Assignment 2

Franklin Ngochi

2023-02-17

```
library("ISLR")
library("caret")
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
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("FNN")
```

#Getting the data from CSV file

```
UB<-read.csv("C:\\Users\\ngoch\\OneDrive\\Documents\\KSU\\Fundamentals of Machine Learning\\UniversalBa
```

#Dummify the "Education" variable

```
UB$Education_1<-ifelse (UB$Education==1, 1, 0)
UB$Education_2<-ifelse (UB$Education==2, 1, 0)
UB$Education_3<-ifelse (UB$Education==3, 1, 0)
```

#Subsetting to only include needed columns

```
Dataset<-select(UB,Age, Experience, Income, Family, CCAvg, Mortgage, Education_1, Education_2, Educatio
head(Dataset)
```

```
##   Age Experience Income Family CCAvg Mortgage Education_1 Education_2
## 1  25          1     49      4   1.6        0           1           0
## 2  45         19     34      3   1.5        0           1           0
## 3  39         15     11      1   1.0        0           1           0
## 4  35          9    100      1   2.7        0           0           1
## 5  35          8     45      4   1.0        0           0           1
## 6  37         13     29      4   0.4      155           0           1
##   Education_3 Personal.Loan Securities.Account CD.Account Online CreditCard
## 1           0             0                  1          0      0          0
## 2           0             0                  1          0      0          0
## 3           0             0                  0          0      0          0
## 4           0             0                  0          0      0          0
## 5           0             0                  0          0      0          1
## 6           0             0                  0          0      1          0
```

#Partitioning the data into training(60%) and validation set (40%)

```
set.seed(321)
Index<-createDataPartition(Dataset$Personal.Loan,p=0.6, list=FALSE)
TrainSet<-Dataset[Index,]
ValidSet<-Dataset[-Index,]
head(TrainSet)
```

```
##    Age Experience Income Family CCAvg Mortgage Education_1 Education_2
## 1   25          1     49      4   1.6        0           1           0
## 2   45         19     34      3   1.5        0           1           0
## 4   35          9    100      1   2.7        0           0           1
## 6   37         13     29      4   0.4      155           0           1
## 9   35         10     81      3   0.6      104           0           1
## 11  65         39    105      4   2.4        0           0           0
##    Education_3 Personal.Loan Securities.Account CD.Account Online CreditCard
## 1            0             0                  1          0      0          0
## 2            0             0                  1          0      0          0
## 4            0             0                  0          0      0          0
## 6            0             0                  0          0      1          0
## 9            0             0                  0          0      1          0
## 11           1             0                  0          0      0          0
```

```
head(ValidSet)
```

```
##    Age Experience Income Family CCAvg Mortgage Education_1 Education_2
## 3   39         15     11      1   1.0        0           1           0
## 5   35          8     45      4   1.0        0           0           1
## 7   53         27     72      2   1.5        0           0           1
## 8   50         24     22      1   0.3        0           0           0
## 10  34          9    180      1   8.9        0           0           0
## 13  48         23    114      2   3.8        0           0           0
##    Education_3 Personal.Loan Securities.Account CD.Account Online CreditCard
## 3            0             0                  0          0      0          0
## 5            0             0                  0          0      0          1
## 7            0             0                  0          0      1          0
## 8            1             0                  0          0      0          1
## 10           1             1                  0          0      0          0
## 13           1             0                  1          0      0          0
```

#Create dataframe for test case

```
TestSet <- data.frame(Age = as.integer(40), Experience = as.integer(10), Income = as.integer(84), Family
```

