Assignment 2

Randheer Gonuguntla

2023-10-01

```
library(class)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(caret)
```

```
## Loading required package: ggplot2

## Loading required package: lattice
```

Loading data set

```
dataset_UniversalBank<-read.csv("C:/Users/sidda/Downloads/UniversalBank.csv")
head(dataset_UniversalBank)
```

```
##   ID Age Experience Income ZIP.Code Family CCAvg Education Mortgage
## 1  1  25          1     49    91107      4   1.6         1        0
## 2  2  45         19     34    90089      3   1.5         1        0
## 3  3  39         15     11    94720      1   1.0         1        0
## 4  4  35          9    100    94112      1   2.7         2        0
## 5  5  35          8     45    91330      4   1.0         2        0
## 6  6  37         13     29    92121      4   0.4         2      155
##   Personal.Loan Securities.Account CD.Account Online CreditCard
## 1             0                  1          0      0          0
## 2             0                  1          0      0          0
## 3             0                  0          0      0          0
## 4             0                  0          0      0          0
## 5             0                  0          0      0          1
## 6             0                  0          0      1          0
```

Removing unwanted columns i.e ID and Zip code

```
dataset_UniversalBank1<-dataset_UniversalBank[,-1]
head(dataset_UniversalBank1)
```

```
##   Age Experience Income ZIP.Code Family CCAvg Education Mortgage Personal.Loan
## 1  25          1     49    91107      4   1.6         1        0             0
## 2  45         19     34    90089      3   1.5         1        0             0
## 3  39         15     11    94720      1   1.0         1        0             0
## 4  35          9    100    94112      1   2.7         2        0             0
## 5  35          8     45    91330      4   1.0         2        0             0
## 6  37         13     29    92121      4   0.4         2      155             0
##   Securities.Account CD.Account Online CreditCard
## 1                  1          0      0          0
## 2                  1          0      0          0
## 3                  0          0      0          0
## 4                  0          0      0          0
## 5                  0          0      0          1
## 6                  0          0      1          0
```

```
dataset_UniversalBank1<-dataset_UniversalBank1[,-4]
head(dataset_UniversalBank1)
```

```
##   Age Experience Income Family CCAvg Education Mortgage Personal.Loan
## 1  25          1     49      4   1.6         1        0             0
## 2  45         19     34      3   1.5         1        0             0
## 3  39         15     11      1   1.0         1        0             0
## 4  35          9    100      1   2.7         2        0             0
## 5  35          8     45      4   1.0         2        0             0
## 6  37         13     29      4   0.4         2      155             0
##   Securities.Account CD.Account Online CreditCard
## 1                  1          0      0          0
## 2                  1          0      0          0
## 3                  0          0      0          0
## 4                  0          0      0          0
## 5                  0          0      0          1
## 6                  0          0      1          0
```

converting personal loan as factor

```
dataset_UniversalBank1$Personal.Loan=as.factor(dataset_UniversalBank1$Personal.Loan)
```

running is.na to check if there are any NA values

```
head(is.na(dataset_UniversalBank1))
```

```
##        Age Experience Income Family CCAvg Education Mortgage Personal.Loan
## [1,] FALSE      FALSE  FALSE  FALSE FALSE     FALSE    FALSE         FALSE
## [2,] FALSE      FALSE  FALSE  FALSE FALSE     FALSE    FALSE         FALSE
## [3,] FALSE      FALSE  FALSE  FALSE FALSE     FALSE    FALSE         FALSE
## [4,] FALSE      FALSE  FALSE  FALSE FALSE     FALSE    FALSE         FALSE
```

```
## [5,] FALSE      FALSE  FALSE  FALSE FALSE     FALSE    FALSE         FALSE
## [6,] FALSE      FALSE  FALSE  FALSE FALSE     FALSE    FALSE         FALSE
##      Securities.Account CD.Account Online CreditCard
## [1,]              FALSE      FALSE  FALSE      FALSE
## [2,]              FALSE      FALSE  FALSE      FALSE
## [3,]              FALSE      FALSE  FALSE      FALSE
## [4,]              FALSE      FALSE  FALSE      FALSE
## [5,]              FALSE      FALSE  FALSE      FALSE
## [6,]              FALSE      FALSE  FALSE      FALSE
```

```
any(is.na(dataset_UniversalBank1))
```

```
## [1] FALSE
```

