# R Code for Classification and Regression Trees

## Classification Tree with rpart (mower)

install.packages("rpart")

library(rpart)

# read data

mower24 <- read.csv("C:/MA 299/R/mower24.csv")

# grow tree (default minsplit = 20 and xval = 10 fold cross validation)

fit <- rpart(ownership ~ income + lotsize,

method="class", data=mower24, control=rpart.control(minsplit = 1, xval = 10))

# show result

printcp(fit) # display the results

# plot tree

plot(fit, uniform=TRUE, main="Classification Tree for mower")

text(fit, use.n=TRUE, all=TRUE, cex=.8)

# create a postscript plot of tree

# You need to install postscript viewer.

# Go to http://download.cnet.com/Postscript-Viewer/3000-2094\_4-10845650.html.

post(fit, file = "c:/MA 299/R/mowertree.ps", title = "Classification Tree for Mower")

# prune the tree

pfit<- prune(fit, cp= fit$cptable[which.min(fit$cptable[,"xerror"]),"CP"])

# show result

printcp(pfit) # display the results

# plot the pruned tree

plot(pfit, uniform=TRUE, main="Pruned Classification Tree for Mower")

text(pfit, use.n=TRUE, all=TRUE, cex=.8)

# create a postscript plot of tree

post(pfit, file = "c:/MA 299/R/prunemowertree.ps", title = "Pruned Classification Tree for Mower")

## Classification Tree with rpart (UniverslBankMc)

# Classification Tree with rpart

install.packages("rpart")

library(rpart)

# read data

BankM <- read.csv("C:/MA 299/R/UniversalBankM.csv")

# grow tree (default minsplit = 1 and xval = 10 fold cross validation)

fit <- rpart(PersonalLoan ~ .,

method="class", data=BankM, control=rpart.control(cp = 0.01, minsplit = 1, xval = 10))

# show result

printcp(fit) # display the results

# plot tree

plot(fit, uniform=TRUE,

main="Classification Tree for UniversalankM")

text(fit, use.n=TRUE, all=TRUE, cex=.8)

# create a postscript plot of tree

# You need to install postscript viewer.

# Go to http://download.cnet.com/Postscript-Viewer/3000-2094\_4-10845650.html.

post(fit, file = "c:/MA 299/R/UniverslBankM.ps",

title = "Classification Tree for UniversalBankM")

## Prune Tree

# prune the tree

pfit<- prune(fit, cp= fit$cptable[which.min(fit$cptable[,"xerror"]),"CP"])

# show result

printcp(pfit) # display the results

# plot the pruned tree

plot(pfit, uniform=TRUE,

main="Pruned Classification Tree for pruneUniversalBankM")

text(pfit, use.n=TRUE, all=TRUE, cex=.8)

# create a postscript plot of tree

post(pfit, file = "c:/MA 299/R/pruneUniversalBankM.ps",

title = "Pruned Classification Tree for UniverslaBankM")

--------------------

## Regression Tree with rpart (Boston Housing)

# Regression Tree with rpart

install.packages("rpart")

library(rpart)

# read data

housing <- read.csv("C:/MA 299/R/housing.csv")

# grow tree (default minsplit = 10 and xval = 10 fold cross validation)

fit <- rpart(medv ~ .,

method="anova", data=housing, control=rpart.control(cp = 0.01, minsplit = 10, xval = 10))

# show result

printcp(fit) # display the results

# plot tree

plot(fit, uniform=TRUE,

main="Regression Tree for Housing")

text(fit, use.n=TRUE, all=TRUE, cex=.8)

# create a postscript plot of tree

# You need to install postscript viewer.

# Go to http://download.cnet.com/Postscript-Viewer/3000-2094\_4-10845650.html.

post(fit, file = "c:/MA 299/R/housing.ps",

title = "Regression Tree for Housing")

## Prune Tree

# prune the tree

pfit<- prune(fit, cp= fit$cptable[which.min(fit$cptable[,"xerror"]),"CP"])

# show result

printcp(pfit) # display the results

# plot the pruned tree

plot(pfit, uniform=TRUE,

main="Pruned Regression Tree for Housing")

text(pfit, use.n=TRUE, all=TRUE, cex=.8)

# create a postscript plot of tree

post(pfit, file = "c:/MA 299/R/pruneHousing.ps",

title = "Pruned Classification Tree for Housing")

--------------------

library(randomForest)

library(caret)

BankRFfull <- read.csv("C:/MA 299/R/UniversalBankRF.csv")

#Separate the data for prediction from the rest

#The data point for prediction is called "new" and the rest is called "BankRF"

new <- BankRFfull[5001,]

BankRF <- BankRFfull[1:5000,]

#Separate BankRF into the training set and the test set

idxs = sample(1:nrow(BankRF), as.integer(0.80\*nrow(BankRF)))

trainBankRF = BankRF[idxs,]

testBankRF = BankRF[-idxs,]

#Apply Random Forest with the training set

rf <- randomForest(as.factor(PersonalLoan) ~ ., data = trainBankRF, ntree = 500,

mtry = 4, nodesize = 5, importance = TRUE)

#Construct variable importance plot

varImpPlot(rf, type = 1)

#Create confusion matrix for the test set

rf.pred <- predict(rf, testBankRF)

confusionMatrix(rf.pred, as.factor(testBankRF$PersonalLoan))

#Predict class for the data point called "new"

result=predict(rf,new,"class")

result