRRRR, agree=0.833, adj=0.583, (0 split)  
## Bl.cromatin splits as LLRLR-RR--, agree=0.800, adj=0.500, (0 split)  
## Normal.nucleoli splits as LRRRRR--RR, agree=0.800, adj=0.500, (0 split)  
## Epith.c.size splits as LLRRRRRRRR, agree=0.767, adj=0.417, (0 split)  
##   
## Node number 10: 12 observations  
## predicted class=benign expected loss=0.08333333 P(node) =0.02919708  
## class counts: 11 1  
## probabilities: 0.917 0.083   
##   
## Node number 11: 18 observations  
## predicted class=malignant expected loss=0.05555556 P(node) =0.04379562  
## class counts: 1 17  
## probabilities: 0.056 0.944

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

##   
## mytree.pred benign malignant  
## benign 170 5  
## malignant 7 90

# 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

#Regularised Discriminant Analysis

myrda <- rda(Class ~ ., breastCance.train)  
myrda.pred <- predict(myrda, breastCance.test)  
table(myrda.pred$class,breastCance.test$Class)

##   
## benign malignant  
## benign 174 3  
## malignant 3 92

#Random Forests

myrf <- randomForest(Class ~ ., breastCance.train)  
myrf.pred <- predict(myrf, breastCance.test)  
head(myrf.pred)

## 5 11 12 13 15 20   
## benign benign benign malignant malignant benign   
## Levels: benign malignant

table(myrf.pred, breastCance.test$Class)

##   
## myrf.pred benign malignant  
## benign 174 3  
## malignant 3 92

combine.classes<-data.frame(myrf.pred, myrda.pred$class,   
 mytree.pred,mysvm.pred,   
 mynb.pred$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)  
str(combine.classes)

## 'data.frame': 272 obs. of 5 variables:  
## $ myrf.pred : num 0 0 0 1 1 0 1 1 0 0 ...  
## $ myrda.pred.class: num 0 0 0 1 1 0 1 1 0 0 ...  
## $ mytree.pred : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ mysvm.pred : num 0 0 0 0 1 0 1 0 0 0 ...  
## $ mynb.pred.class : num 0 0 0 1 1 0 1 1 0 0 ...

combine.cl<-combine.classes[, -c(5,6)]  
majority.vote=rowSums(combine.classes[,-c(5,6)])  
head(majority.vote)

## 5 11 12 13 15 20   
## 1 1 1 3 4 1

combine.classes[,5]<-rowSums(combine.classes[,-c(5,6)])  
combine.classes[,6]<-ifelse(combine.classes[,5]>=4, "malignant", "benign")  
table(combine.classes[,6], breastCance.test$Class)

##   
## benign malignant  
## benign 174 7  
## malignant 3 88

Confusion\_combine <-table(combine.classes[,6], breastCance.test$Class)  
accuracy <- sum(diag(Confusion\_combine))/sum(Confusion\_combine)  
  
cat("\n","accuracy is",accuracy)

##   
## accuracy is 0.9632353

```{r}

# Calculate confusion matrix

confusion <- table(combine.classes[,6], breastCance.test$Class)

# Calculate performance metrics

accuracy <- sum(diag(confusion))/sum(confusion)

precision <- diag(confusion)/colSums(confusion)

recall <- diag(confusion)/rowSums(confusion)

f1\_score <- 2 \* precision \* recall / (precision + recall)

# Print results

cat("Accuracy: ", accuracy, "\n")

cat("Precision: ", precision, "\n")

cat("Recall: ", recall, "\n")

cat("F1 Score: ", f1\_score, "\n")

```

Accuracy: 0.9632353

Precision: 0.9830508 0.9263158

Recall: 0.961326 0.967033

F1 Score: 0.972067 0.9462366