Converting categorical variable into i.e education into dummy variables
converting education into character

```
education<-as.character(dataset_UniversalBank1$Education)

dataset_UniversalBank2<-cbind(dataset_UniversalBank1[,-6],education)
head(dataset_UniversalBank2)
```

```
##   Age Experience Income Family CCAvg Mortgage Personal.Loan Securities.Account
## 1  25          1     49      4   1.6        0             0                  1
## 2  45         19     34      3   1.5        0             0                  1
## 3  39         15     11      1   1.0        0             0                  0
## 4  35          9    100      1   2.7        0             0                  0
## 5  35          8     45      4   1.0        0             0                  0
## 6  37         13     29      4   0.4      155             0                  0
##   CD.Account Online CreditCard education
## 1          0      0          0         1
## 2          0      0          0         1
## 3          0      0          0         1
## 4          0      0          0         2
## 5          0      0          1         2
## 6          0      1          0         2
```

```
dummy_model<-dummyVars("~education",data = dataset_UniversalBank2)
education_dummy<-data.frame(predict(dummy_model,dataset_UniversalBank2))
head(education_dummy)
```

```
##   education1 education2 education3
## 1          1          0          0
## 2          1          0          0
## 3          1          0          0
## 4          0          1          0
## 5          0          1          0
## 6          0          1          0
```

```
dataset_ub_dummy<-cbind(dataset_UniversalBank2[,-12],education_dummy)
head(dataset_ub_dummy)
```

```
##   Age Experience Income Family CCAvg Mortgage Personal.Loan Securities.Account
## 1  25          1     49      4   1.6        0             0                   1
## 2  45         19     34      3   1.5        0             0                   1
## 3  39         15     11      1   1.0        0             0                   0
## 4  35          9    100      1   2.7        0             0                   0
## 5  35          8     45      4   1.0        0             0                   0
## 6  37         13     29      4   0.4      155             0                   0
##   CD.Account Online CreditCard education1 education2 education3
## 1          0      0          0          1          0          0
## 2          0      0          0          1          0          0
## 3          0      0          0          1          0          0
## 4          0      0          0          0          1          0
## 5          0      0          1          0          1          0
## 6          0      1          0          0          1          0
```

dividing data into training and testing set

```
set.seed(3333)
train<-createDataPartition(dataset_ub_dummy$Personal.Loan,p=0.60,list = FALSE)
trainset<-dataset_ub_dummy[train,]
nrow(trainset)
```

```
## [1] 3000
```

```
validationset<-dataset_ub_dummy[-train,]
nrow(validationset)
```

```
## [1] 2000
```

```
testset<-data.frame(Age = 40, Experience = 10, Income = 84, Family = 2, CCAvg = 2,  Mortgage = 0, Securi
      CreditCard = 1,education1 = 0, education2 = 1, education3 = 0)


summary(trainset)
```

```
