FML_Assignment3

Sandhya Cheepurupalli

2022-10-16

Following Packages are required to run the following code

```
library(caret)
```

```
## Loading required package: ggplot2

## Loading required package: lattice
```

```
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(ISLR)
library(e1071)
```

Step-1 Reading the dataset

```
setwd("C:\\Users\\sandh\\Downloads")
mydata<-read.csv(file='UniversalBank.csv')
mydata$Personal.Loan<-as.factor(mydata$Personal.Loan)
mydata$Online<-as.factor(mydata$Online)#converting to factor
mydata$CreditCard<-as.factor(mydata$CreditCard)#converting to factor
```

Step-2 Removing unwanted data from the dataset

```
mydata1<- mydata %>% select(Personal.Loan, Online, CreditCard)
head(mydata1)
```

```
##   Personal.Loan Online CreditCard
## 1             0      0          0
## 2             0      0          0
## 3             0      0          0
## 4             0      0          0
## 5             0      0          1
## 6             0      1          0
```

Step-3 Partitioning dataset into training data and validation data

```
set.seed(7)
Train_Index = createDataPartition(mydata1$Personal.Loan,p=0.60, list=FALSE)
# 60% of remaining data as validation
Train_Data = mydata1[Train_Index,]
Validation_Data = mydata1[-Train_Index,]
head(Train_Data)
```

```
##    Personal.Loan Online CreditCard
## 2              0      0          0
## 6              0      1          0
## 7              0      1          0
## 8              0      0          1
## 9              0      1          0
## 10             1      0          0
```

Question-A Creating Pivot Table with Online as Column and CC & Loan as rows

```
table1<- ftable(Train_Data$Online,Train_Data$Personal.Loan,Train_Data$CreditCard,
                row.vars = c(2,3),dnn=c('Online','Personal.loan','CreditCard'))
table1
```

```
##                          Online    0    1
## Personal.loan CreditCard                 
## 0             0                  779 1133
##               1                  313  487
## 1             0                   74  128
##               1                   33   53
```

Question-B Finding the Probability for P(Loan=1 | P(CC=1), P(Online=1)) from the above Pivot Table

```r
P.loan<-table1[4,2]/(table1[2,2]+table1[4,2])
P.loan
```

```
## [1] 0.09814815
```

Question-C Creating two Pivot Tables - One for Loan as Rows and Online as Columns - Other for Loan as Rows and CC as Columns

```r
table2<- ftable(Train_Data$Online,Train_Data$Personal.Loan,row.vars = c(2),
                dnn=c('Online','Personal.loan'))
table2
```

```
##                Online    0    1
## Personal.loan
## 0                     1092 1620
## 1                      107  181
```

```r
table3<- ftable(Train_Data$CreditCard,Train_Data$Personal.Loan,row.vars = c(2),
                dnn=c('CreditCard','Personal.loan'))
table3
```

```
##                CreditCard    0    1
## Personal.loan
## 0                         1912  800
## 1                          202   86
```

Question-D Finding Probabilities for P(CC=1 | Loan =1), P (Online=1 | Loan=1), P(Loan=1), P(CC=1 | Loan = 0), P(Online=1 | Loan=0), P(Loan=0)

```r
p.CC1.loan1<-table3[2,2]/(table3[2,1]+table3[2,2])
p.CC1.loan1
```

```
## [1] 0.2986111
```

```r
p.online1.loan1<-table2[2,2]/(table2[2,1]+table2[2,2])
p.online1.loan1
```

```
## [1] 0.6284722
```

```r
p.CC1.loan0<-table3[1,2]/(table3[1,1]+table3[1,2])
p.CC1.loan0
```

```
## [1] 0.2949853
```

```
p.online1.loan0<-table2[1,2]/(table2[1,1]+table2[1,2])
p.online1.loan0
```

```
## [1] 0.5973451
```

```
p.loan1<-filter(Train_Data, Personal.Loan==1) %>% count()
p.loan1<-p.loan1[[1]]/3000
p.loan1
```

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

```
p.loan0<-filter(Train_Data, Personal.Loan==0) %>% count()
p.loan0<-p.loan0[[1]]/3000
p.loan0
```

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

Question-E Finding Naive Bayes Probability for P(Loan=1 | CC=1, Online=1) using the Above Probabilities

```
Naive.Bayes<-(p.CC1.loan1*p.online1.loan1*p.loan1)/((p.CC1.loan1*p.online1.loan1*p.loan1)+
                                                     (p.CC1.loan0*p.online1.loan0*p.loan0))
Naive.Bayes
```

```
## [1] 0.1016095
```

Question-F Comparing value obtained from Question-B and Question-E

The probability of P(Loan=1 | CC=1, Online=1) from Question-B is 0.09814815 while from Question-E is 0.1016095.

$$\text{It is clearly seen that the probability of Question-E} > \text{Question-B.}$$

This difference is because, we have considered CC and Online as conditionally independent variables in Question-E.

The probability of B is more accurate as we have not considered CC and Online as conditionally independent variables.

Question-G Finding Probability of P(Loan=1 | CC=1, Online=1) using Naive Bayes Function

```
Question.G<-naiveBayes(Personal.Loan ~ ., data = Train_Data)
Question.G
```

```
##
## 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:
##    Online
## Y           0         1
##   0 0.4026549 0.5973451
##   1 0.3715278 0.6284722
##
##    CreditCard
## Y           0         1
##   0 0.7050147 0.2949853
##   1 0.7013889 0.2986111
```

```r
Ans<-predict(Question.G,Validation_Data, type="raw")
head(Ans)
```

```
##              0          1
## [1,] 0.9111772 0.08882284
## [2,] 0.9111772 0.08882284
## [3,] 0.9111772 0.08882284
## [4,] 0.9097610 0.09023897
## [5,] 0.9111772 0.08882284
## [6,] 0.8999654 0.10003456
```

It is found that the conditional probabilities from Question-D and Question G are equal. Which can be interpreted as Naive Bayes probabilities from Question E and G will also be equal.