Assignment 03

Vamshikrishna Sunnam

2023-03-05

```
BankData <- read.csv("C:/Users/vamsh/OneDrive/Documents/KENT SEM 01/FML/Assignment 3/UniversalBank.csv")
summary(BankData)
```

```
##        ID            Age          Experience       Income          ZIP.Code
##  Min.   :   1   Min.   :23.00   Min.   :-3.0   Min.   :  8.00   Min.   : 9307
##  1st Qu.:1251   1st Qu.:35.00   1st Qu.:10.0   1st Qu.: 39.00   1st Qu.:91911
##  Median :2500   Median :45.00   Median :20.0   Median : 64.00   Median :93437
##  Mean   :2500   Mean   :45.34   Mean   :20.1   Mean   : 73.77   Mean   :93153
##  3rd Qu.:3750   3rd Qu.:55.00   3rd Qu.:30.0   3rd Qu.: 98.00   3rd Qu.:94608
##  Max.   :5000   Max.   :67.00   Max.   :43.0   Max.   :224.00   Max.   :96651
##      Family          CCAvg          Education        Mortgage
##  Min.   :1.000   Min.   : 0.000   Min.   :1.000   Min.   :  0.0
##  1st Qu.:1.000   1st Qu.: 0.700   1st Qu.:1.000   1st Qu.:  0.0
##  Median :2.000   Median : 1.500   Median :2.000   Median :  0.0
##  Mean   :2.396   Mean   : 1.938   Mean   :1.881   Mean   : 56.5
##  3rd Qu.:3.000   3rd Qu.: 2.500   3rd Qu.:3.000   3rd Qu.:101.0
##  Max.   :4.000   Max.   :10.000   Max.   :3.000   Max.   :635.0
##  Personal.Loan   Securities.Account   CD.Account         Online
##  Min.   :0.000   Min.   :0.0000     Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.0000     1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.000   Median :0.0000     Median :0.0000   Median :1.0000
##  Mean   :0.096   Mean   :0.1044     Mean   :0.0604   Mean   :0.5968
##  3rd Qu.:0.000   3rd Qu.:0.0000     3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.0000     Max.   :1.0000   Max.   :1.0000
##    CreditCard
##  Min.   :0.000
##  1st Qu.:0.000
##  Median :0.000
##  Mean   :0.294
##  3rd Qu.:1.000
##  Max.   :1.000
```

```
library(caret)
```

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

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

```
library(e1071)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(class)
library(ggplot2)
```

```
#converting variables
BankData$Personal.Loan <- factor(BankData$Personal.Loan)
BankData$Online <- factor(BankData$Online)
BankData$CreditCard <- factor(BankData$CreditCard)
df= BankData
```

```
#TASK1
set.seed(64060)
Train_index <- createDataPartition(df$Personal.Loan, p = 0.6, list = FALSE)
train.df = df[Train_index,]
validation.df = df[-Train_index,]
mytable <- xtabs(~ CreditCard + Online + Personal.Loan , data = train.df)
ftable(mytable)
```

```
##                   Personal.Loan    0    1
## CreditCard Online
## 0          0                     772   75
##            1                    1152  120
## 1          0                     309   34
##            1                     479   59
```

```
#TASK2
probability = 59/(59+479)
probability
```

```
## [1] 0.1096654
```

```
#TASK3
table(Personal.Loan = train.df$Personal.Loan, Online = train.df$Online)
```

```
##              Online
## Personal.Loan    0    1
##             0 1081 1631
##             1  109  179
```

```r
table(Personal.Loan = train.df$Personal.Loan, CreditCard = train.df$CreditCard)
```

```
##              CreditCard
## Personal.Loan    0    1
##             0 1924  788
##             1  195   93
```

```r
table(Personal.Loan = train.df$Personal.Loan)
```

```
## Personal.Loan
##    0    1
## 2712  288
```

```r
#TASK4
#i. P(CC = 1 | Loan = 1) (the proportion of loan acceptors who have credit cards)
Probablity1 <- 93/(93+195)
Probablity1
```

```
## [1] 0.3229167
```

```r
#ii. P(Online = 1 | Loan = 1)
Probablity2 <- 179/(179+109)
Probablity2
```

```
## [1] 0.6215278
```

```r
#iii. P(Loan = 1) (the proportion of loan acceptors)
Probablity3 <- 288/(288+2712)
Probablity3
```

```
## [1] 0.096
```

```r
#iv. P(CC = 1 | Loan = 0)
Probablity4 <- 788/(788+1924)
Probablity4
```

```
## [1] 0.2905605
```

```r
#v. P(Online = 1 | Loan = 0)
Probablity5 <- 1631/(1631+1081)
Probablity5
```

```
## [1] 0.6014012
```

```r
#vi. P(Loan = 0)
Probablity6 <- 2712/(2712+288)
Probablity6
```

```
## [1] 0.904
```

```r
#TASK5
Task5Probablity <- (Probablity1*Probablity2*Probablity3)/
((Probablity1*Probablity2*Probablity3) +(Probablity4*Probablity5*Probablity6))
Task5Probablity
```

```
## [1] 0.1087106
```

```r
#TASK6
#The values we obtained from questions 2 and 5 are nearly identical.
#The exact method is distinguished from the naive bayes method.
#To predict, we need a similar independent variable and classification, which the naive bayes method do
#due to the fact that we used the same values from the pivot table.
```

```r
#Task7
#Apply naive Bayes to the data. Examine the model output on training data and identify the entry corres
#Contrast this with the figure you obtained in (E).
nb.model <- naiveBayes(Personal.Loan ~ Online + CreditCard, data = train.df)

# convert Online and CreditCard variables in new data to factors with same labels as in training data
To_Predict <- data.frame(Online = factor(1, levels = levels(train.df$Online)),
                         CreditCard = factor(1, levels = levels(train.df$CreditCard)))

# Predict the probability of getting a loan
prob_loan <- predict(nb.model, To_Predict, type = 'raw')[, 2]

# Calculate the probability of having a credit card and being online
prob_cc_online <- mytable[1, 2, 1]/sum(mytable[1, 2, ])

# Calculate the conditional probability of getting a loan given having a credit card and being online
cond_prob <- prob_loan * Task5Probablity/prob_cc_online

cond_prob
```

```
##          1
## 0.01304903
```