##       Age           Experience        Income           Family
##  Min.   :23.00   Min.   :-3.00   Min.   :  8.00   Min.   :1.000
##  1st Qu.:35.00   1st Qu.:10.00   1st Qu.: 39.00   1st Qu.:1.000
##  Median :46.00   Median :20.00   Median : 64.00   Median :2.000
##  Mean   :45.39   Mean   :20.14   Mean   : 73.75   Mean   :2.404
##  3rd Qu.:55.00   3rd Qu.:30.00   3rd Qu.: 98.00   3rd Qu.:3.000
##  Max.   :67.00   Max.   :43.00   Max.   :224.00   Max.   :4.000
##      CCAvg            Mortgage      Personal.Loan Securities.Account
##  Min.   : 0.000   Min.   :  0.00   0:2712        Min.   :0.0000
##  1st Qu.: 0.700   1st Qu.:  0.00   1: 288        1st Qu.:0.0000
##  Median : 1.600   Median :  0.00                 Median :0.0000
##  Mean   : 1.943   Mean   : 57.05                 Mean   :0.1063
##  3rd Qu.: 2.500   3rd Qu.:101.00                 3rd Qu.:0.0000
##  Max.   :10.000   Max.   :617.00                 Max.   :1.0000
##    CD.Account        Online         CreditCard       education1
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
```

```
##  Median :0.000   Median :1.0000   Median :0.0000   Median :0.0000
##  Mean   :0.061   Mean   :0.5997   Mean   :0.2913   Mean   :0.4243
##  3rd Qu.:0.000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##    education2     education3
##  Min.   :0.00   Min.   :0.0000
##  1st Qu.:0.00   1st Qu.:0.0000
##  Median :0.00   Median :0.0000
##  Mean   :0.28   Mean   :0.2957
##  3rd Qu.:1.00   3rd Qu.:1.0000
##  Max.   :1.00   Max.   :1.0000
```

```
summary(validationset)
```

```
##       Age           Experience        Income           Family
##  Min.   :23.00   Min.   :-3.00   Min.   :  8.00   Min.   :1.000
##  1st Qu.:35.00   1st Qu.:10.00   1st Qu.: 39.00   1st Qu.:1.000
##  Median :45.00   Median :20.00   Median : 63.00   Median :2.000
##  Mean   :45.26   Mean   :20.06   Mean   : 73.81   Mean   :2.385
##  3rd Qu.:55.00   3rd Qu.:30.00   3rd Qu.: 98.00   3rd Qu.:3.000
##  Max.   :67.00   Max.   :43.00   Max.   :204.00   Max.   :4.000
##      CCAvg           Mortgage      Personal.Loan Securities.Account
##  Min.   : 0.000   Min.   :  0.00   0:1808        Min.   :0.0000
##  1st Qu.: 0.700   1st Qu.:  0.00   1: 192        1st Qu.:0.0000
##  Median : 1.500   Median :  0.00                 Median :0.0000
##  Mean   : 1.931   Mean   : 55.67                 Mean   :0.1015
##  3rd Qu.: 2.600   3rd Qu.:101.00                 3rd Qu.:0.0000
##  Max.   :10.000   Max.   :635.00                 Max.   :1.0000
##    CD.Account         Online         CreditCard      education1
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.0000   Median :1.0000   Median :0.000   Median :0.0000
##  Mean   :0.0595   Mean   :0.5925   Mean   :0.298   Mean   :0.4115
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
##    education2       education3
##  Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000   Median :0.000
##  Mean   :0.2815   Mean   :0.307
##  3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :1.000
```

```
summary(testset)
```

```
##       Age       Experience     Income       Family      CCAvg      Mortgage
##  Min.   :40   Min.   :10   Min.   :84   Min.   :2   Min.   :2   Min.   :0
##  1st Qu.:40   1st Qu.:10   1st Qu.:84   1st Qu.:2   1st Qu.:2   1st Qu.:0
##  Median :40   Median :10   Median :84   Median :2   Median :2   Median :0
##  Mean   :40   Mean   :10   Mean   :84   Mean   :2   Mean   :2   Mean   :0
##  3rd Qu.:40   3rd Qu.:10   3rd Qu.:84   3rd Qu.:2   3rd Qu.:2   3rd Qu.:0
##  Max.   :40   Max.   :10   Max.   :84   Max.   :2   Max.   :2   Max.   :0
##  Securities.Account   CD.Account     Online    CreditCard   education1
```

```
##  Min.   :0          Min.   :0    Min.   :1   Min.   :1    Min.   :0
##  1st Qu.:0          1st Qu.:0    1st Qu.:1   1st Qu.:1    1st Qu.:0
##  Median :0          Median :0    Median :1   Median :1    Median :0
##  Mean   :0          Mean   :0    Mean   :1   Mean   :1    Mean   :0
##  3rd Qu.:0          3rd Qu.:0    3rd Qu.:1   3rd Qu.:1    3rd Qu.:0
##  Max.   :0          Max.   :0    Max.   :1   Max.   :1    Max.   :0
##    education2   education3
##  Min.   :1    Min.   :0
##  1st Qu.:1    1st Qu.:0
##  Median :1    Median :0
##  Mean   :1    Mean   :0
##  3rd Qu.:1    3rd Qu.:0
##  Max.   :1    Max.   :0
```

