w4-assignment1.R

mukhijas

2019-10-29

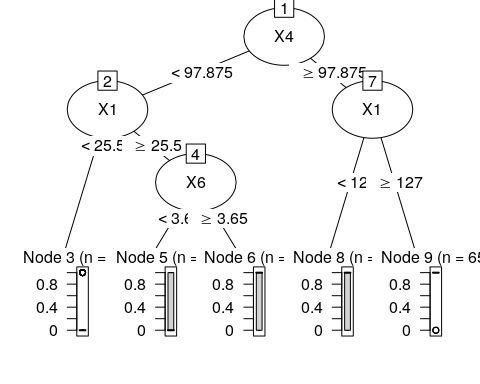
library("readxl")  
library("partykit")

## Loading required package: grid

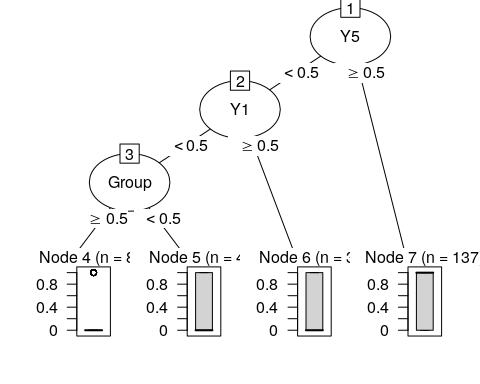
## Loading required package: libcoin

## Loading required package: mvtnorm

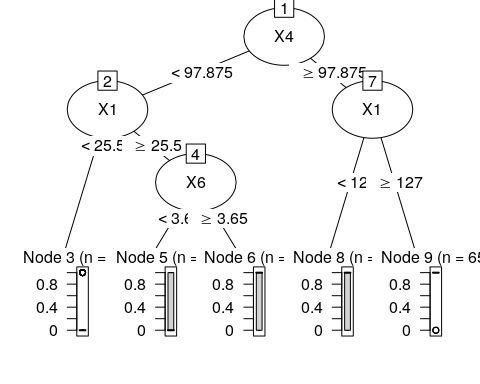
library("rpart")  
  
# Reading data from excel file  
df <- read\_excel("/users/pgrad/mukhijas/Desktop/DA-Labs/labs/data/DT-W4.xlsx")  
  
# Removing ID column  
df = subset(df, select = -c(ID))  
  
  
# Data with all X,Y and Groups  
df\_with\_X\_and\_Y\_all\_groups = data.frame(df)  
DT\_Model\_XY\_all\_groups <- rpart(Response~., data=df\_with\_X\_and\_Y\_all\_groups, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))  
plot(as.party((DT\_Model\_XY\_all\_groups)))



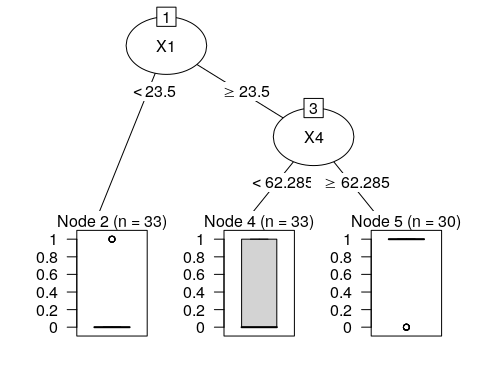
# Data without X.  
df\_without\_X\_all\_groups = data.frame(df)  
df\_without\_X\_all\_groups = within(df\_without\_X\_all\_groups, rm(X1, X2, X3, X4, X5, X6, X7))  
DT\_model\_without\_X\_all\_groups <- rpart(Response~., data=df\_without\_X\_all\_groups, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))  
plot(as.party((DT\_model\_without\_X\_all\_groups)))



