TBANLT 560 Project 2

2023-03-21

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

## Loading required package: mlbench

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

# if you don't have any required package, use the install.packages() command  
# load the data set  
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)  
  
ind <- sample(2, nrow(BreastCancer), replace = TRUE, prob=c(0.8, 0.2))

# create model using svm (support vector machine)  
require(e1071)

## Loading required package: e1071

# svm requires tuning  
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"))  
# display the tuning results (in text format)  
x.svm.tune

##   
## Parameter tuning of 'svm':  
##   
## - sampling method: fixed training/validation set   
##   
## - best parameters:  
## gamma cost  
## 0.015625 4  
##   
## - best performance: 0.01675978

# 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[ind == 1,], cost=4, gamma=0.0625, probability = TRUE)  
x.svm.prob <- predict(x.svm, type="prob", newdata=BreastCancer[ind == 2,])  
table(x.svm.prob)

## x.svm.prob  
## benign malignant   
## 89 59

classifier1 <- c(89,59)

# create model using recursive partitioning on the training data set  
require(rpart)

## Loading required package: rpart

x.rp <- rpart(Class ~ ., data=BreastCancer[ind == 1,])  
# predict classes for the evaluation data set  
x.rp.pred <- predict(x.rp, type="class", newdata=BreastCancer[ind == 2,])  
# score the evaluation data set (extract the probabilities)  
x.rp.prob <- predict(x.rp, type="prob", newdata=BreastCancer[ind == 2,])  
table(x.rp.pred)

## x.rp.pred  
## benign malignant   
## 88 60

classifier2 <- c(88,60)

# create model using conditional inference trees  
require(party)

## Loading required package: party

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

## 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 4.2.3

## 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 4.2.3

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)]  
  
table(x.ct.pred)

## x.ct.pred  
## benign malignant   
## 88 60

classifier3 <- c(88,60)

# create model using random forest and bagging ensemble using conditional inference trees  
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)]  
table(x.cf.pred)

## x.cf.pred  
## benign malignant   
## 88 60

classifier4 <- c(88,60)  
  
  
combine.df <- data.frame(classifier1, classifier2, classifier3, classifier4)  
  
combine.df$vote <- rowSums(combine.df)  
  
combine.df$class <- ifelse(combine.df$vote > 2, 1, 0)  
  
combine.df

## classifier1 classifier2 classifier3 classifier4 vote class  
## 1 89 88 88 88 353 1  
## 2 59 60 60 60 239 1