Normalizing

```
normvar<-c('Age',"Experience","Income","Family","CCAvg","Mortgage","Securities.Account","CD.Account","O
normalization_values<-preProcess(trainset[,normvar],method = c('center','scale'))

trainset.norm<-predict(normalization_values,trainset)
summary(trainset.norm)
```

```
##       Age              Experience           Income             Family
##  Min.   :-1.95962   Min.   :-2.02036   Min.   :-1.4387   Min.   :-1.2209
##  1st Qu.:-0.90931   1st Qu.:-0.88514   1st Qu.:-0.7604   1st Qu.:-1.2209
##  Median : 0.05348   Median :-0.01191   Median :-0.2134   Median :-0.3511
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.84121   3rd Qu.: 0.86133   3rd Qu.: 0.5305   3rd Qu.: 0.5187
##  Max.   : 1.89152   Max.   : 1.99655   Max.   : 3.2873   Max.   : 1.3885
##      CCAvg            Mortgage       Personal.Loan Securities.Account
##  Min.   :-1.1183   Min.   :-0.5597   0:2712        Min.   :-0.3449
##  1st Qu.:-0.7153   1st Qu.:-0.5597   1: 288        1st Qu.:-0.3449
##  Median :-0.1972   Median :-0.5597                 Median :-0.3449
##  Mean   : 0.0000   Mean   : 0.0000                 Mean   : 0.0000
##  3rd Qu.: 0.3210   3rd Qu.: 0.4311                 3rd Qu.:-0.3449
##  Max.   : 4.6389   Max.   : 5.4933                 Max.   : 2.8985
##    CD.Account          Online          CreditCard        education1
##  Min.   :-0.2548   Min.   :-1.2237   Min.   :-0.6411   Min.   :-0.8584
##  1st Qu.:-0.2548   1st Qu.:-1.2237   1st Qu.:-0.6411   1st Qu.:-0.8584
##  Median :-0.2548   Median : 0.8169   Median :-0.6411   Median :-0.8584
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.:-0.2548   3rd Qu.: 0.8169   3rd Qu.: 1.5594   3rd Qu.: 1.1646
##  Max.   : 3.9228   Max.   : 0.8169   Max.   : 1.5594   Max.   : 1.1646
##    education2        education3
##  Min.   :-0.6235   Min.   :-0.6478
##  1st Qu.:-0.6235   1st Qu.:-0.6478
##  Median :-0.6235   Median :-0.6478
##  Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 1.6033   3rd Qu.: 1.5432
##  Max.   : 1.6033   Max.   : 1.5432
```

```
validationset.norm<-predict(normalization_values,validationset)
summary(validationset.norm)
```

```
##       Age              Experience           Income              Family
##  Min.   :-1.95962   Min.   :-2.020356   Min.   :-1.438657   Min.   :-1.2209
##  1st Qu.:-0.90931   1st Qu.:-0.885145   1st Qu.:-0.760390   1st Qu.:-1.2209
##  Median :-0.03405   Median :-0.011905   Median :-0.235279   Median :-0.3511
##  Mean   :-0.01107   Mean   :-0.006928   Mean   : 0.001141   Mean   :-0.0158
##  3rd Qu.: 0.84121   3rd Qu.: 0.861334   3rd Qu.: 0.530508   3rd Qu.: 0.5187
##  Max.   : 1.89152   Max.   : 1.996545   Max.   : 2.849747   Max.   : 1.3885
##      CCAvg              Mortgage         Personal.Loan Securities.Account
##  Min.   :-1.118341   Min.   :-0.55972   0:1808        Min.   :-0.34489
##  1st Qu.:-0.715336   1st Qu.:-0.55972   1: 192        1st Qu.:-0.34489
##  Median :-0.254759   Median :-0.55972                 Median :-0.34489
##  Mean   :-0.006571   Mean   :-0.01362                 Mean   :-0.01568
##  3rd Qu.: 0.378535   3rd Qu.: 0.43113                 3rd Qu.:-0.34489
##  Max.   : 4.638873   Max.   : 5.66987                 Max.   : 2.89855
##    CD.Account            Online          CreditCard         education1
##  Min.   :-0.254835   Min.   :-1.22369   Min.   :-0.64106   Min.   :-0.85841
##  1st Qu.:-0.254835   1st Qu.:-1.22369   1st Qu.:-0.64106   1st Qu.:-0.85841
##  Median :-0.254835   Median : 0.81693   Median :-0.64106   Median :-0.85841
##  Mean   :-0.006266   Mean   :-0.01462   Mean   : 0.01467   Mean   :-0.02596
##  3rd Qu.:-0.254835   3rd Qu.: 0.81693   3rd Qu.: 1.55939   3rd Qu.: 1.16455
##  Max.   : 3.922794   Max.   : 0.81693   Max.   : 1.55939   Max.   : 1.16455
##    education2         education3
##  Min.   :-0.62351   Min.   :-0.64780
##  1st Qu.:-0.62351   1st Qu.:-0.64780
##  Median :-0.62351   Median :-0.64780
##  Mean   : 0.00334   Mean   : 0.02483
##  3rd Qu.: 1.60330   3rd Qu.: 1.54318
##  Max.   : 1.60330   Max.   : 1.54318
```

```
testset.norm<-predict(normalization_values,testset)
summary(testset.norm)
```

```
##       Age              Experience           Income              Family
##  Min.   :-0.4717   Min.   :-0.8851   Min.   :0.2242   Min.   :-0.3511
##  1st Qu.:-0.4717   1st Qu.:-0.8851   1st Qu.:0.2242   1st Qu.:-0.3511
##  Median :-0.4717   Median :-0.8851   Median :0.2242   Median :-0.3511
##  Mean   :-0.4717   Mean   :-0.8851   Mean   :0.2242   Mean   :-0.3511
##  3rd Qu.:-0.4717   3rd Qu.:-0.8851   3rd Qu.:0.2242   3rd Qu.:-0.3511
##  Max.   :-0.4717   Max.   :-0.8851   Max.   :0.2242   Max.   :-0.3511
##      CCAvg              Mortgage       Securities.Account   CD.Account
##  Min.   :0.0331   Min.   :-0.5597   Min.   :-0.3449    Min.   :-0.2548
##  1st Qu.:0.0331   1st Qu.:-0.5597   1st Qu.:-0.3449    1st Qu.:-0.2548
##  Median :0.0331   Median :-0.5597   Median :-0.3449    Median :-0.2548
##  Mean   :0.0331   Mean   :-0.5597   Mean   :-0.3449    Mean   :-0.2548
##  3rd Qu.:0.0331   3rd Qu.:-0.5597   3rd Qu.:-0.3449    3rd Qu.:-0.2548
##  Max.   :0.0331   Max.   :-0.5597   Max.   :-0.3449    Max.   :-0.2548
##      Online         CreditCard       education1         education2
##  Min.   :0.8169   Min.   :1.559   Min.   :-0.8584   Min.   :1.603
##  1st Qu.:0.8169   1st Qu.:1.559   1st Qu.:-0.8584   1st Qu.:1.603
```

```
##  Median :0.8169   Median :1.559   Median :-0.8584   Median :1.603
##  Mean   :0.8169   Mean   :1.559   Mean   :-0.8584   Mean   :1.603
##  3rd Qu.:0.8169   3rd Qu.:1.559   3rd Qu.:-0.8584   3rd Qu.:1.603
##  Max.   :0.8169   Max.   :1.559   Max.   :-0.8584   Max.   :1.603
##    education3
##  Min.   :-0.6478
##  1st Qu.:-0.6478
##  Median :-0.6478
##  Mean   :-0.6478
##  3rd Qu.:-0.6478
##  Max.   :-0.6478
```

