````
---
title: "Assignment3_NaiveBayes"
author: "Ram"
date: "10/18/2021"
output:
  pdf_document: default
  word_document: default
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
loading all the required libraries

```{r}
library(reshape2)
library(gmodels)
library(caret)
library(ISLR)
library(e1071)
library(ggplot2)
library(lattice)
```

Reading the Universal bank CSV file to UB

```{r}
UB<-read.csv("C:/Users/ramne/Desktop/ML Assignment/Naive
Bayes/UniversalBank.csv")
```

changing the numerical variables to categorical variables

```{r}
UB$Personal.Loan<-factor(UB$Personal.Loan)
UB$Online<-factor(UB$Online)
UB$CreditCard<-factor(UB$CreditCard)
```


Split dataset into training (60%) and validation (40%)

```{r}
UBankNew <- sample(2,nrow(UB), replace=TRUE, prob=c(0.6,0.4))
tdata <- UB[UBankNew==1,]
vdata <- UB[UBankNew==2,]
```

A)Create a pivot table for the training data with Online as a column
variable, CC as a row variable
  and Loan as a secondary row variable such that the values inside the
table convey the count.

```{r}
melt_tbank<-
melt(tdata,id=c("CreditCard","Personal.Loan"),variable="Online")
cast_tbank<-dcast(melt_tbank,CreditCard+Personal.Loan~Online)
cast_tbank[,c(1:2,14)]
````
````

```

B)Consider the task of classifying a customer who owns a bank credit card and is actively using
  online banking services. Looking at the pivot table, what is the probability that this customer will
  accept the loan offer? [This is the probability of loan acceptance (Loan = 1) conditional on having
  a bank credit card (CC = 1) and being an active user of online banking services (Online = 1)].

  creating a 3 way cross table in R with the help of table function.

```{r}
a = table(tdata[,c(10,13,14)])
b<- as.data.frame(a)
b
```

It gives the cross tabulation of Online*CreditCard for Personal.Loan=0,Personal.Loan=1 separately as shown above.
There are 558 records where online = 1 and cc = 1 considering Loan=0(51) + Loan=1(507).
52 of them accept the loan, so the conditional probability is 52/546 = 0.09139.

C)Create two separate pivot tables for the training data. One will have Loan (rows) as a function of
  Online (columns) and the other will have Loan (rows) as a function of CC

 Pivot table for Loan (rows) as a function of Online (columns)

```{r}
table(tdata[,c(10,13)])
```

 Pivot table for Loan (rows) as a function of CC

```{r}
table(tdata[,c(10,14)])
```

 D)Compute the following quantities [P(A | B) means â€œthe probability of A given Bâ€]:

 i. P(CC = 1 | Loan = 1) (the proportion of credit card holders among the loan acceptors)

```{r}
P1<-table(tdata[,c(14,10)])
S1<-P1[2,2]/(P1[2,2]+P1[1,2])
S1
```

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

```{r}
P2<-table(tdata[,c(13,10)])
S2<-P2[2,2]/(P2[2,2]+P2[1,2])
S2
```

iii. P(Loan = 1) (the proportion of loan acceptors)

```{r}
P3<-table(tdata[,10])
S3<-P3[2]/(P3[2]+P3[1])
S3
```

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

```{r}
P4<-table(tdata[,c(14,10)])
S4<-P4[2,1]/(P4[2,1]+P4[1,1])
S4
```

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

```{r}
P5<-table(tdata[,c(13,10)])
S5<-P5[2,1]/(P5[2,1]+P5[1,1])
S5
```

vi. P(Loan = 0)

```{r}
P6<-table(tdata[,10])
S6<-P6[1]/(P6[1]+P6[2])
S6
```

E)Use the quantities computed above to compute the naive Bayes probability P(Loan = 1 | CC = 1,Online = 1).

Naive_Bayes_Probability= (S1 * S2 * S3)/ [(S1 * S2 * S3)) + (S4 * S5 * S6)]
= 0.0180/ (0.0180+0.1640) = 0.09890

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

The value obtained from the Pivot table is 0.09139 and the naive bayes probability is 0.09890 which are almost similar
and the value of the pivot table is more accurate because pivot table does not assume that probabilities are independent.

G)Which of the entries in this table are needed for computing P(Loan = 1 | CC = 1, Online = 1)? Run

naive Bayes on the data. Examine the model output on training data, and find the entry that
corresponds to P(Loan = 1 | CC = 1, Online = 1). Compare this to the number you obtained in (E).

Performing Naive Bayes on the training data

```{r}
table(tdata[,c(10,13:14)])
tdata_naive<-tdata[,c(10,13:14)]
UB_NB<-naiveBayes(Personal.Loan~.,data=tdata_naive)
UB_NB
```

Confusion Matrix

```{r}
c_prediction<-predict(UB_NB,newdata =tdata)
confusionMatrix(c_prediction,tdata$Personal.Loan)
CrossTable(x=tdata$Personal.Loan,y=c_prediction, prop.chisq = FALSE)
```

After running the naive Bayes on the data,obtained value is 0.0957
whereas the value obtained from E is 0.0989 which are almost similar