Assignment3\_FML

Eswar Dumpa

2024-03-09

## Loading Packages

library(tinytex)

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

library(class)  
library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

library(ISLR)  
library(tidyverse)

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

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.3 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ lubridate 1.9.3 ✔ tibble 3.2.1  
## ✔ purrr 1.0.2 ✔ tidyr 1.3.0

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ purrr::lift() masks caret::lift()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(knitr)

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

library(e1071)  
library(gmodels)

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

library(ggplot2)  
library("pROC")

## Type 'citation("pROC")' for a citation.  
##   
## Attaching package: 'pROC'  
##   
## The following object is masked from 'package:gmodels':  
##   
## ci  
##   
## The following objects are masked from 'package:stats':  
##   
## cov, smooth, var

## Data Preparation

### Data Preparation

#### Importing & Cleaning Data

We are Importing Data from CSV file and cleaning

Universal\_bank <- read.csv("UniversalBank.csv")  
  
  
  
#Making decision variable into factor as it is a classification model  
Universal\_bank$Personal.Loan<-as.factor(Universal\_bank$Personal.Loan)  
  
  
#Removing unnecessary variables and rearranging the variable as per test data  
Universal\_bank <-Universal\_bank[,c("Personal.Loan","Online","CreditCard")]  
  
  
# Converting Categorical Variables to Factors  
Universal\_bank$Online<-as.factor(Universal\_bank$Online)  
Universal\_bank$CreditCard<-as.factor(Universal\_bank$CreditCard)  
head(Universal\_bank)

## Personal.Loan Online CreditCard  
## 1 0 0 0  
## 2 0 0 0  
## 3 0 0 0  
## 4 0 0 0  
## 5 0 0 1  
## 6 0 1 0

#### Data Partition and Normalization

set.seed(133)  
#Partitioning Data into 60% Training and 40% Validation  
Index\_Train<-createDataPartition(Universal\_bank$Personal.Loan, p=0.6, list=FALSE)  
  
Universal\_bank\_Train <-Universal\_bank[Index\_Train,]  
  
Universal\_bank\_Validation <-Universal\_bank[-Index\_Train,]  
print(paste("No. of rows in Train data is",nrow(Universal\_bank\_Train)))

## [1] "No. of rows in Train data is 3000"

print(paste("No. of rows in Validation data is",nrow(Universal\_bank\_Validation)))

## [1] "No. of rows in Validation data is 2000"

## A & B. Pivot Table and Direct Calculation

### A. Online with CreditCard & Personal.Loan

Building a Pivot Table with Online as column variable and Credit Card as Row Variable Along with Loan as Secondary row variable.

pivot1<-ftable(Universal\_bank\_Train$Online,Universal\_bank\_Train$CreditCard,Universal\_bank\_Train$Personal.Loan, row.vars = c(3,2),dnn=c('Online','CreditCard','Personal.loan'))  
  
print("Pivot Table for the given variables is")

## [1] "Pivot Table for the given variables is"

pivot1

## Online 0 1  
## Personal.loan CreditCard   
## 0 0 809 1109  
## 1 319 475  
## 1 0 72 137  
## 1 34 45

### B. Probability of Loan Given CC and Online

# Pivot Table from question A  
P1<- pivot1[4,2]/(pivot1[2,2]+pivot1[4,2])

$$ P(\frac{Loan=1}{CC=1,Online=1})
=\frac{45}{45+475}= 0.0865384615384615\\
Probability\hspace{.3cm} is \hspace{.3cm} 0.0865384615384615$$

## C,D,E. Pivot Tables and Naive Bayes

### C. Online with Loan & Credit Card with Loan

#### Online with Loan

pivot2<-ftable(Universal\_bank\_Train$Online, Universal\_bank\_Train$Personal.Loan,  
 row.vars = c(2),dnn=c('Online', 'Personal.loan'))  
print("Pivot Table for the given variables is")

## [1] "Pivot Table for the given variables is"

pivot2

## Online 0 1  
## Personal.loan   
## 0 1128 1584  
## 1 106 182

#### CreditCard with Loan

pivot3<-ftable(Universal\_bank\_Train$CreditCard, Universal\_bank\_Train$Personal.Loan,  
 row.vars = c(2),dnn=c('CreditCard', 'Personal.loan'))  
print("Pivot Table for the given variables is")

