Assignment3

2024-03-06

```
library(readr)
library(class)
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
library(ISLR)
library(reshape2)
library(e1071)
library(pROC)
```

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

```
##
## Attaching package: 'pROC'
```

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

```
library(gmodels)
```

```
##
## Attaching package: 'gmodels'
```

```
## The following object is masked from 'package:pROC':
##
##     ci
```

```
UniversalBank <- read_csv("UniversalBank.csv")
```

```
## Rows: 5000 Columns: 14-- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (14): ID, Age, Experience, Income, ZIP Code, Family, CCAvg, Education, M...
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
Training_Index <- createDataPartition(UniversalBank$Experience, p = .5996, list = FALSE)
Training_data <- UniversalBank[Training_Index,]
Validation_data <- UniversalBank[-Training_Index,]
```

```
pivot_UB<- table(Training_data$CreditCard, Training_data$Online, Training_data$`Personal Loan`)
rownames(pivot_UB) <- c("No CC", "Yes CC")
colnames(pivot_UB) <- c("No Online", "Yes Online")
pivot_UB
```

```
## , ,  = 0
##
```

```
##
##          No Online Yes Online
##   No CC        780       1130
##   Yes CC       309        495
##
## , ,  = 1
##
##
##          No Online Yes Online
##   No CC         76        131
##   Yes CC        30         49
```

```
CC_Online_1 <- Training_data$`Personal Loan`[Training_data$CreditCard == "1" & Training_data$Online ==
mean(CC_Online_1)
```

```
## [1] 0.09007353
```

```
pivot_table_online <- table(Training_data$`Personal Loan`, Training_data$Online)
rownames(pivot_table_online) <- c("No Loan", "Loan")
colnames(pivot_table_online) <- c("Offline", "Online")
pivot_table_online
```

```
##
##           Offline Online
##   No Loan    1089   1625
##   Loan        106    180
```

```
pivot_table_cc <- table(Training_data$`Personal Loan`, Training_data$CreditCard)
rownames(pivot_table_cc) <- c("No Loan", "Loan")
colnames(pivot_table_cc) <- c("No Credit Card", "Has Credit Card")
pivot_table_cc
```

```
##
##           No Credit Card Has Credit Card
##   No Loan           1910             804
##   Loan               207              79
```

```
cc_loan_acceptors <- sum(Training_data$CreditCard[Training_data$`Personal Loan` == 1])
loan_acceptors <- sum(Training_data$`Personal Loan` == 1)
prop_cc_loan_acceptors <- cc_loan_acceptors / loan_acceptors
prop_cc_loan_acceptors
```

```
## [1] 0.2762238
```

```
online_loan_acceptors <- sum(Training_data$Online[Training_data$`Personal Loan` == 1])
prop_online_loan_acceptors <- online_loan_acceptors / loan_acceptors
prop_online_loan_acceptors
```

```
## [1] 0.6293706
```

```
total_nrows <- nrow(Training_data)
prop_loan_acceptors <- loan_acceptors / total_nrows
prop_loan_acceptors
```

```
## [1] 0.09533333
```

```
cc_loan_rejector <- sum(Training_data$CreditCard[Training_data$`Personal Loan` == 0])
loan_rejector <- sum(Training_data$`Personal Loan` == 0)
prop_cc1_loan_rejector <- cc_loan_rejector / loan_rejector
```

```r
prop_cc1_loan_rejector
```

```
## [1] 0.2962417
```

```r
online_loan_rejectors <- sum(Training_data$Online[Training_data$`Personal Loan` == 0])
prop_online1_loan_rejector <- online_loan_rejectors / loan_rejector
prop_online1_loan_rejector
```

```
## [1] 0.5987472
```

```r
prop_loan_rejectors <- loan_rejector / total_nrows
prop_loan_rejectors
```

```
## [1] 0.9046667
```

```r
naiveBayes_model <- naiveBayes(`Personal Loan` ~ CreditCard + Online, data = Training_data)
predicted_test_labels <- predict(naiveBayes_model, Validation_data, type = "raw")
roc(Validation_data$`Personal Loan`, predicted_test_labels[,2])
```

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

```
## Setting direction: controls < cases
```

```
##
## Call:
## roc.default(response = Validation_data$`Personal Loan`, predictor = predicted_test_labels[,     2])
##
## Data: predicted_test_labels[, 2] in 1806 controls (Validation_data$`Personal Loan` 0) < 194 cases (Va
## Area under the curve: 0.4816
```

```r
# The one calculated in the pivot table had a better output than this
```