Assignment_3 - Gabriel Arsego

Garsego

2025-10-09

```
set.seed(123)
data <- read.csv("C:\\Users\\arseg\\Downloads\\UniversalBank.csv")
```

Partitioning the data into training (60%) and validation (40%)

```
Index_Train <- createDataPartition(data$Personal.Loan, p = 0.6, list = F)
Training <- data[Index_Train, ]
Validation <- data[-Index_Train, ]
```

Direction A.

```
#Creating the pivot table:
table_train <- table(Training$CreditCard, Training$Personal.Loan,
Training$Online)
ft_table_train <- ftable(table_train)
ft_table_train

##          0    1
##
## 0 0    785 1145
##   1     65  122
## 1 0    317  475
##   1     34   57
```

Direction B.

```
# For Online = 1 and CreditCard = 1
Loan_0 <- ft_table_train[3, 2]
Loan_1 <- ft_table_train[4, 2]
Prob_Loan_given_CC_Online <- Loan_1 / (Loan_0 + Loan_1)
Prob_Loan_given_CC_Online

## [1] 0.1071429
```

Direction C.

```
table_online <- table(Training$Personal.Loan, Training$Online)
table_online

## 
##         0    1
##   0 1102 1620
##   1   99  179

table_credit <- table(Training$Personal.Loan, Training$CreditCard)
table_credit

## 
##         0    1
##   0 1930  792
##   1  187   91
```

Direction D.

```
#i. P(CC = 1 | Loan = 1)
Credit_Card_Holders = 91 / (187 + 91)
Credit_Card_Holders

## [1] 0.3273381

#ii. P(Online = 1 | Loan = 1)
Loan_Online = 179 / (179 + 99)
Loan_Online

## [1] 0.6438849

#iii. P(Loan = 1)
Loan_Yes = (187 + 91)/ (1930 + 792 + 187 + 91)
Loan_Yes

## [1] 0.09266667

#iv. P(CC = 1 | Loan = 0)
Credit_No_Loan = 792 / (1930 + 792)
Credit_No_Loan

## [1] 0.2909625

#v. P(Online = 1 | Loan = 0)
Online_No_Loan = 1620 / (1102 + 1620)
Online_No_Loan

## [1] 0.5951506
```

```
#vi. P(Loan = 0)
Loan_No = (1930 + 792) / (1930 + 792 + 91 + 187)
Loan_No

## [1] 0.9073333
```

Direction E.

```
#P(Loan = 1 | CC = 1, Online = 1)
numerator <- Loan_Yes * Credit_Card_Holders * Loan_Online
denominator <- numerator + (Loan_No * Credit_No_Loan * Online_No_Loan)

Bayes_Probability <- numerator / denominator
Bayes_Probability

## [1] 0.1105637
```

Direction F.

```
#Comparing Bayes_Probability and Prob_Loan_given_CC_Online
Bayes_Probability

## [1] 0.1105637

Prob_Loan_given_CC_Online

## [1] 0.1071429

#The value found in Direction B is likely more accurate because it used the
true observed proportion.
```

Direction G.

```
#Using naive Bayes:
naive_B_model <- naiveBayes(Personal.Loan~CreditCard+Online,
                            data = Training)


naive_B_model

##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##          0          1
## 0.90733333 0.09266667
##
## Conditional probabilities:
##    CreditCard
## Y        [,1]      [,2]
##   0 0.2909625 0.4542897
##   1 0.3273381 0.4700881
##
##    Online
## Y        [,1]      [,2]
##   0 0.5951506 0.4909531
##   1 0.6438849 0.4797134



new_data <- data.frame(CreditCard = 1, Online = 1)
predict(naive_B_model, new_data, type = "raw")

##              0         1
## [1,] 0.8843065 0.1156935

#The second column showing 0.1156935, shows P(Loan = 1 | CC = 1, Online = 1)
when using naiveBayes(). Comparing the values:
##Direction E: 0.1105637
##Direction G: 0.1156935
#The values obtained are really close, meanining that the manual calculations
matched the naive Bayes model.
```