question 1: Classifying the customer

```
set.seed(3333)
new_grid<-expand.grid(k=c(1))
new_model<-train(Personal.Loan~.,data=trainset.norm,method="knn",tuneGrid=new_grid)

new_model
```

```
## k-Nearest Neighbors
##
## 3000 samples
##   13 predictor
##    2 classes: '0', '1'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 3000, 3000, 3000, 3000, 3000, 3000, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.9526142  0.696821
##
## Tuning parameter 'k' was held constant at a value of 1
```

```
predict_test<-predict(new_model,testset.norm)
predict_test
```

```
## [1] 0
## Levels: 0 1
```

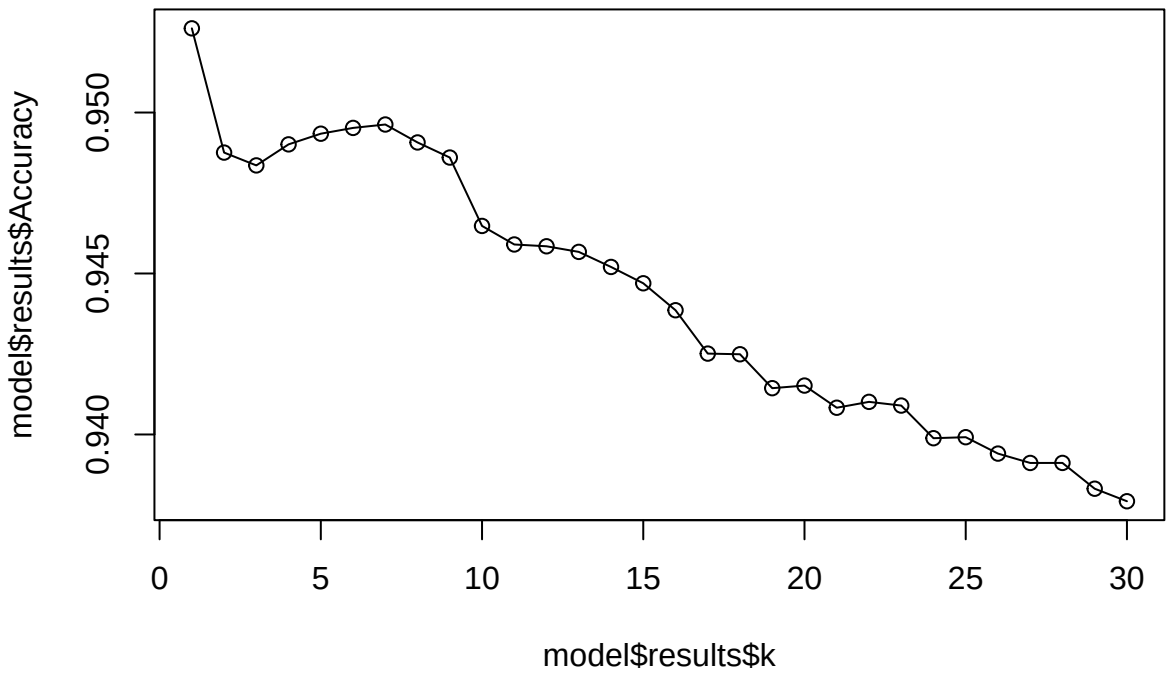
From the above output it can be seen that the customer can be classified into two levels 0 and 1.

question 2: identifying the best k

```
set.seed(3333)
searchGrid <- expand.grid(k=seq(1:30))
model<-train(Personal.Loan~.,data=trainset.norm,method="knn",tuneGrid=searchGrid)
model
```

```
## k-Nearest Neighbors
##
## 3000 samples
##   13 predictor
##    2 classes: '0', '1'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 3000, 3000, 3000, 3000, 3000, 3000, ...
## Resampling results across tuning parameters:
##
##   k   Accuracy   Kappa
##    1  0.9526142  0.6968210
##    2  0.9487558  0.6701117
##    3  0.9483571  0.6609592
##    4  0.9490110  0.6594947
##    5  0.9493416  0.6554003
##    6  0.9495220  0.6495090
##    7  0.9496276  0.6457508
##    8  0.9490702  0.6397724
##    9  0.9486012  0.6322860
##   10  0.9464770  0.6139487
##   11  0.9459014  0.6070735
##   12  0.9458449  0.6042007
##   13  0.9456713  0.6025038
##   14  0.9452035  0.5964677
##   15  0.9446964  0.5909303
##   16  0.9438587  0.5806804
##   17  0.9425118  0.5679100
##   18  0.9424910  0.5677760
##   19  0.9414378  0.5562707
##   20  0.9415189  0.5544854
##   21  0.9408304  0.5479054
##   22  0.9410128  0.5486401
##   23  0.9408976  0.5472808
##   24  0.9398842  0.5360221
##   25  0.9399156  0.5355442
##   26  0.9394064  0.5307370
##   27  0.9391155  0.5284241
##   28  0.9391164  0.5282213
##   29  0.9383155  0.5195490
##   30  0.9379262  0.5147746
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 1.
```

```
plot(model$results$k,model$results$Accuracy, type = 'o')
```



finding the best k

```
best_k <- model$bestTune[[1]]
best_k
```

```
## [1] 1
```

