Assignment 3

Abi

2022-10-17

library(readr)  
UniversalBank <- read.csv("C:/Users/abinaya/Downloads/UniversalBank.csv")  
View(UniversalBank)  
summary(UniversalBank)

## 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

Converting into factor:

UniversalBank$Personal.Loan<-as.factor(UniversalBank$Personal.Loan)  
UniversalBank$Online<-as.factor(UniversalBank$Online)  
UniversalBank$CreditCard<-as.factor(UniversalBank$CreditCard)  
summary(UniversalBank)

## 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 Personal.Loan  
## Min. :1.000 Min. : 0.000 Min. :1.000 Min. : 0.0 0:4520   
## 1st Qu.:1.000 1st Qu.: 0.700 1st Qu.:1.000 1st Qu.: 0.0 1: 480   
## 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   
## Securities.Account CD.Account Online CreditCard  
## Min. :0.0000 Min. :0.0000 0:2016 0:3530   
## 1st Qu.:0.0000 1st Qu.:0.0000 1:2984 1:1470   
## Median :0.0000 Median :0.0000   
## Mean :0.1044 Mean :0.0604   
## 3rd Qu.:0.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000

Loading Packages:

library("caret")

## Loading required package: ggplot2

## Loading required package: lattice

library('class')  
library('ISLR')

## Question A

## Partitioning the data into training (60%) and validation set(40%)  
set.seed(215)  
Train\_Index = createDataPartition(UniversalBank$Personal.Loan,p=0.6, list=FALSE)   
Train.df=UniversalBank[Train\_Index,]  
Validation.df=UniversalBank[-Train\_Index,]

## A Pivot table for the training data with Online as a column variable, CC as a rowvariable, and Loan as a secondary row variable.

mytable <- xtabs(~ Online+CreditCard+Personal.Loan, data =Train.df)  
ftable(mytable)

## Personal.Loan 0 1  
## Online CreditCard   
## 0 0 757 82  
## 1 308 34  
## 1 0 1158 125  
## 1 489 47

## Question B

# Probability of customer accepting loan with the condition on having credit card and using online services = 47/(47+489)= 0.08768

## Question C

# Pivot tables for the loan and online ,loan and creditcard

table(Online=Train.df$Online, Personal.Loan=Train.df$Personal.Loan)

## Personal.Loan  
## Online 0 1  
## 0 1065 116  
## 1 1647 172

table(CreditCard=Train.df$CreditCard, Personal.Loan=Train.df$Personal.Loan)

## Personal.Loan  
## CreditCard 0 1  
## 0 1915 207  
## 1 797 81

## Question D

# Computig the quantities [P(A | B) means “the probability ofA given B”]:

## Question E

## Naive Bayes probability P(Loan = 1 | CC = 1, Online = 1).

# Naive bayes formula p(Y/x1,….xn) = p(x1,x2,…xn)/y\*p(y)/p(x1,x2,…xn)

# y=loan1

# x1= creditcard1

# x2= online1

# p(creditcard1(x1))= (797+81)/(797+81+1915+207)=0.29266 # p(online1(x2)) = (1647+172)/(1647+172+1065+116)=0.6063

# p(loan1/creditcard1,online1)= p(creditcard1/loan1)*p(online1/loan1)*p(loan)

## /p(creditcard1)\*p(online1)

= 0.28125\*0.5972222\*0.096/0.29266\*0.6063  
  
 = 0.09087591

## Question F

# \* Value obtained in pivot table is 0.08768 and naive bayes probabilty is 0.09087, both values are merely same but the most accurate estimate is table value because of the assumption that we made in naive bayes as variables are independent.

## Question G

# Entries for computing g P(Loan = 1 | CC = 1, Online = 1)

library(e1071)  
nb.model<-naiveBayes (Personal.Loan~Online+CreditCard, data=Train.df)  
To\_Predict=data.frame(Online='1', CreditCard='1')  
predict(nb.model,To\_Predict,type='raw')

## 0 1  
## [1,] 0.9091368 0.09086319

# \* Value obtained from naive bias formula (from question E) and the value obtained from computing naive bias by entry are same.