## [1] "Pivot Table for the given variables is"

pivot3

## CreditCard 0 1  
## Personal.loan   
## 0 1918 794  
## 1 209 79

### D.Caculations Based on Pivot Tables

#### Individual Probabilities

##### Probabilities CC Given Loan

###### Probability of CC 1 given Loan 1

pivot3

## CreditCard 0 1  
## Personal.loan   
## 0 1918 794  
## 1 209 79

P2<- pivot3[2,2]/(pivot3[2,1]+pivot3[2,2])  
P2

## [1] 0.2743056

###### Probability of CC 1 given Loan 0

pivot3

## CreditCard 0 1  
## Personal.loan   
## 0 1918 794  
## 1 209 79

P3<- pivot3[1,2]/(pivot3[1,1]+pivot3[1,2])  
P3

## [1] 0.2927729

##### Below Are the results

$$ P(\frac{CC=1}{Loan=1})
=\frac{79}{209+79}= 0.2743056\\$$

$$ P(\frac{CC=1}{Loan=0})
=\frac{794}{1918+794}= 0.2927729\\$$

##### Probabilities Online Given Loan

###### Probability of Online 1 given Loan 1

pivot2

## Online 0 1  
## Personal.loan   
## 0 1128 1584  
## 1 106 182

P4<- pivot2[2,2]/(pivot2[2,1]+pivot2[2,2])  
P4

## [1] 0.6319444

###### Probability of Online 1 given Loan 0

pivot2

## Online 0 1  
## Personal.loan   
## 0 1128 1584  
## 1 106 182

P5<- pivot2[1,2]/(pivot2[1,1]+pivot2[1,2])  
P5

## [1] 0.5840708

##### Below Are the results

$$ P(\frac{Online=1}{Loan=1})
=\frac{182}{106+182}= 0.6319444\\$$

$$ P(\frac{Online=1}{Loan=0})
=\frac{1584}{1128+1584}= 0.5840708\\$$

##### Probability of Loan

P6<-(filter(Universal\_bank\_Train,Personal.Loan==1) %>%count())/nrow(Universal\_bank\_Train)  
P6<-P6[[1]]  
  
P7<-(filter(Universal\_bank\_Train,Personal.Loan==0) %>%count())/nrow(Universal\_bank\_Train)  
P7<-P7[[1]]  
P6

## [1] 0.096

P7

## [1] 0.904

$$ P({Loan=1})
=\frac{288}{3000}= 0.096\\$$

$$ P({Loan=0})
=\frac{2712}{3000}= 0.904\\$$

### E. Naive Bayes

P8 <- (P2\*P4\*P6)/((P2\*P4\*P6)+(P3\*P5\*P7))  
P8

## [1] 0.09718894

$$\text{Naive Bayes Probability is} \hspace{.3cm} 0.09718894$$

## F. Comparision

### Comparision b/w Naive bayes probability and Probability Using Pivot Table

P8-P1

## [1] 0.01065048

* The Probability obtained using pivot table is **0.086538461538461**.
* The Probability obtained using Naive bayes formula is **0.09718894**
* Since in Naive Bayes, We assume **conditional independence**.
* Hence, there is an increase of **0.01065048** in the value of probability

## G. Naive Bayes using R

# Creating Naive Bayes Classifier  
Loan.prob <- naiveBayes(Personal.Loan ~ ., data = Universal\_bank\_Train)   
  
c(Loan.prob$apriori[1]/(Loan.prob$apriori[1]+Loan.prob$apriori[2]),Loan.prob$apriori[2]/(Loan.prob$apriori[1]+Loan.prob$apriori[2]))

## 0 1   
## 0.904 0.096

Loan.prob$tables

## $Online  
## Online  
## Y 0 1  
## 0 0.4159292 0.5840708  
## 1 0.3680556 0.6319444  
##   
## $CreditCard  
## CreditCard  
## Y 0 1  
## 0 0.7072271 0.2927729  
## 1 0.7256944 0.2743056

Since the individual probabilities are matching to the above calculations in Question D .

Naive bayes probability=0.09718894

### Roc Calculation and plot

## predict probabilities  
pred.prob <- predict(Loan.prob, newdata = Universal\_bank\_Validation, type = "raw")  
  
###roc plot  
  
roc(Universal\_bank\_Validation$Personal.Loan,pred.prob[,2])

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

##   
## Call:  
## roc.default(response = Universal\_bank\_Validation$Personal.Loan, predictor = pred.prob[, 2])  
##   
## Data: pred.prob[, 2] in 1808 controls (Universal\_bank\_Validation$Personal.Loan 0) < 192 cases (Universal\_bank\_Validation$Personal.Loan 1).  
## Area under the curve: 0.4668

plot.roc(Universal\_bank\_Validation$Personal.Loan,pred.prob[,2])

## Setting levels: control = 0, case = 1  
## Setting direction: controls < cases