So the best K for the data is 1

question3:confusion matrix

```
library(gmodels)

train_label<-trainset.norm[,7]
validation_label<-validationset.norm[,7]
test_label<-testset.norm[,7]

predicted_validationlabel<-knn(trainset.norm,validationset.norm,cl=train_label,k=5)

CrossTable(x=validation_label,y=predicted_validationlabel,prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
```

```
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  2000
##
##
##                  | predicted_validationlabel
## validation_label |         0 |         1 | Row Total |
## -----------------|-----------|-----------|-----------|
##                0 |      1805 |         3 |      1808 |
##                  |     0.998 |     0.002 |     0.904 |
##                  |     0.971 |     0.021 |           |
##                  |     0.902 |     0.002 |           |
## -----------------|-----------|-----------|-----------|
##                1 |        53 |       139 |       192 |
##                  |     0.276 |     0.724 |     0.096 |
##                  |     0.029 |     0.979 |           |
##                  |     0.026 |     0.070 |           |
## -----------------|-----------|-----------|-----------|
##     Column Total |      1858 |       142 |      2000 |
##                  |     0.929 |     0.071 |           |
## -----------------|-----------|-----------|-----------|
##
##
```

question4:Classifying the given customer with best k

```
set.seed(3333)
bestk_grid<-expand.grid(k=c(best_k))
bestk_model<-train(Personal.Loan~.,data=trainset.norm,method="knn",tuneGrid=bestk_grid)
bestk_model
```

```
## k-Nearest Neighbors
##
## 3000 samples
##   13 predictor
##    2 classes: '0', '1'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 3000, 3000, 3000, 3000, 3000, 3000, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.9526142  0.696821
##
## Tuning parameter 'k' was held constant at a value of 1
```

```
bestk_test<-predict(bestk_model,testset.norm)
bestk_test
```

```
## [1] 0
## Levels: 0 1
```

question5:confusion matrix for validation and training sets
dividing dataset into traning, validation and testing set

```
set.seed(3333)
train1<-createDataPartition(dataset_ub_dummy$Personal.Loan,p=0.50,list = FALSE)
trainset_2<-dataset_ub_dummy[train1,]
middleset<-dataset_ub_dummy[-train1,]
nrow(middleset)
```

```
## [1] 2500
```

```
train2<-createDataPartition(middleset$Personal.Loan,p=0.6,list = FALSE)
validationset_2<-middleset[train2,]
testset_2<-middleset[-train2,]

nrow(trainset_2)
```

```
## [1] 2500
```

```
nrow(validationset_2)
```

```
## [1] 1500
```

```
nrow(testset_2)
```

```
## [1] 1000
```