#Normalizing the data

```
Train.Norm<-TrainSet
Valid.Norm<-ValidSet
Test.Norm<-TestSet
Norm.Data<-preProcess(TrainSet [,1:6], method=c("center", "scale"))
Train.Norm[, 1:6]<-predict(Norm.Data, TrainSet[,1:6])#Training Data
Valid.Norm[,1:6]<-predict(Norm.Data, ValidSet[,1:6])#Validation Data
Test.Norm <-predict(Norm.Data, TestSet)#Validation Data
summary(Train.Norm)
```

```
##       Age               Experience           Income             Family
##  Min.   :-1.95926   Min.   :-2.02646   Min.   :-1.4420   Min.   :-1.2428
##  1st Qu.:-0.82870   1st Qu.:-0.89447   1st Qu.:-0.7689   1st Qu.:-1.2428
##  Median : 0.04096   Median :-0.02371   Median :-0.2043   Median :-0.3690
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.91062   3rd Qu.: 0.84705   3rd Qu.: 0.5123   3rd Qu.: 0.5048
##  Max.   : 1.86725   Max.   : 1.97903   Max.   : 3.2483   Max.   : 1.3786
##      CCAvg              Mortgage         Education_1     Education_2
##  Min.   :-1.1114   Min.   :-0.5604   Min.   :0.000   Min.   :0.0000
##  1st Qu.:-0.7116   1st Qu.:-0.5604   1st Qu.:0.000   1st Qu.:0.0000
##  Median :-0.1976   Median :-0.5604   Median :0.000   Median :0.0000
##  Mean   : 0.0000   Mean   : 0.0000   Mean   :0.408   Mean   :0.2833
##  3rd Qu.: 0.3735   3rd Qu.: 0.4261   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   : 4.5998   Max.   : 5.3488   Max.   :1.000   Max.   :1.0000
##   Education_3     Personal.Loan   Securities.Account   CD.Account
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000     Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000     1st Qu.:0.00000
##  Median :0.0000   Median :0.000   Median :0.0000     Median :0.00000
##  Mean   :0.3087   Mean   :0.098   Mean   :0.1097     Mean   :0.06033
##  3rd Qu.:1.0000   3rd Qu.:0.000   3rd Qu.:0.0000     3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000     Max.   :1.00000
##      Online         CreditCard
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.0000
##  Mean   :0.6053   Mean   :0.2893
##  3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000
```

```
var(Train.Norm[,1:14])
```

```
##                              Age   Experience       Income       Family
## Age                  1.000000000  0.993957874 -0.066305972 -0.039280819
## Experience           0.993957874  1.000000000 -0.057378872 -0.044830908
## Income              -0.066305972 -0.057378872  1.000000000 -0.152500025
## Family              -0.039280819 -0.044830908 -0.152500025  1.000000000
## CCAvg               -0.067416444 -0.066300108  0.646630189 -0.109629442
```

```
## Mortgage           -0.011570080 -0.008814267  0.223349358 -0.016447736
## Education_1        -0.017384381 -0.004452110  0.097255388 -0.055922656
## Education_2        -0.014228029 -0.014850917 -0.057634713  0.060607935
## Education_3         0.031612409  0.019303027 -0.039620675 -0.004685278
## Personal.Loan      -0.003524051 -0.003776466  0.154887688  0.018307374
## Securities.Account  0.002231682  0.001666677 -0.004940472  0.004677023
## CD.Account          0.006096927  0.006466316  0.043363815  0.004532897
## Online              0.001575653  0.001929163  0.000300238  0.003800124
## CreditCard          0.014668165  0.015174115 -0.002254922  0.001868983
##                           CCAvg      Mortgage   Education_1   Education_2
## Age                -0.067416444 -0.0115700800 -0.0173843807 -0.0142280288
## Experience         -0.066300108 -0.0088142667 -0.0044521102 -0.0148509167
## Income              0.646630189  0.2233493580  0.0972553880 -0.0576347131
## Family             -0.109629442 -0.0164477358 -0.0559226564  0.0606079348
## CCAvg               1.000000000  0.1224779189  0.0711136744 -0.0366222195
## Mortgage            0.122477919  1.0000000000  0.0181068656 -0.0187073637
## Education_1         0.071113674  0.0181068656  0.2416165388 -0.1156385462
## Education_2        -0.036622219 -0.0187073637 -0.1156385462  0.2031232633
## Education_3        -0.034491455  0.0006004981 -0.1259779927 -0.0874847171
## Personal.Loan       0.110521023  0.0477492500 -0.0216578860  0.0075691897
## Securities.Account  0.004314701 -0.0043211810 -0.0010776926  0.0005946427
## CD.Account          0.030312906  0.0255672711 -0.0022834278  0.0005724130
## Online             -0.010872239 -0.0025721435 -0.0036438813  0.0041569412
## CreditCard         -0.010253707 -0.0050302909  0.0009523174 -0.0029787707
##                     Education_3 Personal.Loan Securities.Account   CD.Account
## Age                0.0316124095  -0.003524051       0.0022316822  0.006096927
## Experience         0.0193030269  -0.003776466       0.0016666765  0.006466316
## Income            -0.0396206749   0.154887688      -0.0049404717  0.043363815
## Family            -0.0046852784   0.018307374       0.0046770232  0.004532897
## CCAvg             -0.0344914549   0.110521023       0.0043147012  0.030312906
## Mortgage           0.0006004981   0.047749250      -0.0043211810  0.025567271
## Education_1       -0.1259779927  -0.021657886      -0.0010776926 -0.002283428
## Education_2       -0.0874847171   0.007569190       0.0005946427  0.000572413
## Education_3        0.2134627098   0.014088696       0.0004830499  0.001711015
## Personal.Loan      0.0140886962   0.088425475       0.0015865288  0.021761254
## Securities.Account 0.0004830499   0.001586529       0.0976724464  0.022390908
## CD.Account         0.0017110148   0.021761254       0.0223909081  0.056712126
## Online            -0.0005130599  -0.001656552       0.0009487607  0.019818162
## CreditCard         0.0020264533   0.000312104      -0.0010639102  0.031887518
##                          Online    CreditCard
## Age                0.0015756531  0.0146681654
## Experience         0.0019291626  0.0151741149
## Income             0.0003002380 -0.0022549225
## Family             0.0038001238  0.0018689834
## CCAvg             -0.0108722392 -0.0102537073
## Mortgage          -0.0025721435 -0.0050302909
## Education_1       -0.0036438813  0.0009523174
## Education_2        0.0041569412 -0.0029787707
## Education_3       -0.0005130599  0.0020264533
## Personal.Loan     -0.0016565522  0.0003121040
## Securities.Account 0.0009487607 -0.0010639102
## CD.Account         0.0198181616  0.0318875181
## Online             0.2389845504  0.0038581749
## CreditCard         0.0038581749  0.2056881183
```

