

HW_06

Preet Dabhi

3/29/2022

```
#####
# Company      : Stevens
# Project      : HW_06
# Purpose      : C5.0 and Random Forest methodology
# First Name   : Preet
# Last Name    : Dabhi
# Id           : 10459151
# Date         : 03/29/2022

## Delete all the objects from your R- environment.
#####

library(C50)

## Warning: package 'C50' was built under R version 4.1.3

library(randomForest)

## Warning: package 'randomForest' was built under R version 4.1.3

## randomForest 4.7-1

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

rm(list=ls())

# 6.1

bc_c50 <- read.csv("D:/SEM 3/CS 513/HW_02/breast-cancer-wisconsin.csv",header=TRUE, sep=", ")

View(bc_c50)

#random index creation
idx<-sort(sample(nrow(bc_c50),as.integer(.70*nrow(bc_c50))))
training_data<-bc_c50[idx,]
testing_data<-bc_c50[-idx,]

#install.packages("C50")
```

```

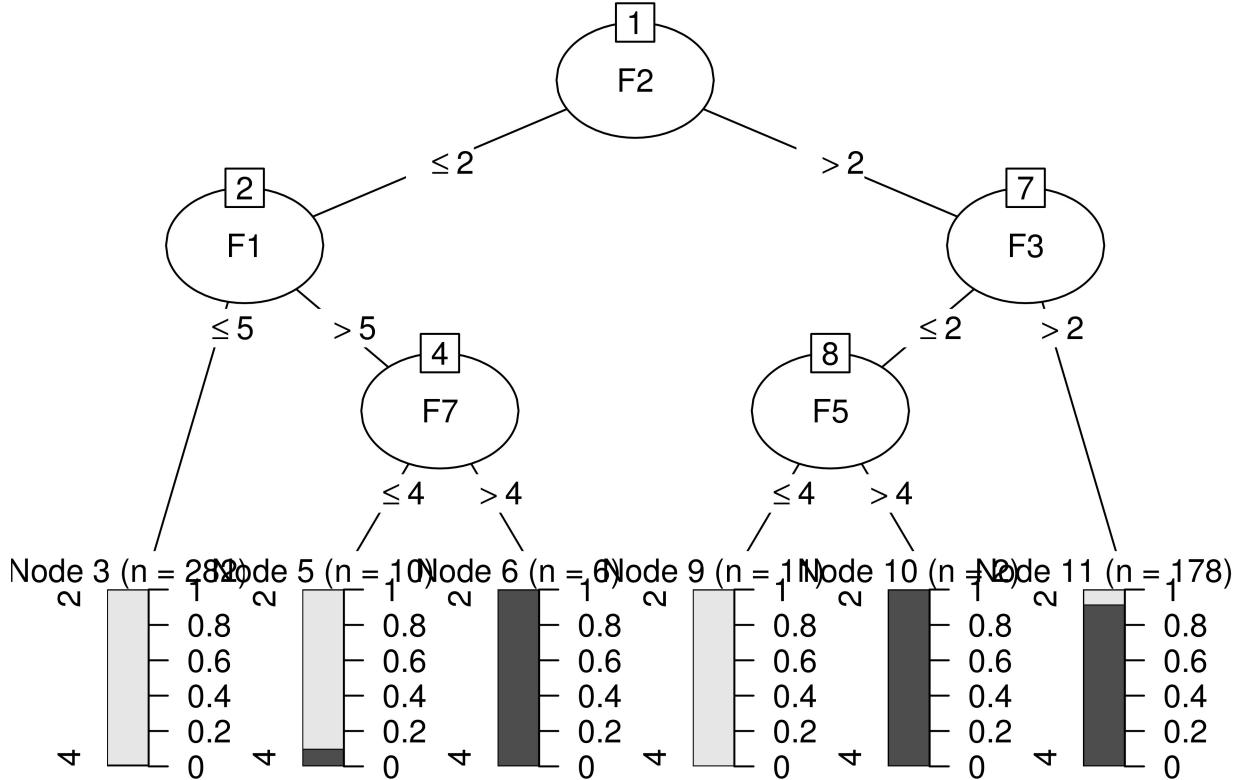
library(C50)
c50class<-C5.0(factor(Class)~., data=training_data)
summary(c50class)

##
## Call:
## C5.0.formula(formula = factor(Class) ~ ., data = training_data)
##
##
## C5.0 [Release 2.07 GPL Edition]      Tue Mar 29 11:55:41 2022
## -----
##
## Class specified by attribute 'outcome'
##
## Read 489 cases (11 attributes) from undefined.data
##
## Decision tree:
##
## F2 <= 2:
## :...F1 <= 5: 2 (282/3)
## :   F1 > 5:
## :     ...F7 <= 4: 2 (10/1)
## :       F7 > 4: 4 (6)
## F2 > 2:
## :...F3 > 2: 4 (178/15)
##     F3 <= 2:
##       ...F5 <= 4: 2 (11)
##         F5 > 4: 4 (2)
##
##
## Evaluation on training data (489 cases):
##
##      Decision Tree
## -----
##      Size      Errors
##
##        6    19( 3.9%)  <<
##
##
##      (a)    (b)    <-classified as
##      ----  ----
##        299    15    (a): class 2
##          4    171    (b): class 4
##
##
## Attribute usage:
##
## 100.00% F2
## 60.94% F1
## 39.06% F3
## 3.27% F7
## 2.66% F5
##

