ML Assignment 3

Abhinay Chiranjeetgh Marneni

2022-10-11

Loading library functions Packages

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

```
## Loading required package: ggplot2

## Loading required package: lattice
```

```
library(ISLR)
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(gmodels)
library(ggplot2)
library(knitr)
library(e1071)
library(pROC)
```

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

##
## Attaching package: 'pROC'

## The following object is masked from 'package:gmodels':
##
##     ci

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

Importing the Uviversal bank.csv file and to exarct the data into factor variables

```
data1<-read.csv("C:/Users/abhin/OneDrive/Documents/Assigments Buss 1sem/ML/UniversalBank.csv",header=TRU
head(data1)
```

```
##   ID Age Experience Income ZIP.Code Family CCAvg Education Mortgage
## 1  1  25          1     49    91107      4   1.6         1        0
## 2  2  45         19     34    90089      3   1.5         1        0
## 3  3  39         15     11    94720      1   1.0         1        0
## 4  4  35          9    100    94112      1   2.7         2        0
## 5  5  35          8     45    91330      4   1.0         2        0
## 6  6  37         13     29    92121      4   0.4         2      155
##   Personal.Loan Securities.Account CD.Account Online CreditCard
## 1             0                  1          0      0          0
## 2             0                  1          0      0          0
## 3             0                  0          0      0          0
## 4             0                  0          0      0          0
## 5             0                  0          0      0          1
## 6             0                  0          0      1          0
```

```
data1$Personal.Loan <- as.factor(data1$Personal.Loan)
data1$Online <- as.factor(data1$Online)
data1$CreditCard <- as.factor(data1$CreditCard)
is.factor(data1$Personal.Loan)
```

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

```
is.factor(data1$Online)
```

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

```
is.factor(data1$CreditCard)
```

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

Partition the collecting data

```
set.seed(400)
data_partition<-createDataPartition(data1$Personal.Loan,p=.6,list=FALSE,times=1)
train<-data1[data_partition,]
valid<-data1[-data_partition,]
head(train)
```

```
##    ID Age Experience Income ZIP.Code Family CCAvg Education Mortgage
## 1   1  25          1     49    91107      4   1.6         1        0
## 2   2  45         19     34    90089      3   1.5         1        0
```

```
## 6    6  37          13      29     92121      4   0.4          2      155
## 8    8  50          24      22     93943      1   0.3          3        0
## 9    9  35          10      81     90089      3   0.6          2      104
## 10 10  34           9     180     93023      1   8.9          3        0
##    Personal.Loan Securities.Account CD.Account Online CreditCard
## 1              0                  1          0      0          0
## 2              0                  1          0      0          0
## 6              0                  0          0      1          0
## 8              0                  0          0      0          1
## 9              0                  0          0      1          0
## 10             1                  0          0      0          0
```

```
head(valid)
```

```
##    ID Age Experience Income ZIP.Code Family CCAvg Education Mortgage
## 3   3  39         15     11    94720      1   1.0         1        0
## 4   4  35          9    100    94112      1   2.7         2        0
## 5   5  35          8     45    91330      4   1.0         2        0
## 7   7  53         27     72    91711      2   1.5         2        0
## 11 11  65         39    105    94710      4   2.4         3        0
## 12 12  29          5     45    90277      3   0.1         2        0
##    Personal.Loan Securities.Account CD.Account Online CreditCard
## 3              0                  0          0      0          0
## 4              0                  0          0      0          0
## 5              0                  0          0      0          1
## 7              0                  0          0      1          0
## 11             0                  0          0      0          0
## 12             0                  0          0      1          0
```