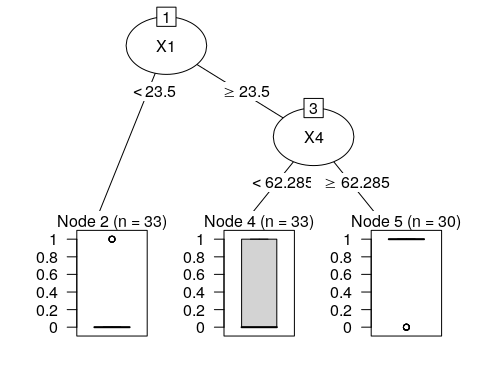
# Data without Y.  
df\_without\_Y\_all\_groups = data.frame(df)  
df\_without\_Y\_all\_groups = within(df\_without\_Y\_all\_groups, rm(Y1, Y2, Y3, Y4, Y5, Y6, Y7))  
DT\_model\_without\_Y\_all\_groups <- rpart(Response~., data=df\_without\_Y\_all\_groups, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))  
plot(as.party((DT\_model\_without\_Y\_all\_groups)))



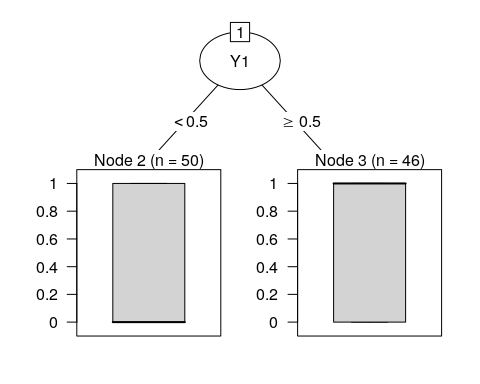
# Data with X and Y for group 0  
df\_with\_X\_and\_Y\_group\_0 = data.frame(df)  
df\_with\_X\_and\_Y\_group\_0 = df\_with\_X\_and\_Y\_group\_0[df\_with\_X\_and\_Y\_group\_0$Group == 0,]  
DT\_Model\_XY\_group\_0 <- rpart(Response~., data=df\_with\_X\_and\_Y\_group\_0, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))  
plot(as.party((DT\_Model\_XY\_group\_0)))



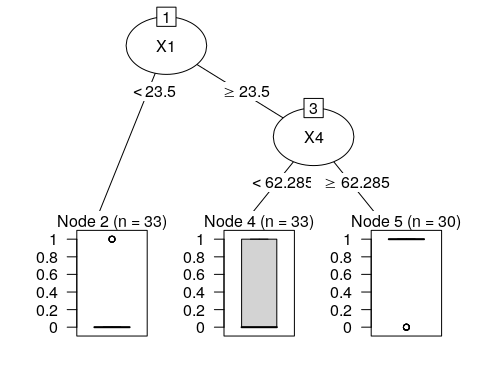
# Data with X and group 0  
df\_with\_X\_group\_0 = data.frame(df)  
df\_with\_X\_group\_0 = df\_with\_X\_group\_0[df\_with\_X\_group\_0$Group == 0,]  
df\_with\_X\_group\_0 = within(df\_with\_X\_group\_0, rm(Y1, Y2, Y3, Y4, Y5, Y6, Y7))  
DT\_Model\_X\_group\_0 <- rpart(Response~., data=df\_with\_X\_group\_0, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))  
plot(as.party((DT\_Model\_X\_group\_0)))



