Assignment 3 - Fundamentals of Machine Learning

2022-10-14

library(ISLR)  
library(forecast)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

library(class)  
library(psych)   
library(caret)

## Loading required package: ggplot2

##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

## Loading required package: lattice

library(FNN)

##   
## Attaching package: 'FNN'

## The following objects are masked from 'package:class':  
##   
## knn, knn.cv

library(melt)  
library(MASS)   
library(reshape2)   
library(reshape)

##   
## Attaching package: 'reshape'

## The following objects are masked from 'package:reshape2':  
##   
## colsplit, melt, recast

## The following object is masked from 'package:class':  
##   
## condense

library(e1071)

bank = read.csv("~/Desktop/Fundamentals of Machine Learning/Assignment 2/UniversalBank.csv")  
bank$Online = as.factor(bank$Online)  
bank$Personal.Loan = as.factor(bank$Personal.Loan)  
bank$CreditCard = as.factor(bank$CreditCard)

train.index <- sample(row.names(bank), dim(bank)[1]\*0.6)   
test.index <- setdiff(row.names(bank), train.index)   
train.df <- bank[train.index, ]  
test.df <- bank[test.index, ]  
train <- bank[train.index, ]  
test <- bank[train.index,]

melted.bank = melt(train, id=c("CreditCard","Personal.Loan"), variable= "Online")  
recast.bank=dcast(melted.bank, CreditCard+Personal.Loan~Online)

## Aggregation function missing: defaulting to length

recast.bank[ ,c(1:2,14)]

## CreditCard Personal.Loan Online  
## 1 0 0 1897  
## 2 0 1 205  
## 3 1 0 809  
## 4 1 1 89

Note: These numbers were all reading different everytime I knit, even without making any additional changes so I went ahead and left it as it was reading the final time I ran the numbers.

78/(1871+216+835+78)= 2.8%

melted.bank1 = melt(train, id=c("Personal.Loan"), variable = "Online")  
melted.bank2 = melt(train, id=c("CreditCard"), variable = "Online")  
recast.bank1 = dcast(melted.bank1, Personal.Loan~Online)

## Aggregation function missing: defaulting to length

recast.bank2 = dcast(melted.bank2, CreditCard~Online)

## Aggregation function missing: defaulting to length

PersonalLoanPivot = recast.bank1[, c(1,13)]  
CreditCardPivot = recast.bank2[, c(1,14)]  
  
PersonalLoanPivot

## Personal.Loan Online  
## 1 0 2706  
## 2 1 294

CreditCardPivot

## CreditCard Online  
## 1 0 2102  
## 2 1 898

table(train[, c(14,10)])

## Personal.Loan  
## CreditCard 0 1  
## 0 1897 205  
## 1 809 89

table(train[, c(13,10)])

## Personal.Loan  
## Online 0 1  
## 0 1095 114  
## 1 1611 180

table(train[, c(10)])

##   
## 0 1   
## 2706 294

Note: These numbers were all reading different everytime I knit, even without making any additional changes so I went ahead and left it as it was reading the final time I ran the numbers.

1. 83/(83+199) = 29.4%
2. 167/(167+115) = 59.2%
3. 282/(282+2718) = 9.4%
4. 811/(811+1907) = 29.8%
5. 1599/(1599+1119) = 58.8%
6. 2718/(2718+282) = 90.6%

((83/(83+199))\*(167/(167+115))\*(282/(282+2718)))/(((83/(83+199))\*(167/(167+115))\*(282/(282+2718)))+((811/(811+1907))\*(1599/(1599+1119))\*2718/(2718+282)))

## [1] 0.09339891

naive.train = train.df[, c(10,13:14)]  
naive.test = test.df[, c(10,13:14)]  
naivebayes = naiveBayes(Personal.Loan~.,data=naive.train)  
naivebayes

##   
## Naive Bayes Classifier for Discrete Predictors  
##   
## Call:  
## naiveBayes.default(x = X, y = Y, laplace = laplace)  
##   
## A-priori probabilities:  
## Y  
## 0 1   
## 0.902 0.098   
##   
## Conditional probabilities:  
## Online  
## Y 0 1  
## 0 0.4046563 0.5953437  
## 1 0.3877551 0.6122449  
##   
## CreditCard  
## Y 0 1  
## 0 0.7010347 0.2989653  
## 1 0.6972789 0.3027211

The naive bayes output is the same as the previous methods. (.294)(.592)(.094)/((.294)(.592)(.094)+(.298)(.588)(.906)) = .09 In comparison to E, they are about the same.