Assignment 3 Naive Bayes FML

2025-06-15

#Assignment 3 Naive Bayes FML

#by David Wilkinson

#Load Libraries and Data

library(caret)  
library(dplyr)  
library(ISLR)  
library(e1071)

#load universal bank data  
library(readr)  
ub.df <- read.csv("C:\\Users\\david\\OneDrive\\Documents\\Kent State University MSBA\\Fundamentals of Machine Learning June 2025\\UniversalBank.csv")  
ub.df <- data.frame(ub.df)  
dim(ub.df)

## [1] 5000 14

ub.df\_select <- select(ub.df, CreditCard, Online, Personal.Loan)  
dim(ub.df\_select)

## [1] 5000 3

#Partition Data into Training and Test Sets

set.seed(123)  
#Divide data into test and train  
Index\_Train<-createDataPartition(ub.df\_select$Personal.Loan, p=0.6, list=FALSE)  
Train <-ub.df\_select[Index\_Train,]  
Test <-ub.df\_select[-Index\_Train,]

#Pivot Table

#make pivot table using melt and cast  
library(reshape)  
melted\_df <- melt(Train,  
 id = c("CreditCard", "Personal.Loan"),   
 measure = "Online")  
pivot\_cast <- cast(melted\_df, CreditCard + Personal.Loan ~ variable, sum)  
pivot\_cast

## CreditCard Personal.Loan Online  
## 1 0 0 1145  
## 2 0 1 122  
## 3 1 0 475  
## 4 1 1 57

#P(Loan Acceptance | Bank Credit Card = 1, Online Banking Services =1)

p\_loan\_given\_cc <- 57/sum(475+57)  
p\_loan\_given\_cc

## [1] 0.1071429

#Personal Loan Based on Online Banking Use Pivot Table

library(reshape)  
melted\_loan\_online <- melt(Train,  
 id = "Personal.Loan",   
 measure = "Online")  
loan\_online\_pivot <- cast(melted\_loan\_online, Personal.Loan ~ variable, sum)  
loan\_online\_pivot

## Personal.Loan Online  
## 1 0 1620  
## 2 1 179

#Personal Loan Based on Credit Card Pivot Table

library(reshape)  
melted\_loan\_cc <- melt(Train,  
 id = "Personal.Loan",   
 measure = "CreditCard")  
loan\_cc\_pivot <- cast(melted\_loan\_cc, Personal.Loan ~ variable, sum)  
loan\_cc\_pivot

## Personal.Loan CreditCard  
## 1 0 792  
## 2 1 91

#Manual computation of the following conditional probabilities

library(dplyr)  
#P(CC = 1 | Loan = 1) (the proportion of credit card holders among the loan acceptors)  
p\_cc1\_given\_loan1 <- nrow(filter(ub.df\_select, CreditCard == 1 & Personal.Loan == 1))/nrow(filter(ub.df\_select, Personal.Loan == 1))  
p\_cc1\_given\_loan1

## [1] 0.2979167

#P(Online = 1 | Loan = 1)  
p\_online1\_given\_loan1 <- nrow(filter(ub.df\_select, Online == 1 & Personal.Loan == 1))/nrow(filter(ub.df\_select, Personal.Loan == 1))  
p\_online1\_given\_loan1

## [1] 0.60625

#P(Loan = 1) (the proportion of loan acceptors)  
p\_loan1 <- nrow(filter(ub.df\_select,Personal.Loan == 1))/nrow(filter(ub.df\_select, Personal.Loan == 0|1))  
p\_loan1

## [1] 0.096

#P(CC = 1 | Loan = 0)  
p\_cc1\_given\_loan0 <- nrow(filter(ub.df\_select, CreditCard == 1 & Personal.Loan == 0))/nrow(filter(ub.df\_select, Personal.Loan == 0))  
p\_cc1\_given\_loan0

## [1] 0.2935841

#P(Online = 1 | Loan = 0)  
p\_online1\_given\_loan0 <- nrow(filter(ub.df\_select, Online == 1 & Personal.Loan == 0))/nrow(filter(ub.df\_select, Personal.Loan == 0))  
p\_online1\_given\_loan0

## [1] 0.5957965

#P(Loan = 0)  
p\_loan0 <- nrow(filter(ub.df\_select,Personal.Loan == 0))/nrow(filter(ub.df\_select, Personal.Loan == 0|1))  
p\_loan0

## [1] 0.904

#Naive Bayes Manual Calculation - P(Loan = 1 | CC= 1, Online = 1).

manual.nb <- (p\_cc1\_given\_loan1 \* p\_online1\_given\_loan1 \* p\_loan1)/((p\_cc1\_given\_loan1 \* p\_online1\_given\_loan1 \* p\_loan1) + (p\_cc1\_given\_loan0 \* p\_online1\_given\_loan0 \* p\_loan0))  
manual.nb

## [1] 0.09881706

This value from the manual computation is very close to the one calculated in Question B. However, the answer based on the first pivot table is more correct, since it is an exact Bayes’ classification, not naive.

#Perform Naive Bayes on Personal.Loan given CreditCard and Online Banking Use

nb\_model <-naiveBayes(Personal.Loan~ CreditCard + Online, data = Train)  
nb\_model

##   
## Naive Bayes Classifier for Discrete Predictors  
##   
## Call:  
## naiveBayes.default(x = X, y = Y, laplace = laplace)  
##   
## A-priori probabilities:  
## Y  
## 0 1   
## 0.90733333 0.09266667   
##   
## Conditional probabilities:  
## CreditCard  
## Y [,1] [,2]  
## 0 0.2909625 0.4542897  
## 1 0.3273381 0.4700881  
##   
## Online  
## Y [,1] [,2]  
## 0 0.5951506 0.4909531  
## 1 0.6438849 0.4797134

Entries 3 and 4 in the first pivot table are most necessary for calculating P(Loan 1| CreditCard = 1, Online)