Machine Learning Assignment 3

AKANKSHA NADUKULA

2022-10-18

```
univbank <- read.csv("C:/Users/VIJAY KUMAR/Downloads/UniversalBank (1).csv")

#Converting the factor variables.
univbank$Personal.Loan <- as.factor(univbank$Personal.Loan)
univbank$Online <- as.factor(univbank$Online)
univbank$CreditCard <- as.factor(univbank$CreditCard)

#Checking the data to find any null variables
head(is.na(univbank))

#loading required libraries
library(ggplot2)
library(caret)
library(reshape2)
library(melt)
library(ISLR)
library(class)


#Partition of data to 60:40
set.seed(300)
univbank.part <-createDataPartition (univbank$Personal.Loan, p = 0.6, list =
F)
train <-univbank[univbank.part,]
validate <- univbank[-univbank.part,]


#Normalization of the dataset
norm<-preProcess(train[,-c(10,13:14)],method=c("center","scale"))
train_n <-predict(norm,train)
validate_n <-predict(norm,validate)

#A. Making a pivot table with Online as a column variable, CC as a row
variable, and Loan as a secondary row variable using the training data.
table <- ftable(train_n[,c(14,10,13)])
table

#B. The pivot table may be used to analyze the likelihood that this consumer
will accept the loan offer, which comes out to
50/(50+489)
#which is
```

```
#C. The training data is split into two different pivot tables. In one, the
personal loan (rows) are a function of the online (columns) while in the
other, the personal loan (rows) are a function of the credit card.

melt1 <- melt(train_n,id=c("Personal.Loan"),variable="Online")
melt2 <- melt(train_n,id=c("Personal.Loan"), variable="CreditCard")
cast1 <- dcast(melt1, Personal.Loan~ Online)
cast2 <- dcast(melt2, Personal.Loan~ CreditCard)
```

```
#D. Calculating the given quantities [P(A | B) means “the probability of A
given B

ftable(train_n[,c(10,13)])
ftable(train_n[,c(10,14)])
ftable(train_n[,10])
#1. P(CC = 1 | Loan = 1) =
(89/89+199)
#2. P(Online = 1 | Loan = 1)  =
(164/164+124)
#3. P(Loan = 1) =
(288/288+2712)
#4. P(CC = 1 | Loan = 0) =
(805/805+1907)
#5. P(Online = 1 | Loan = 0) =
(1616/ 1616+1096)
#6. P(Loan = 0) =
(2712/ 2712+288)


#E. Making use of the quantities computed above to compute the naive Bayes
probability P(Loan = 1 | CC = 1, Online = 1).
(0.309*0.569*0.096)/(0.309*0.569*0.096) + (0.296*0.595*0.904)
```

F. On comparing the naive bayes probability value i.e. 1.159212 with the one obtained from the pivot table in (B) i.e.0.092, it is evident that the naive bayes probability is higher than that of the pivot table.

```
#G. Running Naive Bayes on the data
library(naivebayes)

naive_b<- naive_bayes(Personal.Loan~Online+CreditCard,data=train_n)
naive_b

# For the consumer who is accepting the loan, using a credit card, and using
online banking, the value generated by running the Naive Bayes Model is
0.096, which is less than that of the value obtained in E.
```