Assignment 2

The assignments including couple of function. First use table function to to get confusion matrices and the result shows there are 119 true negative where both class and scored the class are 0. 27 true positive where both class and scored the class are 1 and there are 30 false negative where class is 1 and scored class is 0 and 5 false positive. Then we compute accuracy which is the total number of right prediction divide by total number of data points the result is
0.8066298. After that we calculate Error rate which is total number of wrong prediction divide by total number of data points and we get 0.1933702.
Then we compute Precision which is total number of correct prediction of positive outcome divide by total number of predicted positive outcome and we get 0.84375. We also compute Sensitivity and Specificity. We get 0.4736842 for Sensitivity and 0.9596774 for Specificity. Using both metrics, we compute F1 score which equals 0.6067416. The F1 score is bonded by 0 and 1 because the formula for the score is 2TP divide by sum of 2TP, FN and FP. All the matrices are positive. In extreme case if the model is very good with no wrong prediction then the score is 1 since FN and FP are 0 and if the model is very bad with all wrong predictions then the score will be 0.

After compute all matrices we plot the ROC curve. Then we use confusionMatrix function to reproduce all the calculating above. The result are rounded to 4 digits but it's very close to our calculation with accuracy equals 0.8066 and Sensitivity and Specificity equal to 0.4737 and 0.9597 respectively.
The ROC curve also looks similar which both AUC equals 0.85

Appendix

Assignment2

```
library(caret)
## Warning: package 'caret' was built under R version 3.4.4
## Loading required package: lattice
## Loading required package: ggplot2
library(pROC)
## Warning: package 'pROC' was built under R version 3.4.4
## Type 'citation("pROC")' for a citation.
##
## Attaching package: 'pROC'
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
library(ggplot2)
```

Q3

```
Accuracy <- function(data) {
tb = table(data$class,data$scored.class)
TN=tb[1,1]
TP=tb[2,2]
FN=tb[2,1]
FP=tb[1,2]

return((TP+TN)/(TP+FP+TN+FN))

}
Accuracy(data)
## [1] 0.8066298
```

Q4

```
Class_Error_Rate <- function(data) {
tb = table(data$class,data$scored.class)
TN=tb[1,1]
TP=tb[2,2]
FN=tb[2,1]
FP=tb[1,2]
```

```
return((FP+FN)/(TP+FP+TN+FN))

}
Class_Error_Rate(data)
## [1] 0.1933702
```

Q5

```
Precision <- function(data) {
tb = table(data$class,data$scored.class)
TP=tb[2,2]
FP=tb[1,2]

return((TP)/(TP+FP))

}
Precision(data)
## [1] 0.84375
```

Q6

```
Sensitivity <- function(data) {
tb = table(data$class,data$scored.class)
TP=tb[2,2]
FN=tb[2,1]

return((TP)/(TP+FN))

}

Sensitivity(data)
## [1] 0.4736842
```

Q7

```
Specificity <- function(data) {
tb = table(data$class,data$scored.class)
TN=tb[1,1]
TP=tb[2,2]
FN=tb[2,1]
FP=tb[1,2]

return((TN)/(TN+FP))

}
Specificity(data)
## [1] 0.9596774
```

Q8

```
F1_score <- function(data){
```

```
Sensitivity <- Sensitivity(data)
Precision   <-Precision(data)

return((2*Sensitivity*Precision)/(Sensitivity+Precision))

}
F1_score(data)
## [1] 0.6067416
```

Q10

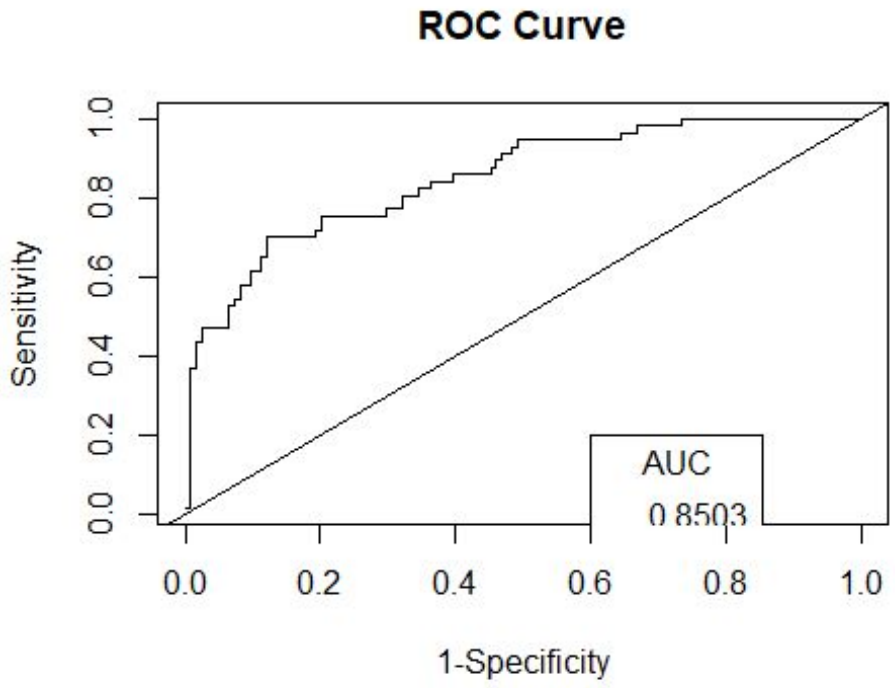
```
ROC = function(labels, scores){
  labels = labels[order(scores, decreasing=TRUE)]
  result =data.frame(TPR=cumsum(labels)/sum(labels), FPR=cumsum(!labels)
/sum(!labels), labels)

  FPR_df = c(diff(result$FPR), 0)
  TPR_df = c(diff(result$TPR), 0)
  AUC = round(sum(result$TPR * FPR_df) + sum(TPR_df * FPR_df)/2,4)

  plot(result$FPR,result$TPR,type="l",main ="ROC Curve",ylab="Sensitivi
ty",xlab="1-Specificity")
  abline(a=0,b=1)
  legend(.6,.2,AUC,title = "AUC")

}

ROC(data$class,data$scored.probability)
```



Q11

```
Calss_Mat <-function(data){
Accuracy<-Accuracy(data)
Class_Error_Rate<-Class_Error_Rate(data)
Precision<-Precision(data)
Sensitivity<-Sensitivity(data)
Specificity<-Specificity(data)
F1_score<-F1_score(data)
ROC(data$class,data$scored.probability)
p1<-paste("Accuracy: ",Accuracy)
p2<-paste("Class Error Rate: ",Class_Error_Rate)
p3<-paste("Precision: ",Precision)
p4<-paste("Sensitivity: ",Sensitivity)
p5<-paste("Specificity: ",Specificity)
p6<-paste("F1_score: ",F1_score)
print(p1)
print(p2)
print(p3)
print(p4)
print(p5)
print(p6)
}
Calss_Mat(data)
```