Project2

Rex Cory

3/10/2021

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(e1071)

## Warning: package 'e1071' was built under R version 4.0.3

library(klaR)

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

## Loading required package: MASS

## Warning: package 'MASS' was built under R version 4.0.3

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

library(nnet)  
library(MASS)  
library(rpart)  
library(mlbench)

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

library(randomForest)

## Warning: package 'randomForest' was built under R version 4.0.3

## randomForest 4.6-14

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

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':  
##   
## combine

library(party)

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

## Loading required package: grid

## Loading required package: mvtnorm

## Warning: package 'mvtnorm' was built under R version 4.0.3

## Loading required package: modeltools

## Warning: package 'modeltools' was built under R version 4.0.3

## 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 4.0.3

library(ipred)

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

library(ROCR)

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

data(BreastCancer)

# remove missing values  
BreastCancer <- na.omit(BreastCancer)   
# remove the unique identifier  
BreastCancer$Id <- NULL   
# partition the data set for 80% training and 20% evaluation   
set.seed(2)  
  
ind <- sample(2, nrow(BreastCancer), replace = TRUE, prob=c(0.8, 0.2))

x.rp <- rpart(Class ~ ., data=BreastCancer[ind == 1,])  
x.rp.pred <- predict(x.rp, type="class", newdata=BreastCancer[ind == 2,])  
x.rp.prob <- predict(x.rp, type="prob", newdata=BreastCancer[ind == 2,])

x.ct <- ctree(Class ~ ., data=BreastCancer[ind == 1,])  
x.ct.pred <- predict(x.ct, newdata=BreastCancer[ind == 2,])  
x.ct.prob <- 1- unlist(treeresponse(x.ct, BreastCancer[ind == 2,]), use.names=F)[seq(1,nrow(BreastCancer[ind == 2,])\*2,2)]

x.cf <- cforest(Class ~ ., data=BreastCancer[ind == 1,], control = cforest\_unbiased(mtry = ncol(BreastCancer)-2))  
x.cf.pred <- predict(x.cf, newdata=BreastCancer[ind == 2,])  
x.cf.prob <- 1- unlist(treeresponse(x.cf, BreastCancer[ind == 2,]), use.names=F)[seq(1,nrow(BreastCancer[ind == 2,])\*2,2)]

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

#Tune  
x.svm.tune <- tune(svm, Class~., data = BreastCancer[ind == 1,], ranges = list(gamma = 2^(-8:1), cost = 2^(0:4)), tunecontrol = tune.control(sampling = "fix"))  
  
#SVM  
x.svm <- svm(Class~., data = BreastCancer[ind == 1,], cost=4, gamma=0.0625, probability = TRUE)  
x.svm.prob <- predict(x.svm, type="prob", newdata=BreastCancer[ind == 2,], probability = TRUE)

## plot ROC curves to compare the performance of the individual classifiers  
##  
  
# Output the plot to a PNG file for display on web.  
png(filename="roc\_curve\_5\_models.png", width=700, height=700)  
  
  
# create an ROCR prediction object from rpart() probabilities  
x.rp.prob.rocr <- prediction(x.rp.prob[,2], BreastCancer[ind == 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[ind == 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[ind == 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[ind == 2,'Class'])  
x.ip.perf <- performance(x.ip.prob.rocr, "tpr","fpr")  
plot.new()  
plot(x.ip.perf, col=5, add=TRUE)  
# svm  
x.svm.prob.rocr <- prediction(attr(x.svm.prob, "probabilities")[,2], BreastCancer[ind == 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()

## png   
## 2