Normalizing the data

```
norm <- preProcess(train[,-c(10,13,14)], method=c("center","scale"))
train_norm <-predict(norm,train)
valid_norm<- predict(norm,valid)
head(valid_norm)
```

```
##           ID        Age Experience       Income   ZIP.Code     Family      CCAvg
## 3  -1.725518 -0.5561250 -0.4479061 -1.36985086  0.6818904 -1.2265941 -0.5266860
## 4  -1.724820 -0.9044833 -0.9696035  0.56597031  0.4205022 -1.2265941  0.4526340
## 5  -1.724122 -0.9044833 -1.0565530 -0.63032367 -0.7755203  1.3887451 -0.5266860
## 7  -1.722726  0.6631291  0.5954885 -0.04305208 -0.6117228 -0.3548144 -0.2386507
## 11 -1.719934  1.7082040  1.6388832  0.67472431  0.6775912  1.3887451  0.2798128
## 12 -1.719235 -1.4270208 -1.3174017 -0.63032367 -1.2282205  0.5169654 -1.0451495
##     Education   Mortgage Personal.Loan Securities.Account CD.Account Online
## 3  -1.0591770 -0.5619176             0         -0.3448852 -0.2442874      0
## 4   0.1282412 -0.5619176             0         -0.3448852 -0.2442874      0
## 5   0.1282412 -0.5619176             0         -0.3448852 -0.2442874      0
## 7   0.1282412 -0.5619176             0         -0.3448852 -0.2442874      1
## 11  1.3156593 -0.5619176             0         -0.3448852 -0.2442874      0
## 12  0.1282412 -0.5619176             0         -0.3448852 -0.2442874      1
```

```
##    CreditCard
## 3           0
## 4           0
## 5           1
## 7           0
## 11          0
## 12          0
```

```
head(train_norm)
```

```
##           ID         Age Experience     Income    ZIP.Code     Family
## 1  -1.726914 -1.77537906 -1.6651999 -0.5433205 -0.87139129  1.3887451
## 2  -1.726216 -0.03358755 -0.1001079 -0.8695825 -1.30904441  0.5169654
## 6  -1.723424 -0.73030415 -0.6218053 -0.9783365 -0.43545783  1.3887451
## 8  -1.722028  0.40186033  0.3346399 -1.1305921  0.34784667 -1.2265941
## 9  -1.721330 -0.90448330 -0.8826539  0.1527051 -1.30904441  0.5169654
## 10 -1.720632 -0.99157288 -0.9696035  2.3060343 -0.04767481 -1.2265941
##         CCAvg  Education   Mortgage Personal.Loan Securities.Account CD.Account
## 1  -0.1810436 -1.0591770 -0.5619176             0          2.8985492 -0.2442874
## 2  -0.2386507 -1.0591770 -0.5619176             0          2.8985492 -0.2442874
## 6  -0.8723283  0.1282412  0.9264659             0         -0.3448852 -0.2442874
## 8  -0.9299354  1.3156593 -0.5619176             0         -0.3448852 -0.2442874
## 9  -0.7571142  0.1282412  0.4367397             0         -0.3448852 -0.2442874
## 10  4.0242717  1.3156593 -0.5619176             1         -0.3448852 -0.2442874
##    Online CreditCard
## 1       0          0
## 2       0          0
## 6       1          0
## 8       0          1
## 9       1          0
## 10      0          0
```

A. Generate a pivot table for the training data utilising Online as a column variable, CC and Loan as Variable

```
table_loan<-table(train_norm$CreditCard,train_norm$Personal.Loan,train_norm$Online)
View(table_loan)
```

B.The probability of loan acceptance (Loan = 1) conditional on having a bank credit card (CC = 1) and being an active user of online banking services (Online = 1) = (47 /(473+47))=0.09038461538

C.Making two pivot table for training data that can assist Online with Loan & Credit Card with Loan

1. Online with Loan

```
table_1<-table(train_norm$Personal.Loan,train_norm$Online)
View(table_1)
```

2. CreditCard with Loan

```
table_2<-table(train_norm$Personal.Loan,train_norm$CreditCard)
View(table_2)
```

loan =1 probabaility of loan

```
table_3<-table(train_norm$Personal.Loan)
prop<-prop.table(table_3)
View(prop)
```

D. the table Caculations Based on Pivot Tables, Individual Probabilities CC Given Loan

i. P(CC = 1 | Loan = 1) (the proportion of credit card holders among the loan acceptors) =85/288=0.2951

ii. P(Online = 1 | Loan = 1)= 171/288=0.5937

iii. P(Loan = 1) (the proportion of loan acceptors) =0.096

iv. P(CC = 1 | Loan = 0) =799/(799+1913)=0.2946

v. P(Online = 1 | Loan = 0) =1629/(1629+1083)=0.6006

vi. P(Loan = 0) =0.904

E. Naive Bayes

Naive Bayes = P(Loan=1/CC=1,Online=1) = P(CC=1/Loan=1)P(Online=1/Loan=1)P(Loa =0.09514777745

Naive Bayes Probability is 0.09514777745

F. Compared of probabilities using a pivot table and naive bayes probability

The Probability obtained using pivot table is 0.09038461538

The Probability obtained using Naive bayes formula is 0.09514777745

Calculation of increase the actual value of probability is 0.00476316207

G.Applying the above-calculated values to determine the naive Bayes probability P(Loan = 1 | CC = 1, Online = 1)

```
model<-naiveBayes(Personal.Loan~CreditCard+Online,data=train_norm)

model
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
```

```
## Y
##     0     1
## 0.904 0.096
##
## Conditional probabilities:
##    CreditCard
## Y           0         1
##   0 0.7053835 0.2946165
##   1 0.7048611 0.2951389
##
##    Online
## Y           0         1
##   0 0.3993363 0.6006637
##   1 0.4062500 0.5937500
```