normalizing trainset_2,validationset_2,testset_2

```
normvar<-c('Age',"Experience","Income","Family","CCAvg","Mortgage","Securities.Account","CD.Account","O
normalization_values_2<-preProcess(trainset_2[,normvar],method = c('center','scale'))

trainset.norm_2<-predict(normalization_values_2,trainset_2)
summary(trainset.norm_2)
```

```
##       Age              Experience          Income            Family
##  Min.   :-1.94902   Min.   :-2.00817   Min.   :-1.4405   Min.   :-1.2228
##  1st Qu.:-0.90505   1st Qu.:-0.88074   1st Qu.:-0.7582   1st Qu.:-1.2228
##  Median : 0.05192   Median :-0.01349   Median :-0.2079   Median :-0.3541
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.83490   3rd Qu.: 0.85375   3rd Qu.: 0.5404   3rd Qu.: 0.5146
##  Max.   : 1.87887   Max.   : 1.98118   Max.   : 3.1817   Max.   : 1.3833
##      CCAvg            Mortgage       Personal.Loan Securities.Account
```

```
##  Min.   :-1.1048   Min.   :-0.5621   0:2260        Min.   :-0.3529
##  1st Qu.:-0.7090   1st Qu.:-0.5621   1: 240        1st Qu.:-0.3529
##  Median :-0.2566   Median :-0.5621                 Median :-0.3529
##  Mean   : 0.0000   Mean   : 0.0000                 Mean   : 0.0000
##  3rd Qu.: 0.3653   3rd Qu.: 0.4352                 3rd Qu.:-0.3529
##  Max.   : 4.5493   Max.   : 5.2566                 Max.   : 2.8323
##    CD.Account          Online          CreditCard        education1
##  Min.   :-0.2588   Min.   :-1.2194   Min.   :-0.6384   Min.   :-0.8557
##  1st Qu.:-0.2588   1st Qu.:-1.2194   1st Qu.:-0.6384   1st Qu.:-0.8557
##  Median :-0.2588   Median : 0.8197   Median :-0.6384   Median :-0.8557
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.:-0.2588   3rd Qu.: 0.8197   3rd Qu.: 1.5659   3rd Qu.: 1.1682
##  Max.   : 3.8623   Max.   : 0.8197   Max.   : 1.5659   Max.   : 1.1682
##    education2        education3
##  Min.   :-0.6235   Min.   :-0.6502
##  1st Qu.:-0.6235   1st Qu.:-0.6502
##  Median :-0.6235   Median :-0.6502
##  Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 1.6032   3rd Qu.: 1.5375
##  Max.   : 1.6032   Max.   : 1.5375
```

```
validationset.norm_2<-predict(normalization_values_2,validationset_2)
summary(validationset.norm_2)
```

```
##       Age              Experience          Income              Family
##  Min.   :-1.94902   Min.   :-2.00817   Min.   :-1.440548   Min.   :-1.22276
##  1st Qu.:-0.90505   1st Qu.:-0.88074   1st Qu.:-0.758216   1st Qu.:-1.22276
##  Median :-0.03508   Median :-0.01349   Median :-0.251970   Median :-0.35407
##  Mean   :-0.01188   Mean   :-0.01055   Mean   : 0.006641   Mean   :-0.02224
##  3rd Qu.: 0.92190   3rd Qu.: 0.85375   3rd Qu.: 0.474384   3rd Qu.: 0.51461
##  Max.   : 1.87887   Max.   : 1.98118   Max.   : 2.873551   Max.   : 1.38329
##      CCAvg             Mortgage        Personal.Loan Securities.Account
##  Min.   :-1.10475   Min.   :-0.56205   0:1356        Min.   :-0.3529
##  1st Qu.:-0.70897   1st Qu.:-0.56205   1: 144        1st Qu.:-0.3529
##  Median :-0.25665   Median :-0.56205                 Median :-0.3529
##  Mean   :-0.02199   Mean   :-0.03423                 Mean   :-0.0705
##  3rd Qu.: 0.30876   3rd Qu.: 0.37707                 3rd Qu.:-0.3529
##  Max.   : 4.54931   Max.   : 5.41155                 Max.   : 2.8323
##    CD.Account          Online          CreditCard         education1
##  Min.   :-0.25881   Min.   :-1.21941   Min.   :-0.63835   Min.   :-0.85569
##  1st Qu.:-0.25881   1st Qu.:-1.21941   1st Qu.:-0.63835   1st Qu.:-0.85569
##  Median :-0.25881   Median : 0.81974   Median :-0.63835   Median :-0.85569
##  Mean   :-0.03077   Mean   : 0.03535   Mean   : 0.01999   Mean   :-0.01241
##  3rd Qu.:-0.25881   3rd Qu.: 0.81974   3rd Qu.: 1.56590   3rd Qu.: 1.16818
##  Max.   : 3.86233   Max.   : 0.81974   Max.   : 1.56590   Max.   : 1.16818
##    education2        education3
##  Min.   :-0.62348   Min.   :-0.650162
##  1st Qu.:-0.62348   1st Qu.:-0.650162
##  Median :-0.62348   Median :-0.650162
##  Mean   : 0.01039   Mean   : 0.003208
##  3rd Qu.: 1.60325   3rd Qu.: 1.537463
##  Max.   : 1.60325   Max.   : 1.537463
```

```
testset.norm_2<-predict(normalization_values_2,testset_2)
summary(testset.norm_2)
```

```
##       Age              Experience            Income             Family
##  Min.   :-1.94902   Min.   :-2.008167   Min.   :-1.44055   Min.   :-1.22276
##  1st Qu.:-0.81805   1st Qu.:-0.880743   1st Qu.:-0.76372   1st Qu.:-1.22276
##  Median :-0.03508   Median :-0.013494   Median :-0.20795   Median :-0.35407
##  Mean   :-0.01037   Mean   :-0.006296   Mean   : 0.02598   Mean   :-0.01529
##  3rd Qu.: 0.83490   3rd Qu.: 0.853755   3rd Qu.: 0.62846   3rd Qu.: 0.51461
##  Max.   : 1.87887   Max.   : 1.894453   Max.   : 3.31377   Max.   : 1.38329
##      CCAvg              Mortgage        Personal.Loan Securities.Account
##  Min.   :-1.10475   Min.   :-0.56205   0:904         Min.   :-0.352926
##  1st Qu.:-0.70897   1st Qu.:-0.56205   1: 96         1st Qu.:-0.352926
##  Median :-0.20010   Median :-0.56205                 Median :-0.352926
##  Mean   :-0.01217   Mean   :-0.02391                 Mean   : 0.003822
##  3rd Qu.: 0.30876   3rd Qu.: 0.41580                 3rd Qu.:-0.352926
##  Max.   : 3.75774   Max.   : 5.58582                 Max.   : 2.832324
##    CD.Account            Online           CreditCard         education1
##  Min.   :-0.258807   Min.   :-1.21941   Min.   :-0.63835   Min.   :-0.85569
##  1st Qu.:-0.258807   1st Qu.:-1.21941   1st Qu.:-0.63835   1st Qu.:-0.85569
##  Median :-0.258807   Median : 0.81974   Median :-0.63835   Median :-0.85569
##  Mean   :-0.003297   Mean   :-0.06525   Mean   : 0.01852   Mean   :-0.01781
##  3rd Qu.:-0.258807   3rd Qu.: 0.81974   3rd Qu.: 1.56590   3rd Qu.: 1.16818
##  Max.   : 3.862331   Max.   : 0.81974   Max.   : 1.56590   Max.   : 1.16818
##    education2          education3
##  Min.   :-0.623485   Min.   :-0.6502
##  1st Qu.:-0.623485   1st Qu.:-0.6502
##  Median :-0.623485   Median :-0.6502
##  Mean   :-0.008907   Mean   : 0.0280
##  3rd Qu.: 1.603247   3rd Qu.: 1.5375
##  Max.   : 1.603247   Max.   : 1.5375
```

