Assignment_3

Tejasvini Mavuleti

3/6/2022

R Markdown

UniversalBank <- read.csv("C:/Users/mavul/OneDrive/Desktop/UniversalBank.csv") View(UniversalBank)

library(dplyr) library(reshape) library(reshape2) library(ggplot2) library(caret) library(ISLR) library(naivebayes) library(lattice)

UniversalBank$Personal. Loan < -as. factor(UniversalBank$Personal.Loan) UniversalBank$Online = as. factor(UniversalBank$Online) UniversalBank$CreditCard = as. factor(UniversalBank$CreditCard)

Parting the data as training 60% and testing 40%

set.seed(64060) Index <- createDataPartition(UniversalBank$Income, p=0.6, list = FALSE) Train_Data <- UniversalBank[Index,] Test_Data <- UniversalBank[-Index,]

A. Creating a pivot table

set.seed(64060) Melt_Train <- melt(Train_Data,id=c("CreditCard","Personal.Loan"),variable= "Online") cast_Train <- dcast(Melt_Train,CreditCard+Personal.Loan~Online) cast_Train <-cast_Train[c(1,2,14)] cast_Train

---

B. The probability that this customer will accept the loan offer

#P(Loan=1 | CC=1, Online=1) (85)/(811) 0.1048

C.Create two separate pivot tables for the training data

set.seed(64060) Melt_Train1 <- melt(Train_Data,id=c("Personal.Loan"),variable = "Online") cast_Train1 <- dcast(Melt_Train1,Personal.Loan~Online) cast_Train1 <- cast_Train1[c(1,13)] cast_Train1

set.seed(64060) Melt_Train2 <- melt(Train_Data,id=c("CreditCard"),variable = "Online") cast_Train2 <- dcast(Melt_Train2,CreditCard~Online) cast_Train2 <-cast_Train2[c(1,14)] cast_Train2

Train_Data1 <- Train_Data[c(13,10,14)] table(Train_Data1[,c(3,2)]) table(Train_Data1[,c(1,2)]) table(Train_Data1[,c(2)])

D.Compute the following quantities

i. #P(CC = 1 | Loan = 1)

(85)/(85+169) 0.334

ii. #P(Online = 1 | Loan = 1)

(152)/(152+102) 0.598

iii. #P(Loan = 1)

(254)/(2748+254) 0.084

iv. #P(CC = 1 | Loan = 0)

(811)/(811+1937) 0.291

v. #P(Online = 1 | Loan = 0)

(1659)/(1659+1089) 0.603

vi. #P(Loan = 0)

(2748)/(2748+254) 0.915

E.Use the quantities computed above to compute the naive Bayes probability

((0.334*0.598*0.084)/((0.334*0.598*0.084)+(0.291*0.603*0.915))) 0.09460

F.Compare this value with the one obtained from the pivot table in (B). Which is a more accurate estimate.

0.09460 are very similar to the 0.1048 the difference between the exact method and the naive-Bayes method is the exact method would need the exact same independent variable classifications to predict, where the naive bayes method does not.

G .Examine the model output on training data

library(e1071)

set.seed(64060) naivebayes <- naiveBayes(Personal.Loan~.,data=Train_Data1) naivebayes

(((0.334)*(0.598)*(0.084))/((0.334*0.598*0.084)+(0.295*0.603*0.915)))
(0.3159)*(0.5972)*(0.097)/((0.3159)*(0.5972)*(0.097) + (0.2971)*(0.6006)*(0.902))

– Values from the naive Bayes model probability 0.0934 is very similar to value of E that is 0.094.–