```

```
##  
## Time: 0.0 secs
```

```
plot(c50class)
```



```
prediction<-predict(c50class,testing_data,type="class")
```

```
#creates frequency table  
table(actual=testing_data[,11],prediction)
```

```
## prediction  
## actual 2 4  
## 2 134 10  
## 4 4 62
```

```
#percentage accuracy  
match<-(testing_data[,11]==prediction)*100  
acc<-sum(match)/length(match)  
acc
```

```
## [1] 93.33333
```

```

#error-rate
error<-testing_data[,11]!=prediction
error_rate<-sum(error)/length(error)
error_rate

## [1] 0.06666667

#6.2
rm(list=ls())

#load the library
library(randomForest)

bc_randomForest<-read.csv("D:/SEM 3/CS 513/HW_02/breast-cancer-wisconsin.csv",header=TRUE, sep=",")
View(bc_randomForest)

#factoring
bc_randomForest$Class <- factor(bc_randomForest$Class, levels = c(2,4),labels = c("Benign", "Malignant"))
bc_randomForest<-na.omit(bc_randomForest)

#dividing test and training data
index<-sort(sample(nrow(bc_randomForest),as.integer(.70*nrow(bc_randomForest))))
trainingdataset<-bc_randomForest[index,]
testdataset<-bc_randomForest[-index,]

dev.off()

## null device
##          1

fit <- randomForest( Class~, data=trainingdataset, importance=TRUE, ntree=1000)
importance(fit)

##           Benign Malignant MeanDecreaseAccuracy MeanDecreaseGini
## Sample -1.307588 10.037234          7.691286      3.986071
## F1     21.196027 27.212726         29.867882     14.671631
## F2     25.206866 23.698090         35.873010     56.370256
## F3     16.860709 23.813136         28.455347     50.086760
## F4     11.888937 10.211202         15.474016      5.601860
## F5     14.768918 15.668633         21.512044     22.965387
## F6     19.942383 22.947082         30.435318     16.343370
## F7     14.178365 21.021501         25.010413     22.294338
## F8     15.641686 13.907521         19.707813     20.230726
## F9     8.332922  9.164338          11.427581     1.722011

varImpPlot(fit)
prediction<- predict(fit, testdataset)
table(actual=testdataset$Class,prediction)

##           prediction
## actual      Benign Malignant
##   Benign        121       7
##   Malignant      7       75

```

```
#find the error rate
wrongdata<- (testdataset$Class!=prediction)
error<-sum(wrongdata)/length(wrongdata)
error
```

```
## [1] 0.06666667
```