Assignment 03-Naive Bayes

Chathurani Ekanayake

2023-03-05

```
library(readxl)
UniversalBank <- read_excel("F:/1st sem/ML/Assignment 03/UniversalBank.xlsx")
View(UniversalBank)
```

install.packages(class) install.packages(e1071) install.packages("reshape") install.packages("reshape2") install.packages("reshape2") install.packages("gmodels") install.packages("fastDummies")

```
library(class)
library(e1071)
library(reshape)
```

```
##
## Attaching package: 'reshape'
```

```
## The following object is masked from 'package:class':
##
##     condense
```

```
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following objects are masked from 'package:reshape':
##
##     colsplit, melt, recast
```

```
library(gmodels)
library(fastDummies)
library(caret)
```

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

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

```
library(readr)
library(dplyr)
```

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

```
## The following object is masked from 'package:reshape':
##
##     rename
```

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

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

```
library(ISLR)
```

#Select the required variables

```
UniversalBank<-UniversalBank[,c(10,13,14)]
t(t(names(UniversalBank)))
```

```
##      [,1]
## [1,] "Personal Loan"
## [2,] "Online"
## [3,] "CreditCard"
```

#Change the column names

```
colnames(UniversalBank)[1]="Loan"
colnames(UniversalBank)[3]="CC"
t(t(names(UniversalBank)))
```

```
##      [,1]
## [1,] "Loan"
## [2,] "Online"
## [3,] "CC"
```

#use 60% of the data in training set & 40% for validation set

```
set.seed(123)
Train_data<-createDataPartition(UniversalBank$Loan,p=.6, list = FALSE)
Train_UB1<-UniversalBank[Train_data,]
```

```
Valid_UB1=UniversalBank[-Train_data, ]
attach(UniversalBank)
```

A. Create a Pivot table

```
ftable(CC,Loan,Online)
```

```
##         Online    0    1
## CC Loan
## 0  0           1300 1893
##    1            128  209
## 1  0            527  800
##    1             61   82
```

B. The probability of customer accepts loan offer

```
P0<-82/882
```

C. Creating two pivot tables: (Loan, Online) & (Loan,CC)

```
addmargins(table(Loan,Online))
```

```
##      Online
## Loan     0    1  Sum
##   0   1827 2693 4520
##   1    189  291  480
##   Sum 2016 2984 5000
```

```
addmargins(table(Loan,CC))
```

```
##      CC
## Loan     0    1  Sum
##   0   3193 1327 4520
##   1    337  143  480
##   Sum 3530 1470 5000
```

D. Calculate different probabilities of A given B

```
##1. p(CC=1| Loan = 1)
P1 <- 143/480

##2. P(Online = 1 | Loan = 1)
P2 <- 291/480

##3.P(Loan = 1)
addmargins(table(Loan))
```

```
## Loan
##    0    1  Sum
## 4520  480 5000
```

```
p3 <- 480/5000

##4.P(CC = 1 | Loan = 0)
p4 <-1327/4520

##5. P(Online = 1 | Loan = 0)
p5<-2693/4520

##6.P(Loan = 0)
p6<-4520/5000
```

E. Naive Bayes of y P(Loan = 1 | CC = 1, Online = 1).

```
P7 = ((P1*P2)*p3/((P1*P2)*p3+(p4*p5)*p6))
P7
```

```
## [1] 0.09881706
```

F. Compare this value with the one obtained from the pivot table in (b)

```
#Answer: The value for P7 = 0.098 & P0 = 0.092 Due to decimals those two values are sligh
tly different, the two values for E. and B. are almost same.Naive Bayes method does not n
eeded the exact same independent variable as do by the other method of B.So this cause fo
r this slight difference of values.
```

G. Run naive Bayes on the data & Compare this to the number you obtained in (e).

```
P8 <- naiveBayes(Loan~.,data = Train_UB1)
P8
```

```
##
## 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:
##    Online
## Y        [,1]      [,2]
##   0 0.5951506 0.4909531
##   1 0.6438849 0.4797134
##
##    CC
## Y        [,1]      [,2]
##   0 0.2909625 0.4542897
##   1 0.3273381 0.4700881
```

```
# Answer :Here, the Naive Bayes provides the same answer as in question (E) .09
```