DATA 621 Homework 2

Critical Thinking Group 1

October 06, 2021

Contents

Prepared for:

Prof. Dr. Nasrin Khansari

City University of New York, School of Professional Studies - Data 621

DATA 621 – Business Analytics and Data Mining

Home Work 2

Prepared by:

Critical Thinking Group 1

Vic Chan

Gehad Gad

Evan McLaughlin

Bruno de Melo

Anjal Hussan

Zhouxin Shi

Sie Siong Wong

Data Source

```
cm <- read.csv("https://raw.githubusercontent.com/ahussan/DATA_621_Group1/main/HW2/classification-outpu
```

Data Explained and Confusion Matrix

From the confusion matrix table below, we've actual class value versus predicted class value. There are 119 true negative, 27 true positive, 30 false negative, and 5 false positive.

```
conf_mtx <- cm %>% dplyr::select(class, scored.class) %>% table()
conf_mtx
```

```
##      scored.class
## class   0   1
##     0 119   5
##     1  30  27
```

Function for Accuracy of Predictions

```
accuracy <- function(x){

  numerator <- x[2,2] + x[1,1]
  denominator <- sum(x)
  return(numerator/denominator)
}
```

Function for Classification Error Rate of Predictions

```
error_rate <- function(x){

  numerator <- x[1,2] + x[2,1]
  denominator <- sum(x)
  return(numerator/denominator)
}
```

```
accuracy(conf_mtx) + error_rate(conf_mtx)
```

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

Function for Precisions of Predictions

```
precision <- function(x){

  numerator <- x[1,1]
  denominator <- x[1,1] + x[1,2]
  return(numerator/denominator)
}
```

Function for Sensitivity of Predictions

```
sensitivity <- function(x){

  numerator <- x[1,1]
  denominator <- x[1,1] + x[2,1]
  return(numerator/denominator)
}
```

Function for Specificity of Predictions

```
specificity <- function(x){

  numerator <- x[2,2]
  denominator <- x[2,2] + x[1,2]
  return(numerator/denominator)
}
```

F1 Score of Predictions

```
f1_score <- function(x){

  numerator <- 2 * precision(x) * sensitivity(x)
  denominator <- precision(x) + sensitivity(x)
  return(numerator/denominator)
}
```

Investigation of the pROC R package.

We used the pROC R package to generate an ROC curve for the data set.

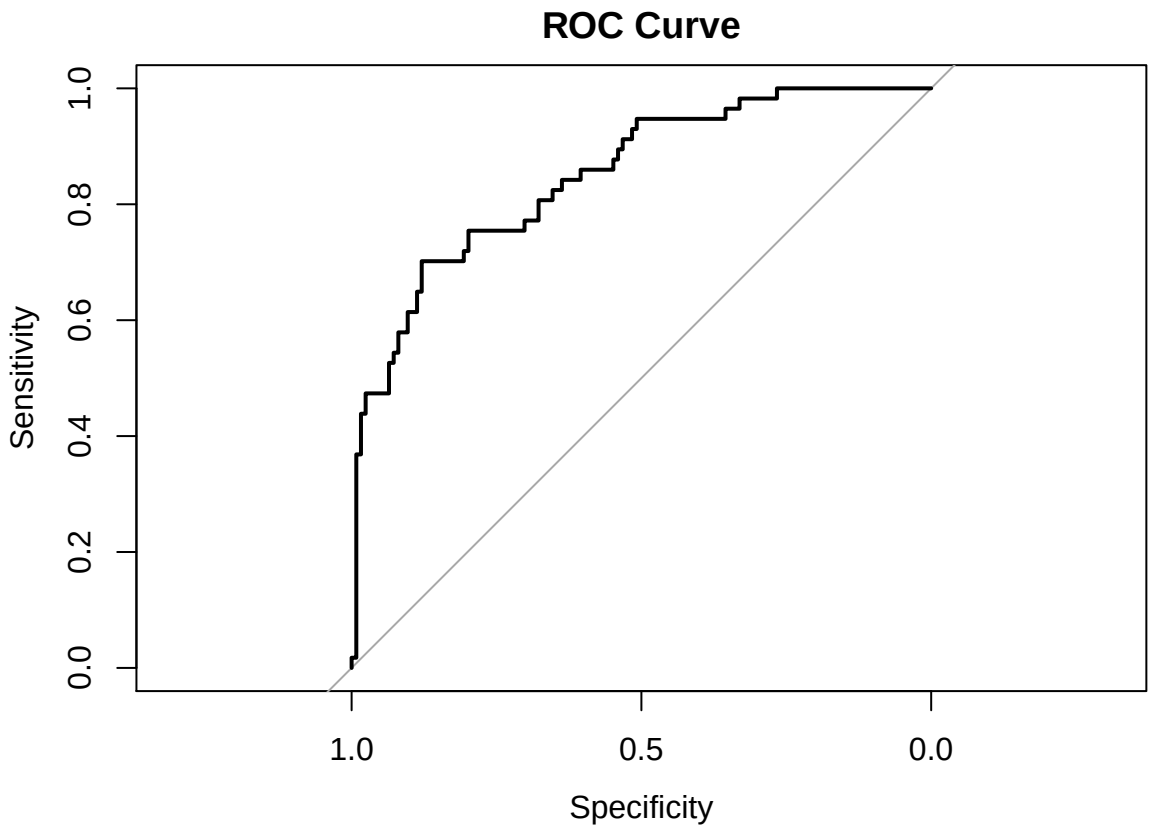
```
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
rcurve <- roc(cm$class~cm$scored.probability)
```

```
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
```

```
plot(rcurve, main="ROC Curve")
```



Best Threshold value using pROC package is {Threshold = 0.375117,fpr = 0.120968,tpr = 0.701754}

Note: The second method (using auc) predicts better than first method (using distance from (0,1))