Mall_HW5

Prashant Mall

3/17/2022

```
#CS513-HW5
#First Name: Prashant Pramodkumar
#Last Name: Mall
#CWID: 10459371
#HW Topic: Dtree

rm(list=ls())

library(class)
library(rpart)

#Read File
df <- read.csv("/Users/prashantmall1997/Library/CloudStorage/OneDrive-Personal/Coding/Stevens-Courses/CS
```

```
head(df, n=5)
```

```
##    Sample F1 F2 F3 F4 F5 F6 F7 F8 F9 Class
## 1 1000025  5  1  1  1  2  1  3  1  1     2
## 2 1002945  5  4  4  5  7 10  3  2  1     2
## 3 1015425  3  1  1  1  2  2  3  1  1     2
## 4 1016277  6  8  8  1  3  4  3  7  1     2
## 5 1017023  4  1  1  3  2  1  3  1  1     2
```

```
#Summary of each column
n <- as.numeric(as.character(df$F6))
```

```
## Warning: NAs introduced by coercion
```

```
df$F6 <- n
summary(df, na.rm = TRUE)
```

```
##      Sample               F1              F2              F3
##  Min.   :   61634   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
##  1st Qu.:  870688   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 1.000
##  Median : 1171710   Median : 4.000   Median : 1.000   Median : 1.000
##  Mean   : 1071704   Mean   : 4.418   Mean   : 3.134   Mean   : 3.207
##  3rd Qu.: 1238298   3rd Qu.: 6.000   3rd Qu.: 5.000   3rd Qu.: 5.000
##  Max.   :13454352   Max.   :10.000   Max.   :10.000   Max.   :10.000
##
##        F4              F5              F6              F7
```

```
##  Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
##  1st Qu.: 1.000   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 2.000
##  Median : 1.000   Median : 2.000   Median : 1.000   Median : 3.000
##  Mean   : 2.807   Mean   : 3.216   Mean   : 3.545   Mean   : 3.438
##  3rd Qu.: 4.000   3rd Qu.: 4.000   3rd Qu.: 6.000   3rd Qu.: 5.000
##  Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000
##                                    NA's   :16
##        F8               F9            Class
##  Min.   : 1.000   Min.   : 1.000   Min.   :2.00
##  1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.:2.00
##  Median : 1.000   Median : 1.000   Median :2.00
##  Mean   : 2.867   Mean   : 1.589   Mean   :2.69
##  3rd Qu.: 4.000   3rd Qu.: 1.000   3rd Qu.:4.00
##  Max.   :10.000   Max.   :10.000   Max.   :4.00
##
```

```
#Remove rows with missing values
df <- na.omit(df)

#Labels to Factor Class
df$Class<- factor(df$Class , levels = c("2","4") , labels = c("Benign","Malignant"))
is.factor(df$Class)
```

```
## [1] TRUE
```

```
#Train and Test - ratio 70% to 30%
df<- df[2:11]
size <- floor(0.70 * nrow(df))

#Set Seed
set.seed(123)
random <- sample(seq_len(nrow(df)), size = size)

#70% Data - Train
train <- df[random, ]

#30% Data - Test
test <- df[-random, ]

#CART
cart <- rpart(Class ~ ., data = train, method = "class")

#Predicting Class - Test Set
predicted <- predict(cart, test, type = "class")
print(length(predicted))
```

```
## [1] 205
```

```
print(length(test$Class))
```

```
## [1] 205
```

```r
#Confusion Matrix
conf_matrix <- table(predicted,test$Class)
print(conf_matrix)
```

```
##
## predicted   Benign Malignant
##   Benign       136         9
##   Malignant      3        57
```

```r
#Accuracy
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy(conf_matrix)
```

```
## [1] 94.14634
```