```
summary(Valid.Norm)
```

```
##       Age              Experience           Income             Family
##  Min.   :-1.95926   Min.   :-2.02646   Min.   :-1.44200   Min.   :-1.24282
##  1st Qu.:-0.91566   1st Qu.:-0.89447   1st Qu.:-0.79058   1st Qu.:-1.24282
##  Median :-0.04600   Median :-0.02371   Median :-0.26944   Median :-0.36903
##  Mean   :-0.04144   Mean   :-0.03651   Mean   :-0.03442   Mean   :-0.05665
##  3rd Qu.: 0.82365   3rd Qu.: 0.75997   3rd Qu.: 0.53399   3rd Qu.: 0.50476
##  Max.   : 1.86725   Max.   : 1.89196   Max.   : 3.11797   Max.   : 1.37855
##      CCAvg              Mortgage         Education_1     Education_2
##  Min.   :-1.11136   Min.   :-0.56041   Min.   :0.000   Min.   :0.0000
##  1st Qu.:-0.71158   1st Qu.:-0.56041   1st Qu.:0.000   1st Qu.:0.0000
##  Median :-0.25468   Median :-0.56041   Median :0.000   Median :0.0000
##  Mean   :-0.01142   Mean   :-0.04825   Mean   :0.436   Mean   :0.2765
##  3rd Qu.: 0.31644   3rd Qu.: 0.37817   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   : 3.97161   Max.   : 5.52123   Max.   :1.000   Max.   :1.0000
##   Education_3     Personal.Loan   Securities.Account   CD.Account
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000     Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000     1st Qu.:0.0000
##  Median :0.0000   Median :0.000   Median :0.0000     Median :0.0000
##  Mean   :0.2875   Mean   :0.093   Mean   :0.0965     Mean   :0.0605
##  3rd Qu.:1.0000   3rd Qu.:0.000   3rd Qu.:0.0000     3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000     Max.   :1.0000
##      Online        CreditCard
##  Min.   :0.000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:0.000
##  Median :1.000   Median :0.000
##  Mean   :0.584   Mean   :0.301
##  3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :1.000   Max.   :1.000
```

```
var(Valid.Norm[,1:14])
```

```
##                              Age   Experience        Income       Family
## Age                 0.9840159175  0.982983343 -0.0391087375 -0.058524094
## Experience          0.9829833433  0.992617829 -0.0309288242 -0.065615143
## Income             -0.0391087375 -0.030928824  0.9977069726 -0.167198213
## Family             -0.0585240935 -0.065615143 -0.1671982127  1.012688896
## CCAvg              -0.0285663154 -0.025600424  0.6413993992 -0.109441165
## Mortgage           -0.0142931587 -0.013575344  0.1675164735 -0.026911690
## Education_1        -0.0073862352  0.003710594  0.1236082390 -0.061956582
## Education_2         0.0029569086  0.001711506 -0.0578550214  0.065699374
## Education_3         0.0044293266 -0.005422101 -0.0657532177 -0.003742791
## Personal.Loan      -0.0005118333  0.000102017  0.1375728430  0.017705315
## Securities.Account -0.0040082007 -0.003729978  0.0051397161  0.007869150
## CD.Account         -0.0043658097 -0.003537619  0.0360354756  0.001635900
## Online              0.0138640914  0.013664522  0.0165298727  0.006311931
## CreditCard         -0.0129907337 -0.012305893  0.0009076062  0.010833878
##                             CCAvg      Mortgage   Education_1   Education_2
## Age                -0.028566315 -0.0142931587 -0.0073862352  0.0029569086
## Experience         -0.025600424 -0.0135753440  0.0037105943  0.0017115064
## Income              0.641399399  0.1675164735  0.1236082390 -0.0578550214
```

```
## Family             -0.109441165 -0.0269116896 -0.0619565823  0.0656993736
## CCAvg               0.991049208  0.0831597444  0.0868624388 -0.0464567193
## Mortgage            0.083159744  0.8714975473  0.0251480852 -0.0069481725
## Education_1         0.086862439  0.0251480852  0.2460270135 -0.1206143072
## Education_2        -0.046456719 -0.0069481725 -0.1206143072  0.2001478239
## Education_3        -0.040405720 -0.0181999127 -0.1254127064 -0.0795335168
## Personal.Loan       0.103963159  0.0302052529 -0.0215587794  0.0122916458
## Securities.Account  0.004952259  0.0020703832  0.0044282141  0.0008181591
## CD.Account          0.035718975  0.0134840622 -0.0008784392  0.0007721361
## Online              0.011742996 -0.0039237492  0.0078799400  0.0050265133
## CreditCard          0.007855158 -0.0001415863  0.0067673837 -0.0017273637
##                      Education_3 Personal.Loan Securities.Account    CD.Account
## Age                 0.0044293266 -0.0005118333      -0.0040082007 -0.0043658097
## Experience         -0.0054221007  0.0001020170      -0.0037299784 -0.0035376185
## Income             -0.0657532177  0.1375728430       0.0051397161  0.0360354756
## Family             -0.0037427913  0.0177053151       0.0078691504  0.0016359003
## CCAvg              -0.0404057195  0.1039631588       0.0049522595  0.0357189755
## Mortgage           -0.0181999127  0.0302052529       0.0020703832  0.0134840622
## Education_1        -0.1254127064 -0.0215587794       0.0044282141 -0.0008784392
## Education_2        -0.0795335168  0.0122916458       0.0008181591  0.0007721361
## Education_3         0.2049462231  0.0092671336      -0.0052463732  0.0001063032
## Personal.Loan       0.0092671336  0.0843931966       0.0025267634  0.0228849425
## Securities.Account -0.0052463732  0.0025267634       0.0872313657  0.0241738369
## CD.Account          0.0001063032  0.0228849425       0.0241738369  0.0568681841
## Online             -0.0129064532  0.0046903452       0.0031455728  0.0216788394
## CreditCard         -0.0050400200  0.0005072536      -0.0035482741  0.0278034017
##                          Online    CreditCard
## Age                 0.013864091 -0.0129907337
## Experience          0.013664522 -0.0123058927
## Income              0.016529873  0.0009076062
## Family              0.006311931  0.0108338782
## CCAvg               0.011742996  0.0078551584
## Mortgage           -0.003923749 -0.0001415863
## Education_1         0.007879940  0.0067673837
## Education_2         0.005026513 -0.0017273637
## Education_3        -0.012906453 -0.0050400200
## Personal.Loan       0.004690345  0.0005072536
## Securities.Account  0.003145573 -0.0035482741
## CD.Account          0.021678839  0.0278034017
## Online              0.243065533 -0.0032856428
## CreditCard         -0.003285643  0.2105042521
```