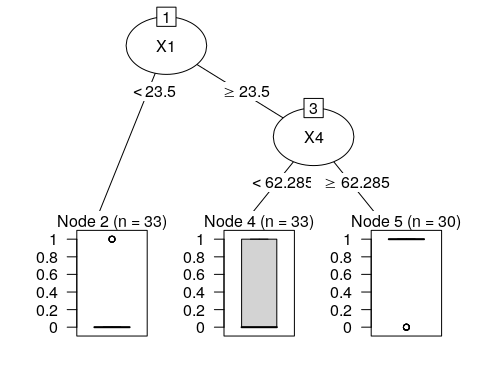
# Data with Y and group 0  
df\_with\_Y\_group\_0 = data.frame(df)  
df\_with\_Y\_group\_0 = df\_with\_Y\_group\_0[df\_with\_Y\_group\_0$Group == 0,]  
df\_with\_Y\_group\_0 = within(df\_with\_Y\_group\_0, rm(X1, X2, X3, X4, X5, X6, X7))  
DT\_Model\_Y\_group\_0 <- rpart(Response~., data=df\_with\_Y\_group\_0, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))  
plot(as.party((DT\_Model\_Y\_group\_0)))



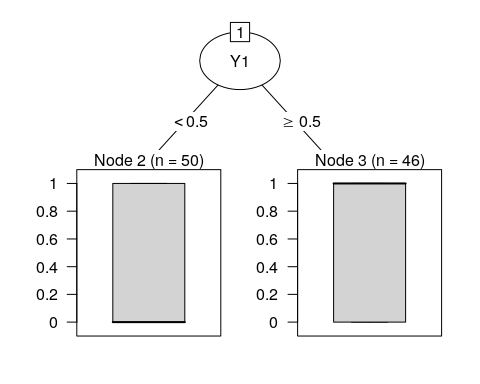
# Data with X, Y and group 1  
df\_with\_X\_and\_Y\_group\_1 = data.frame(df)  
df\_with\_X\_and\_Y\_group\_1 = df\_with\_X\_and\_Y\_group\_1[df\_with\_X\_and\_Y\_group\_1$Group == 0,]  
DT\_Model\_XY\_group\_1 <- rpart(Response~., data=df\_with\_X\_and\_Y\_group\_1, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))  
plot(as.party((DT\_Model\_XY\_group\_1)))



# Data with X and group 1  
df\_with\_X\_group\_1 = data.frame(df)  
df\_with\_X\_group\_1 = df\_with\_X\_group\_1[df\_with\_X\_group\_1$Group == 0,]  
df\_with\_X\_group\_1 = within(df\_with\_X\_group\_1, rm(Y1, Y2, Y3, Y4, Y5, Y6, Y7))  
DT\_Model\_X\_group\_1 <- rpart(Response~., data=df\_with\_X\_group\_1, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))  
plot(as.party((DT\_Model\_X\_group\_1)))



# Data with Y and group 1  
df\_with\_Y\_group\_1 = data.frame(df)  
df\_with\_Y\_group\_1 = df\_with\_Y\_group\_1[df\_with\_Y\_group\_1$Group == 0,]  
df\_with\_Y\_group\_1 = within(df\_with\_Y\_group\_1, rm(X1, X2, X3, X4, X5, X6, X7))  
DT\_Model\_Y\_group\_1 <- rpart(Response~., data=df\_with\_Y\_group\_1, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))  
plot(as.party((DT\_Model\_Y\_group\_1)))



set.seed(42)  
df = subset(df, select = -c(Group))  
model <- rpart(Response ~., data=df, control=rpart.control())  
print(model$cptable)

## CP nsplit rel error xerror xstd  
## 1 0.27172457 0 1.0000000 1.0053654 0.005097732  
## 2 0.06184006 1 0.7282754 0.7435479 0.051057072  
## 3 0.03512004 2 0.6664354 0.7215905 0.053522078  
## 4 0.02898245 3 0.6313153 0.7395593 0.056839008  
## 5 0.01672064 5 0.5733504 0.6881534 0.057966503  
## 6 0.01241449 8 0.5208745 0.6814719 0.062210719  
## 7 0.01000000 9 0.5084600 0.7197336 0.065886710

minError <- which.min(model$cptable [, "xerror"])  
cp <- model$cptable [minError,"CP"]  
Model\_Pruned <- prune(model, cp=cp)  
plot(as.party(Model\_Pruned))

