Assignment 5

Siddharth Harsukh Pansuria | CWID: 20005837

3/19/2022

Clearing environmental variables.

```
rm(list = ls())
```

Loading the breast cancer dataset and replacing '?' with NA

```
data <- read.csv('breast-cancer-wisconsin.csv')
```

Displaying top 10 rows of the data loaded

```
head(data, n=10)
```

```
##      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
## 6   1017122  8 10 10  8  7 10  9  7  1     4
## 7   1018099  1  1  1  1  2 10  3  1  1     2
## 8   1018561  2  1  2  1  2  1  3  1  1     2
## 9   1033078  2  1  1  1  2  1  1  1  5     2
## 10  1033078  4  2  1  1  2  1  2  1  1     2
```

Removing first column since it only contains ID

```
data <- data[-1]
```

Setting column 'Class' as a factor of benign and malignant

```
data$Class <- factor(data$Class, levels = c(2,4), labels = c("benign", "malignant"))
```

Displaying top 10 rows of the data loaded

```
head(data, n = 10)
```

```
##     F1 F2 F3 F4 F5 F6 F7 F8 F9      Class
## 1   5  1  1  1  2  1  3  1  1     benign
## 2   5  4  4  5  7 10  3  2  1     benign
## 3   3  1  1  1  2  2  3  1  1     benign
## 4   6  8  8  1  3  4  3  7  1     benign
## 5   4  1  1  3  2  1  3  1  1     benign
## 6   8 10 10  8  7 10  9  7  1  malignant
## 7   1  1  1  1  2 10  3  1  1     benign
## 8   2  1  2  1  2  1  3  1  1     benign
## 9   2  1  1  1  2  1  1  1  5     benign
## 10  4  2  1  1  2  1  2  1  1     benign
```

Splitting the dataset into training and test data

```
id <- sort(sample(nrow(data), as.integer(.70 * nrow(data))))
train_data <- data[id,]
test_data <- data[-id,]
```

Importibng library and implementing CART

```
library(rpart)
cart <- rpart(Class ~ ., data = train_data, method = "class")
```

Predicting on the test data

```
predict_CART <- predict(cart, test_data, type = "class")
```

Frequency table for CART

```
table(CART = predict_CART, Class = test_data$Class)
```

```
##            Class
## CART        benign malignant
##   benign       141         2
##   malignant      9        58
```

Calculating the error rate

```
wrong_preds <- sum(predict_CART != test_data$Class)
```

Number of wrong predictions

```
error_rate <- wrong_preds/length(predict_CART)
print(paste("Error Rate:" , error_rate))
```

Error Rate in Naïve Bayes predictions

```
## [1] "Error Rate: 0.0523809523809524"
```