#Performing KNN classification with k=1

```
nn1 <-knn(train=Train.Norm[,-10], test=Valid.Norm[,-10], cl=Train.Norm[,10], k=1, prob=TRUE)
confusionMatrix(nn1, as.factor(Valid.Norm$Personal.Loan))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1798   64
##          1   16  122
```

```
##
##                Accuracy : 0.96
##                  95% CI : (0.9505, 0.9682)
##     No Information Rate : 0.907
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.7318
##
##  Mcnemar's Test P-Value : 1.482e-07
##
##             Sensitivity : 0.9912
##             Specificity : 0.6559
##          Pos Pred Value : 0.9656
##          Neg Pred Value : 0.8841
##              Prevalence : 0.9070
##          Detection Rate : 0.8990
##    Detection Prevalence : 0.9310
##       Balanced Accuracy : 0.8235
##
##        'Positive' Class : 0
##
```

```
head (nn1)
```

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

#Classifying test customer with k=1 (Question 1)

```
PredNN <-knn(train=Train.Norm[,-10], test=Test.Norm, cl=Train.Norm[,10], k=1, prob=TRUE)
PredNN
```

```
## [1] 0
## attr(,"prob")
## [1] 1
## attr(,"nn.index")
##      [,1]
## [1,] 2443
## attr(,"nn.dist")
##           [,1]
## [1,] 0.4773902
## Levels: 0
```

```
#Test customer would classify as 0 with a probability of 100%
```

