Assignment 3

SAI ROHITH GUDA

10-14-2023

### Task a

# Load necessary libraries  
library(reshape2)  
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

# Load necessary library  
library(dplyr)  
  
# Read the CSV file  
data <- read.csv("UniversalBank.csv")  
  
# Select the relevant columns  
selected\_data <- data %>% select(Online, CreditCard, Personal.Loan)  
  
# Set seed for reproducibility  
set.seed(42)  
  
# Split data into training and validation sets  
sample\_size <- floor(0.6 \* nrow(selected\_data))  
train\_index <- sample(seq\_len(nrow(selected\_data)), size = sample\_size)  
  
train\_data <- selected\_data[train\_index, ]  
validation\_data <- selected\_data[-train\_index, ]  
  
pivot\_table <- table(train\_data$CreditCard, train\_data$Online, train\_data$Personal.Loan)  
print(paste("Pivot table for CreditCard, Online, and Personal.Loan:\n", pivot\_table))

## [1] "Pivot table for CreditCard, Online, and Personal.Loan:\n 803"   
## [2] "Pivot table for CreditCard, Online, and Personal.Loan:\n 312"   
## [3] "Pivot table for CreditCard, Online, and Personal.Loan:\n 1125"  
## [4] "Pivot table for CreditCard, Online, and Personal.Loan:\n 472"   
## [5] "Pivot table for CreditCard, Online, and Personal.Loan:\n 77"   
## [6] "Pivot table for CreditCard, Online, and Personal.Loan:\n 32"   
## [7] "Pivot table for CreditCard, Online, and Personal.Loan:\n 131"   
## [8] "Pivot table for CreditCard, Online, and Personal.Loan:\n 48"

### Task b

prob\_acceptance <- pivot\_table[2, 2, 2] / (pivot\_table[2, 2, 1] + pivot\_table[2, 2, 2])  
print(paste("Probability of loan acceptance given CC=1 and Online=1:", round(prob\_acceptance, 4)))

## [1] "Probability of loan acceptance given CC=1 and Online=1: 0.0923"

### Task c

pivot\_loan\_online <- table(train\_data$Personal.Loan, train\_data$Online)  
  
pivot\_loan\_cc <- table(train\_data$Personal.Loan, train\_data$CreditCard)  
  
print("Pivot table for Personal.Loan and Online:")

## [1] "Pivot table for Personal.Loan and Online:"

print(pivot\_loan\_online)

##   
## 0 1  
## 0 1115 1597  
## 1 109 179

print("\nPivot table for Personal.Loan and CreditCard:")

## [1] "\nPivot table for Personal.Loan and CreditCard:"

print(pivot\_loan\_cc)

##   
## 0 1  
## 0 1928 784  
## 1 208 80

### Task d

# P(CC = 1 | Loan = 1)  
p\_cc\_given\_loan1 <- pivot\_loan\_cc[2, 2] / (pivot\_loan\_cc[2, 1] + pivot\_loan\_cc[2, 2])  
  
# P(Online = 1 | Loan = 1)  
p\_online\_given\_loan1 <- pivot\_loan\_online[2, 2] / (pivot\_loan\_online[2, 1] + pivot\_loan\_online[2, 2])  
  
# P(Loan = 1)  
p\_loan1 <- sum(train\_data$Personal.Loan == 1) / nrow(train\_data)  
  
# P(CC = 1 | Loan = 0)  
p\_cc\_given\_loan0 <- pivot\_loan\_cc[1, 2] / (pivot\_loan\_cc[1, 1] + pivot\_loan\_cc[1, 2])  
  
# P(Online = 1 | Loan = 0)  
p\_online\_given\_loan0 <- pivot\_loan\_online[1, 2] / (pivot\_loan\_online[1, 1] + pivot\_loan\_online[1, 2])  
  
# P(Loan = 0)  
p\_loan0 <- sum(train\_data$Personal.Loan == 0) / nrow(train\_data)  
  
print(paste("P(CC = 1 | Loan = 1):", round(p\_cc\_given\_loan1, 4)))

## [1] "P(CC = 1 | Loan = 1): 0.2778"

print(paste("P(Online = 1 | Loan = 1):", round(p\_online\_given\_loan1, 4)))

## [1] "P(Online = 1 | Loan = 1): 0.6215"

print(paste("P(Loan = 1):", round(p\_loan1, 4)))

## [1] "P(Loan = 1): 0.096"

print(paste("P(CC = 1 | Loan = 0):", round(p\_cc\_given\_loan0, 4)))

## [1] "P(CC = 1 | Loan = 0): 0.2891"

print(paste("P(Online = 1 | Loan = 0):", round(p\_online\_given\_loan0, 4)))

## [1] "P(Online = 1 | Loan = 0): 0.5889"

print(paste("P(Loan = 0):", round(p\_loan0, 4)))

## [1] "P(Loan = 0): 0.904"

### Task e

p\_cc <- sum(train\_data$CreditCard == 1) / nrow(train\_data)  
p\_online <- sum(train\_data$Online == 1) / nrow(train\_data)  
  
# Naive Bayes Probability  
p\_naive\_bayes <- (p\_cc\_given\_loan1 \* p\_online\_given\_loan1 \* p\_loan1) / (p\_cc \* p\_online)  
print(paste("Naive Bayes Probability P(Loan = 1 | CC = 1, Online = 1):", round(p\_naive\_bayes, 4)))

## [1] "Naive Bayes Probability P(Loan = 1 | CC = 1, Online = 1): 0.0972"

### Task f

comparison <- ifelse(abs(prob\_acceptance - p\_naive\_bayes) < 0.01, "similar", "different")  
print(paste("The value from the pivot table in (b) is", round(prob\_acceptance, 4),   
 "and the value from the naive Bayes estimate in (e) is", round(p\_naive\_bayes, 4),  
 ". The two values are", comparison, "."))

## [1] "The value from the pivot table in (b) is 0.0923 and the value from the naive Bayes estimate in (e) is 0.0972 . The two values are similar ."

### Task g

#install.packages("e1071")  
library(e1071)  
  
model <- naiveBayes(Personal.Loan ~ CreditCard + Online, data=train\_data)  
predicted\_probs <- predict(model, newdata=data.frame(CreditCard=1, Online=1), type="raw")  
p\_naive\_bayes\_R <- predicted\_probs[2]  
print(paste("Predicted Probability using naiveBayes for P(Loan = 1 | CC = 1, Online = 1):", round(p\_naive\_bayes\_R, 4)))

## [1] "Predicted Probability using naiveBayes for P(Loan = 1 | CC = 1, Online = 1): 0.0962"