ASSIGNMENT 3(FML)

Sri Chandana

2023-11-06

# It imports the "Online\_Retail" dataset from the specified file path  
  
Uni\_Bank1 <- read.csv("C:/Users/srich/OneDrive/Desktop/R programming/UniversalBank-1.csv")

# It converts the 'Personal.Loan' column in the 'Uni\_Bank1' data frame to a factor  
# It converts the 'Online' column in the 'Uni\_Bank1' data frame to a factor.  
# It converts the 'CreditCard' column in the 'Uni\_Bank1' data frame to a factor.  
Uni\_Bank1$Personal.Loan <- as.factor(Uni\_Bank1$Personal.Loan)  
Uni\_Bank1$Online <- as.factor(Uni\_Bank1$Online)  
Uni\_Bank1$CreditCard <- as.factor(Uni\_Bank1$CreditCard)

summary(Uni\_Bank1)

## 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

# Loads all the mentioned libraries  
library(ggplot2)  
library(caret)

## Warning: package 'caret' was built under R version 4.3.2

## Loading required package: lattice

library(reshape2)  
library(melt)

## Warning: package 'melt' was built under R version 4.3.2

library(ISLR)  
library(class)

# Set a random seed for reproducibility  
set.seed(23)  
# Using the 'createDataPartition' function to split the 'Personal.Loan' column of the 'Uni\_Bank1' data frame.  
splitIndex <- createDataPartition(Uni\_Bank1$Personal.Loan, p = 0.6, list = FALSE)  
# It creates the training dataset by selecting the rows indicated by 'splitIndex'  
# It creates the validation dataset by selecting the rows that are not in the 'splitIndex'  
  
train\_data <- Uni\_Bank1[splitIndex, ]  
valid\_data <- Uni\_Bank1[-splitIndex, ]

# It checks the dimensions of the training and validation dataset.  
dim(train\_data)

## [1] 3000 14

dim(valid\_data)

## [1] 2000 14

# A) Create 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.  
Table1 <- ftable(train\_data[,c(14,10,13)])  
Table1

## Online 0 1  
## CreditCard Personal.Loan   
## 0 0 773 1127  
## 1 82 114  
## 1 0 315 497  
## 1 39 53

# B)Consider the task of classifying a customer who owns a bank credit card and is actively using online banking services. Looking at the pivot table, what is the probability that this customer will accept the loan offer?  
  
# According to the pivot table, it is possible to calculate the probability of a customer accepting the loan offer as 53 / (53 + 497), which equals 0.096.

# C) Create two separate pivot tables for the training data. One will have Loan (rows) as a function of Online (columns) and the other will have Loan (rows) as a function of CC  
  
# Two separate pivot tables were created using the training data.  
  
melt1 <- melt(train\_data,id=c("Personal.Loan"),variable="Online")

## Warning: attributes are not identical across measure variables; they will be  
## dropped

melt2 <- melt(train\_data,id=c("Personal.Loan"), variable="CreditCard")

## Warning: attributes are not identical across measure variables; they will be  
## dropped

cast1 = dcast(melt1, Personal.Loan~Online)

## Aggregation function missing: defaulting to length

cast2 <- dcast(melt2, Personal.Loan~CreditCard)

## Aggregation function missing: defaulting to length

# D) Compute the following quantities [P(A | B) means “the probability of A given B”:  
  
ftable(train\_data[,c(10,13)])

## Online 0 1  
## Personal.Loan   
## 0 1088 1624  
## 1 121 167

ftable(train\_data[,c(10,14)])

## CreditCard 0 1  
## Personal.Loan   
## 0 1900 812  
## 1 196 92

#1. P(CC = 1 | Loan = 1) = (92/92+196) = 0.319  
#2. P(Online = 1 | Loan = 1) = (167/167+121) = 0.579  
#3. P(Loan = 1) = (288/288+2712) = 0.096  
#4. P(CC = 1 | Loan = 0) = (812/812+1900) = 0.299  
#5. P(Online = 1 | Loan = 0) = (1624/ 1624+1088) = 0.598  
#6. P(Loan = 0) = (2712/ 2712+288) = 0.904

# E) Use the quantities computed above to compute the naive Bayes probability P(Loan = 1 | CC = 1, Online = 1)  
  
# Probability P(Loan = 1 | CC = 1, Online = 1)  
# (0. 319\* 0.579\* 0.096) / (0.319 \* 0.579 \* 0.096) + (0.299 \* 0.598 \* 0.904) = 0.098

# F) Compare this value with the one obtained from the pivot table in (B). Which is a more accurate estimate?  
  
# The probability value calculated in part B is 0.096, and in the current question, it's approximately 0.098. These values have a very small difference. The calculation in part B is more accurate and specific, as it was based on less complex and dependent information.

# G) Which of the entries in this table are needed for computing 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).  
  
library(naivebayes)

## naivebayes 0.9.7 loaded

naive\_b<- naive\_bayes(Personal.Loan~Online+CreditCard,data=train\_data)  
naive\_b

##   
## ================================== Naive Bayes ==================================   
##   
## Call:   
## naive\_bayes.formula(formula = Personal.Loan ~ Online + CreditCard,   
## data = train\_data)  
##   
## ---------------------------------------------------------------------------------   
##   
## Laplace smoothing: 0  
##   
## ---------------------------------------------------------------------------------   
##   
## A priori probabilities:   
##   
## 0 1   
## 0.904 0.096   
##   
## ---------------------------------------------------------------------------------   
##   
## Tables:   
##   
## ---------------------------------------------------------------------------------   
## ::: Online (Bernoulli)   
## ---------------------------------------------------------------------------------   
##   
## Online 0 1  
## 0 0.4011799 0.4201389  
## 1 0.5988201 0.5798611  
##   
## ---------------------------------------------------------------------------------   
## ::: CreditCard (Bernoulli)   
## ---------------------------------------------------------------------------------   
##   
## CreditCard 0 1  
## 0 0.7005900 0.6805556  
## 1 0.2994100 0.3194444  
##   
## ---------------------------------------------------------------------------------