Assignment 3

Sai Prasad

2024-03-04

Importing packages of caret, e1071, ISLR, reshape2,table, melt, cast and dataset.

library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

library(e1071)  
library(ISLR)  
library(tables)

## Warning: package 'tables' was built under R version 4.3.3

library(melt)

## Warning: package 'melt' was built under R version 4.3.3

library(CAST)

## Warning: package 'CAST' was built under R version 4.3.3

# Here I had created dataset as dataset\_universal   
dataset\_universal<-read.csv("C:\\Users\\sidda\\Downloads\\UniversalBank (1).csv")  
#converting variables into factor  
dataset\_universal$Personal.Loan<-factor(dataset\_universal$Personal.Loan)  
dataset\_universal$Online<-factor(dataset\_universal$Online)  
dataset\_universal$CreditCard<-factor(dataset\_universal$CreditCard)  
  
# Here I was doing partition of data into training and validation sets. And named as training\_set and vaslidating\_set   
set.seed(5097)  
train\_universal<-createDataPartition(dataset\_universal$Personal.Loan,p=0.6,list = FALSE)  
training\_set<-dataset\_universal[train\_universal,]  
validating\_set<-dataset\_universal[-train\_universal,]  
nrow(training\_set)

## [1] 3000

nrow(validating\_set)

## [1] 2000

# Question A

Creating a pivot table for the training data with Online as a column variable, CC as a row variable, and Loan as a secondary row variable. The values inside the table should convey the count.

table<-xtabs(~CreditCard+Personal.Loan+Online,data=training\_set)  
ftable(table)

## Online 0 1  
## CreditCard Personal.Loan   
## 0 0 765 1176  
## 1 78 121  
## 1 0 310 461  
## 1 36 53

# Question B

Consider the task of classifying a customer who owns a bank credit card and is acƟvely using online banking services. Looking at the pivot table, what is the probability that this customer will accept the loan offer? [This is the probability of loan acceptance (Loan = 1) condiƟonal on having a bank credit card (CC = 1) and being an acƟve user of online banking services (Online = 1)].

53/(53+461)

## [1] 0.1031128

# Quetion C

Create two separate pivot tables for the training data. One will have Loan (rows) as a funcƟon of Online (columns) and the other will have Loan (rows) as a funcƟon of CC.

table(Personal.Loan=training\_set$Personal.Loan,  
 Online=training\_set$Online)

## Online  
## Personal.Loan 0 1  
## 0 1075 1637  
## 1 114 174

table(Personal.Loan=training\_set$Personal.Loan,  
 CreditCard=training\_set$CreditCard)

## CreditCard  
## Personal.Loan 0 1  
## 0 1941 771  
## 1 199 89

table(Personal.Loan=training\_set$Personal.Loan)

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

#Question D. Compute the following quantities [P(A | B) means “the probability ofA given B”]: i. P(CC = 1 | Loan = 1) (the proporƟon of credit card holders among the loan acceptors) ii. P(Online = 1 | Loan = 1) iii. P(Loan = 1) (the proporƟon of loan acceptors) iv. P(CC = 1 | Loan = 0) v. P(Online = 1 | Loan = 0) vi. P(Loan = 0)

#i.P(CC = 1 | Loan = 1)  
P1=89/(89+199)  
P1

## [1] 0.3090278

#ii. P(Online = 1 | Loan = 1)   
P2=174/(174+114)  
P2

## [1] 0.6041667

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

## [1] 0.096

#iv. P(CC = 1 | Loan = 0)   
P4=771/(771+1941)  
P4

## [1] 0.284292

#v. P(Online = 1 | Loan = 0)   
P5=1637/(1637+1075)  
P5

## [1] 0.6036136

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

## [1] 0.904

#Question E Use the quanƟƟes computed above to compute the naive Bayes probability P(Loan = 1 | CC = 1, Online = 1).

(P1\*P2\*P3)/((P1\*P2\*P3)+(P4\*P5\*P6))

## [1] 0.1035734

#Question F Compare this value with the one obtained from the pivot table in (B). Which is a more accurate estimate? #Answer The likelihood obtained from the pivot table is 0.1031128, while the probability calculated through naive Bayes is 0.1035734. Naive Bayes operates under the assumption of attribute independence, which can affect its accuracy. Therefore, the probability derived from the pivot table is deemed to be more precise than that from naive Bayes.

#Question G Which of the entries in this table are needed for compuƟng P(Loan = 1 | CC = 1, Online = 1)? Run naive Bayes on the data. Examine the model output on training data, and find the entry that corresponds to P(Loan = 1 | CC = 1, Online = 1). Compare this to the number you obtained in (E). Here we are creating a data frame named as test\_data. And converting into factors

Naive\_bayes\_model<-naiveBayes(Personal.Loan~CreditCard+Online,data = training\_set)  
test\_data<-data.frame(CreditCard=1,Online=1)  
test\_data$CreditCard<-factor(test\_data$CreditCard)  
test\_data$Online<-factor(test\_data$Online)  
predict(Naive\_bayes\_model,test\_data,type = 'raw')

## 0 1  
## [1,] 0.8964266 0.1035734