Assignment_3

Venkata Naga Siddartha Gutha

2022-10-17

```
library(caret)
```

```
## Loading required package: ggplot2

## Loading required package: lattice
```

```
library(e1071)
library(ISLR)
library(reshape2)
#loading data set
dataset_ub<-read.csv("C:/Users/sidda/Downloads/UniversalBank.csv")
#converting variables into factor
dataset_ub$Personal.Loan<-factor(dataset_ub$Personal.Loan)
dataset_ub$Online<-factor(dataset_ub$Online)
dataset_ub$CreditCard<-factor(dataset_ub$CreditCard)

#partition of data in to training and validation sets
set.seed(555)
train<-createDataPartition(dataset_ub$Personal.Loan,p=0.6,list = FALSE)
train_set<-dataset_ub[train,]
validation_set<-dataset_ub[-train,]
nrow(train_set)
```

```
## [1] 3000
```

```
nrow(validation_set)
```

```
## [1] 2000
```

```
#Question A
table<-xtabs(~CreditCard+Personal.Loan+Online,data=train_set)
ftable(table)
```

```
##                          Online    0    1
## CreditCard Personal.Loan
## 0          0                     788 1131
##            1                      69  121
## 1          0                     310  483
##            1                      44   54
```

```r
#Question B

54/(54+483)
```

```
## [1] 0.1005587
```

```r
#Question c
table(Personal.Loan=train_set$Personal.Loan,
      Online=train_set$Online)
```

```
##              Online
## Personal.Loan    0    1
##             0 1098 1614
##             1  113  175
```

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

```
##              CreditCard
## Personal.Loan    0    1
##             0 1919  793
##             1  190   98
```

```r
table(Personal.Loan=train_set$Personal.Loan)
```

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

```r
#Question D
#i.P(CC = 1 | Loan = 1)
P1=98/(98+190)
P1
```

```
## [1] 0.3402778
```

```r
#ii. P(Online = 1 | Loan = 1)
P2=175/(175+113)
P2
```

```
## [1] 0.6076389
```

```r
#iii. P(Loan = 1)
P3=288/(288+2712)
P3
```

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

```r
#iv. P(CC = 1 | Loan = 0)
P4=793/(793+1919)
P4
```

```
## [1] 0.2924041
```

```r
#v. P(Online = 1 | Loan = 0)
P5=1614/(1614+1098)
P5
```

```
## [1] 0.5951327
```

```r
#vi. P(Loan = 0)
P6=2712/(288+2712)
P6
```

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

```r
#Question E
#the naive Bayes probability P(Loan = 1 | CC = 1, Online = 1).
(P1*P2*P3)/((P1*P2*P3)+(P4*P5*P6))
```

```
## [1] 0.1120411
```

##Question F

probability from pivot table is 0.1005587 and the naive Bayes probability is 0.1120411. Naive Bayes makes an assumption that attributes are independent of each other. So probability from pivot table is more accurate than the naive Bayes probability

```r
#Question G
Naive_model<-naiveBayes(Personal.Loan~CreditCard+Online,data = train_set)
test<-data.frame(CreditCard=1,Online=1)
test$CreditCard<-factor(test$CreditCard)
test$Online<-factor(test$Online)
predict(Naive_model,test,type = 'raw')
```

```
##              0         1
## [1,] 0.8879589 0.1120411
```

Probability of test data is same as the probability obtained in the question E which is equal to 0.1120411. This implies that Naive bayes algorithm has predicted same as that of calculated probability