confusion matrix

```
library(gmodels)

train_label_2<-trainset.norm_2[,7]
validation_label_2<-validationset.norm_2[,7]
test_label_2<-testset.norm_2[,7]

predicted_validationlabel_2<-knn(trainset.norm_2,validationset.norm_2,cl=train_label_2,k=best_k)

predicted_testlabel_2<-knn(trainset.norm_2,testset.norm_2,cl=train_label_2,k=best_k)

confusionmatrix_1<-CrossTable(x=validation_label_2,y=predicted_validationlabel_2,prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
```

```
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1500
##
##
##                    | predicted_validationlabel_2
## validation_label_2 |         0 |         1 | Row Total |
## -------------------|-----------|-----------|-----------|
##                  0 |      1351 |         5 |      1356 |
##                    |     0.996 |     0.004 |     0.904 |
##                    |     0.983 |     0.040 |           |
##                    |     0.901 |     0.003 |           |
## -------------------|-----------|-----------|-----------|
##                  1 |        23 |       121 |       144 |
##                    |     0.160 |     0.840 |     0.096 |
##                    |     0.017 |     0.960 |           |
##                    |     0.015 |     0.081 |           |
## -------------------|-----------|-----------|-----------|
##       Column Total |      1374 |       126 |      1500 |
##                    |     0.916 |     0.084 |           |
## -------------------|-----------|-----------|-----------|
##
##
```

```
confusionmatrix_2<-CrossTable(x=test_label_2,y=predicted_testlabel_2,prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1000
##
##
##              | predicted_testlabel_2
## test_label_2 |         0 |         1 | Row Total |
## -------------|-----------|-----------|-----------|
##            0 |       899 |         5 |       904 |
##              |     0.994 |     0.006 |     0.904 |
##              |     0.986 |     0.057 |           |
##              |     0.899 |     0.005 |           |
## -------------|-----------|-----------|-----------|
##            1 |        13 |        83 |        96 |
##              |     0.135 |     0.865 |     0.096 |
##              |     0.014 |     0.943 |           |
```

```
##              |     0.013 |     0.083 |           |
## -------------|-----------|-----------|-----------|
## Column Total |       912 |        88 |      1000 |
##              |     0.912 |     0.088 |           |
## -------------|-----------|-----------|-----------|
##
##
```

```
validation_table<-table(validation_label_2,predicted_validationlabel_2)
confusionMatrix(validation_table)
```

```
## Confusion Matrix and Statistics
##
##                   predicted_validationlabel_2
## validation_label_2    0    1
##                  0 1351    5
##                  1   23  121
##
##                Accuracy : 0.9813
##                  95% CI : (0.9731, 0.9876)
##     No Information Rate : 0.916
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8861
##
##  Mcnemar's Test P-Value : 0.001315
##
##             Sensitivity : 0.9833
##             Specificity : 0.9603
##          Pos Pred Value : 0.9963
##          Neg Pred Value : 0.8403
##              Prevalence : 0.9160
##          Detection Rate : 0.9007
##    Detection Prevalence : 0.9040
##       Balanced Accuracy : 0.9718
##
##        'Positive' Class : 0
##
```

```
test_table<-table(test_label_2,predicted_testlabel_2)
confusionMatrix(test_table)
```

```
## Confusion Matrix and Statistics
##
##             predicted_testlabel_2
## test_label_2   0   1
##            0 899   5
##            1  13  83
##
##                Accuracy : 0.982
##                  95% CI : (0.9717, 0.9893)
##     No Information Rate : 0.912
##     P-Value [Acc > NIR] : < 2e-16
```

```
##
##                   Kappa : 0.8923
##
##  Mcnemar's Test P-Value : 0.09896
##
##             Sensitivity : 0.9857
##             Specificity : 0.9432
##          Pos Pred Value : 0.9945
##          Neg Pred Value : 0.8646
##              Prevalence : 0.9120
##          Detection Rate : 0.8990
##    Detection Prevalence : 0.9040
##       Balanced Accuracy : 0.9645
##
##        'Positive' Class : 0
##
```

When comparing the confusion matrices of the validation and testing sets, it can be seen that the validation set's accuracy and sensitivity are marginally higher than the test set's.