#Hypertuning to find best k (Question 2)

```
sqrt(NROW(Train.Norm))
```

```
## [1] 54.77226
```

```
accuracy.df <- data.frame(k = seq(1, 55, 1), accuracy = rep(0, 55))
for (i in 1:55) {
  prediction<-knn(train=Train.Norm[,-10], test=Valid.Norm[,-10],   cl=Train.Norm[,10], k=i, prob=TRUE)
  accuracy.df[i, 2] <-confusionMatrix(prediction, as.factor(Valid.Norm[,10]))$overall[1]
  }
 accuracy.df
```

```
##     k accuracy
## 1   1   0.9600
## 2   2   0.9575
## 3   3   0.9655
## 4   4   0.9560
## 5   5   0.9595
## 6   6   0.9530
## 7   7   0.9560
## 8   8   0.9515
## 9   9   0.9520
## 10 10   0.9495
## 11 11   0.9500
## 12 12   0.9475
## 13 13   0.9500
## 14 14   0.9440
## 15 15   0.9480
## 16 16   0.9445
## 17 17   0.9470
## 18 18   0.9445
## 19 19   0.9465
## 20 20   0.9410
## 21 21   0.9440
## 22 22   0.9390
## 23 23   0.9405
## 24 24   0.9375
## 25 25   0.9400
## 26 26   0.9370
## 27 27   0.9385
## 28 28   0.9340
## 29 29   0.9370
## 30 30   0.9335
## 31 31   0.9360
## 32 32   0.9325
## 33 33   0.9340
## 34 34   0.9325
## 35 35   0.9335
## 36 36   0.9330
## 37 37   0.9330
## 38 38   0.9305
## 39 39   0.9325
## 40 40   0.9295
## 41 41   0.9315
## 42 42   0.9290
## 43 43   0.9290
## 44 44   0.9285
## 45 45   0.9285
```

```
## 46 46   0.9285
## 47 47   0.9290
## 48 48   0.9275
## 49 49   0.9285
## 50 50   0.9275
## 51 51   0.9280
## 52 52   0.9275
## 53 53   0.9275
## 54 54   0.9270
## 55 55   0.9270
```

```
#The choice of k that balances between overfitting and ignoring predictor information is k=3
```

#Confusion Matrix for k=3 (Question 3)

```
nn3<-knn(train=Train.Norm[,-10], test=Valid.Norm[,-10], cl=Train.Norm[,10], k=3, prob=TRUE)
  confusionMatrix(nn3, as.factor(Valid.Norm$Personal.Loan))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1810   65
##          1    4  121
##
##                Accuracy : 0.9655
##                  95% CI : (0.9565, 0.9731)
##     No Information Rate : 0.907
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.7602
##
##  Mcnemar's Test P-Value : 5.08e-13
##
##             Sensitivity : 0.9978
##             Specificity : 0.6505
##          Pos Pred Value : 0.9653
##          Neg Pred Value : 0.9680
##              Prevalence : 0.9070
##          Detection Rate : 0.9050
##    Detection Prevalence : 0.9375
##       Balanced Accuracy : 0.8242
##
##        'Positive' Class : 0
##
```

#Classify customer using the best k (Question 4)

```
PredNN3 <-knn(train=Train.Norm[,-10], test=Test.Norm, cl=Train.Norm[,10], k=3, prob=TRUE)
PredNN3
```

```
## [1] 0
```

```
## attr(,"prob")
## [1] 1
## attr(,"nn.index")
##      [,1] [,2] [,3]
## [1,] 2443 2067 1021
## attr(,"nn.dist")
##           [,1]      [,2]      [,3]
## [1,] 0.4773902 0.6338079 0.7058524
## Levels: 0
```

```
#Test customer would classify as "0"
```

#Repartition the data into training, validation, and test sets

```
set.seed(321)
Training.Index_1 = createDataPartition(Dataset$Personal.Loan, p= 0.5 , list=FALSE)
Training.Data_1  = Dataset [Training.Index_1,] #50% Training data
Remaining.Data = Dataset[-Training.Index_1,] #50% remaining data [training + validation]
Validation.Index_1 = createDataPartition(Remaining.Data$Personal.Loan, p= 0.6, list=FALSE)
Validation.Data_1 = Remaining.Data[Validation.Index_1,] #Validation data
Test.Data_1 = Remaining.Data[-Validation.Index_1,] #Test data
```

#Normalizing the data

```
train.norm_1 <- Training.Data_1
valid.norm_1 <- Validation.Data_1
test.norm_1 <- Test.Data_1
rem_data.norm_1 <- Remaining.Data

norm.values_1 <- preProcess(Training.Data_1[,1:6], method=c("center", "scale"))

train.norm_1[, 1:6] <- predict(norm.values_1, train.norm_1[, 1:6])  #Normalized Training Data
valid.norm_1[, 1:6] <- predict(norm.values_1, valid.norm_1[, 1:6])#Normalized Validation Data
test.norm_1[, 1:6] <- predict(norm.values_1, test.norm_1[, 1:6]) #Normalized Test Data
summary(train.norm_1)
```

```
##       Age              Experience          Income             Family
##  Min.   :-1.96686   Min.   :-2.03332   Min.   :-1.4388   Min.   :-1.2349
##  1st Qu.:-0.83672   1st Qu.:-0.90170   1st Qu.:-0.7672   1st Qu.:-1.2349
##  Median : 0.03262   Median : 0.05581   Median :-0.2039   Median :-0.3601
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.90195   3rd Qu.: 0.83924   3rd Qu.: 0.5111   3rd Qu.: 0.5147
##  Max.   : 1.85823   Max.   : 1.97085   Max.   : 3.2411   Max.   : 1.3895
##      CCAvg            Mortgage        Education_1     Education_2
##  Min.   :-1.1053   Min.   :-0.5529   Min.   :0.000   Min.   :0.0000
##  1st Qu.:-0.7082   1st Qu.:-0.5529   1st Qu.:0.000   1st Qu.:0.0000
##  Median :-0.1976   Median :-0.5529   Median :0.000   Median :0.0000
##  Mean   : 0.0000   Mean   : 0.0000   Mean   :0.412   Mean   :0.2804
##  3rd Qu.: 0.3696   3rd Qu.: 0.4192   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   : 4.5675   Max.   : 5.3276   Max.   :1.000   Max.   :1.0000
##   Education_3     Personal.Loan    Securities.Account   CD.Account
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000     Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000     1st Qu.:0.0000
```

```
##  Median :0.0000   Median :0.0000   Median :0.0000     Median :0.0000
##  Mean   :0.3076   Mean   :0.0944   Mean   :0.1152     Mean   :0.0616
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000     3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000     Max.   :1.0000
##      Online         CreditCard
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.0000
##  Mean   :0.6052   Mean   :0.2924
##  3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000
```

```
var(train.norm_1)
```

```
##                              Age   Experience        Income       Family
## Age                 1.000000000  0.993865664 -6.579987e-02 -0.043545961
## Experience          0.993865664  1.000000000 -5.647708e-02 -0.049949619
## Income             -0.065799873 -0.056477076  1.000000e+00 -0.146044032
## Family             -0.043545961 -0.049949619 -1.460440e-01  1.000000000
## CCAvg              -0.070342389 -0.069436140  6.439289e-01 -0.121305463
## Mortgage           -0.022092863 -0.019130405  2.305317e-01  0.005330148
## Education_1        -0.019743384 -0.006568209  1.046106e-01 -0.058442081
## Education_2        -0.013810072 -0.014020849 -6.463914e-02  0.054423514
## Education_3         0.033553456  0.020589058 -3.997149e-02  0.004018567
## Personal.Loan      -0.005059926 -0.005248492  1.498640e-01  0.019205218
## Securities.Account  0.003237233  0.002531145 -6.491463e-03  0.005061610
## CD.Account          0.008236990  0.008629640  4.410424e-02  0.004765598
## Online              0.002736990  0.003836329 -9.115895e-05  0.006038282
## CreditCard          0.013646098  0.014132258 -5.477936e-03 -0.002058222
##                            CCAvg     Mortgage  Education_1   Education_2
## Age                -0.070342389 -0.022092863 -0.019743384 -0.0138100719
## Experience         -0.069436140 -0.019130405 -0.006568209 -0.0140208493
## Income              0.643928885  0.230531731  0.104610632 -0.0646391431
## Family             -0.121305463  0.005330148 -0.058442081  0.0544235137
## CCAvg               1.000000000  0.133764248  0.073995871 -0.0414890631
## Mortgage            0.133764248  1.000000000  0.020204108 -0.0185841192
## Education_1         0.073995871  0.020204108  0.242352941 -0.1155710284
## Education_2        -0.041489063 -0.018584119 -0.115571028  0.2018565826
## Education_3        -0.032506808 -0.001619988 -0.126781913 -0.0862855542
## Personal.Loan       0.100770841  0.046439936 -0.020901160  0.0055324530
## Securities.Account  0.002345128 -0.003215187 -0.001062825  0.0004981192
## CD.Account          0.028762794  0.023984627 -0.003380552  0.0011278111
## Online             -0.010750703 -0.006163855 -0.002943577  0.0043036415
## CreditCard         -0.014410272 -0.008566771 -0.002869948 -0.0007892757
##                     Education_3 Personal.Loan Securities.Account   CD.Account
## Age                0.0335534561 -5.059926e-03       0.0032372326  0.008236990
## Experience         0.0205890578 -5.248492e-03       0.0025311449  0.008629640
## Income            -0.0399714888  1.498640e-01      -0.0064914631  0.044104243
## Family             0.0040185669  1.920522e-02       0.0050616104  0.004765598
## CCAvg             -0.0325068077  1.007708e-01       0.0023451276  0.028762794
## Mortgage          -0.0016199884  4.643994e-02      -0.0032151868  0.023984627
## Education_1       -0.1267819128 -2.090116e-02      -0.0010628251 -0.003380552
## Education_2       -0.0862855542  5.532453e-03       0.0004981192  0.001127811
## Education_3        0.2130674670  1.536871e-02       0.0005647059  0.002252741
```

```
## Personal.Loan       0.0153687075  8.552285e-02       0.0015257303  0.022193838
## Securities.Account  0.0005647059  1.525730e-03       0.1019697479  0.023313005
## CD.Account          0.0022527411  2.219384e-02       0.0233130052  0.057828571
## Online             -0.0013600640  6.914766e-05       0.0014815526  0.021128131
## CreditCard          0.0036592237 -4.027211e-04      -0.0020853141  0.032000960
##                           Online    CreditCard
## Age                  2.736990e-03  0.0136460979
## Experience           3.836329e-03  0.0141322584
## Income              -9.115895e-05 -0.0054779363
## Family               6.038282e-03 -0.0020582221
## CCAvg               -1.075070e-02 -0.0144102718
## Mortgage            -6.163855e-03 -0.0085667707
## Education_1         -2.943577e-03 -0.0028699480
## Education_2          4.303641e-03 -0.0007892757
## Education_3         -1.360064e-03  0.0036592237
## Personal.Loan        6.914766e-05 -0.0004027211
## Securities.Account   1.481553e-03 -0.0020853141
## CD.Account           2.112813e-02  0.0320009604
## Online               2.390286e-01  0.0066421769
## CreditCard           6.642177e-03  0.2069850340
```

```
summary(valid.norm_1)
```

```
##       Age               Experience          Income             Family
##  Min.   :-1.96686   Min.   :-2.03332   Min.   :-1.43882   Min.   :-1.23487
##  1st Qu.:-0.92365   1st Qu.:-0.90170   1st Qu.:-0.78884   1st Qu.:-1.23487
##  Median :-0.05432   Median :-0.03123   Median :-0.26885   Median :-0.36007
##  Mean   :-0.04719   Mean   :-0.04069   Mean   :-0.03705   Mean   :-0.05914
##  3rd Qu.: 0.81502   3rd Qu.: 0.75219   3rd Qu.: 0.53280   3rd Qu.: 0.51473
##  Max.   : 1.85823   Max.   : 1.88380   Max.   : 3.11106   Max.   : 1.38954
##      CCAvg              Mortgage         Education_1      Education_2
##  Min.   :-1.10529   Min.   :-0.55292   Min.   :0.0000   Min.   :0.000
##  1st Qu.:-0.70819   1st Qu.:-0.55292   1st Qu.:0.0000   1st Qu.:0.000
##  Median :-0.25437   Median :-0.55292   Median :0.0000   Median :0.000
##  Mean   :-0.02513   Mean   :-0.04451   Mean   :0.4267   Mean   :0.272
##  3rd Qu.: 0.31291   3rd Qu.: 0.38110   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   : 4.00024   Max.   : 5.49913   Max.   :1.0000   Max.   :1.000
##   Education_3     Personal.Loan     Securities.Account   CD.Account
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000    Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000    1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000    Median :0.00000
##  Mean   :0.3013   Mean   :0.09733   Mean   :0.09867    Mean   :0.06333
##  3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000    Max.   :1.00000
##      Online         CreditCard
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.0000
##  Mean   :0.5987   Mean   :0.2913
##  3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000
```

```
var(valid.norm_1)
```

```
##                                Age   Experience       Income        Family
## Age                  0.9681981329  0.964112110 -0.042124592 -0.0466909244
## Experience           0.9641121099  0.970181431 -0.035308984 -0.0515075147
## Income              -0.0421245917 -0.035308984  0.976073451 -0.1540250410
## Family              -0.0466909244 -0.051507515 -0.154025041  1.0121202077
## CCAvg               -0.0371782487 -0.034103350  0.613028717 -0.0896178584
## Mortgage             0.0180637273  0.019183787  0.146037012 -0.0476401320
## Education_1         -0.0148744251 -0.004033164  0.106290487 -0.0514494401
## Education_2          0.0185317892  0.016104937 -0.040847003  0.0622388826
## Education_3         -0.0036573641 -0.012071773 -0.065443484 -0.0107894425
## Personal.Loan        0.0039452487  0.004405510  0.146835195  0.0208785750
## Securities.Account  -0.0060393135 -0.005860282  0.013609578  0.0053036809
## CD.Account           0.0001281678  0.000947898  0.040884158  0.0025211159
## Online               0.0168395101  0.017454874  0.016428052 -0.0005322356
## CreditCard          -0.0159968503 -0.015476424 -0.006566572  0.0219850648
##                            CCAvg     Mortgage  Education_1  Education_2
## Age                 -0.037178249  0.018063727 -0.014874425  0.018531789
## Experience          -0.034103350  0.019183787 -0.004033164  0.016104937
## Income               0.613028717  0.146037012  0.106290487 -0.040847003
## Family              -0.089617858 -0.047640132 -0.051449440  0.062238883
## CCAvg                0.938276300  0.067918122  0.073323424 -0.021857004
## Mortgage             0.067918122  0.849690347  0.016583598 -0.007267807
## Education_1          0.073323424  0.016583598  0.244785412 -0.116130754
## Education_2         -0.021857004 -0.007267807 -0.116130754  0.198148099
## Education_3         -0.051466420 -0.009315791 -0.128654659 -0.082017345
## Personal.Loan        0.114781042  0.031598909 -0.022210362  0.017537025
## Securities.Account   0.012197962 -0.006471350  0.001236380  0.003831888
## CD.Account           0.038521367  0.010245323 -0.001690016  0.002775183
## Online               0.006767346 -0.011452915  0.002570603  0.006500334
## CreditCard           0.006825313 -0.005786096  0.005701579 -0.001910607
##                     Education_3 Personal.Loan Securities.Account    CD.Account
## Age                -0.003657364   0.003945249       -0.006039314  0.0001281678
## Experience         -0.012071773   0.004405510       -0.005860282  0.0009478980
## Income             -0.065443484   0.146835195        0.013609578  0.0408841584
## Family             -0.010789442   0.020878575        0.005303681  0.0025211159
## CCAvg              -0.051466420   0.114781042        0.012197962  0.0385213667
## Mortgage           -0.009315791   0.031598909       -0.006471350  0.0102453226
## Education_1        -0.128654659  -0.022210362        0.001236380 -0.0016900156
## Education_2        -0.082017345   0.017537025        0.003831888  0.0027751835
## Education_3         0.210672004   0.004673338       -0.005068268 -0.0010851679
## Personal.Loan       0.004673338   0.087918168        0.007067823  0.0251856793
## Securities.Account -0.005068268   0.007067823        0.088990883  0.0257682900
## CD.Account         -0.001085168   0.025185679        0.025768290  0.0593617968
## Online             -0.009070936   0.004399377        0.004934845  0.0220991772
## CreditCard         -0.003790972  -0.001690905       -0.005415166  0.0269001557
##                           Online   CreditCard
## Age                  0.0168395101 -0.015996850
## Experience           0.0174548736 -0.015476424
## Income               0.0164280518 -0.006566572
## Family              -0.0005322356  0.021985065
## CCAvg                0.0067673465  0.006825313
```

```
## Mortgage           -0.0114529149 -0.005786096
## Education_1         0.0025706026  0.005701579
## Education_2         0.0065003336 -0.001910607
## Education_3        -0.0090709362 -0.003790972
## Personal.Loan       0.0043993774 -0.001690905
## Securities.Account  0.0049348455 -0.005415166
## CD.Account          0.0220991772  0.026900156
## Online              0.2404251723 -0.005081610
## CreditCard         -0.0050816100  0.206595953
```

```
summary(test.norm_1)
```

```
##       Age              Experience          Income             Family
##  Min.   :-1.87992   Min.   :-1.9463   Min.   :-1.43882   Min.   :-1.23487
##  1st Qu.:-0.92365   1st Qu.:-0.9888   1st Qu.:-0.78884   1st Qu.:-1.23487
##  Median :-0.14125   Median :-0.1183   Median :-0.23635   Median :-0.36007
##  Mean   :-0.05371   Mean   :-0.0496   Mean   :-0.01317   Mean   : 0.02222
##  3rd Qu.: 0.81502   3rd Qu.: 0.8392   3rd Qu.: 0.53280   3rd Qu.: 1.38954
##  Max.   : 1.85823   Max.   : 1.8838   Max.   : 2.82940   Max.   : 1.38954
##      CCAvg              Mortgage          Education_1     Education_2
##  Min.   :-1.105292   Min.   :-0.552919   Min.   :0.000   Min.   :0.000
##  1st Qu.:-0.708194   1st Qu.:-0.552919   1st Qu.:0.000   1st Qu.:0.000
##  Median :-0.197641   Median :-0.552919   Median :0.000   Median :0.000
##  Mean   : 0.008016   Mean   :-0.005433   Mean   :0.426   Mean   :0.294
##  3rd Qu.: 0.312913   3rd Qu.: 0.428752   3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   : 3.943515   Max.   : 4.889156   Max.   :1.000   Max.   :1.000
##   Education_3   Personal.Loan   Securities.Account   CD.Account
##  Min.   :0.00   Min.   :0.000   Min.   :0.000      Min.   :0.000
##  1st Qu.:0.00   1st Qu.:0.000   1st Qu.:0.000      1st Qu.:0.000
##  Median :0.00   Median :0.000   Median :0.000      Median :0.000
##  Mean   :0.28   Mean   :0.098   Mean   :0.086      Mean   :0.053
##  3rd Qu.:1.00   3rd Qu.:0.000   3rd Qu.:0.000      3rd Qu.:0.000
##  Max.   :1.00   Max.   :1.000   Max.   :1.000      Max.   :1.000
##      Online        CreditCard
##  Min.   :0.000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:0.000
##  Median :1.000   Median :0.000
##  Mean   :0.573   Mean   :0.302
##  3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :1.000   Max.   :1.000
```

```
var(test.norm_1)
```

```
##                              Age    Experience       Income       Family
## Age                 1.011982398  1.0135563063 -0.048695414 -0.054739151
## Experience          1.013556306  1.0267712706 -0.039174980 -0.062412811
## Income             -0.048695414 -0.0391749800  1.010349755 -0.194830076
## Family             -0.054739151 -0.0624128112 -0.194830076  1.018867442
## CCAvg              -0.026003048 -0.0236108802  0.664809181 -0.108370325
## Mortgage           -0.034099036 -0.0335615346  0.203209930 -0.044291970
## Education_1         0.004265583  0.0161099038  0.116285103 -0.070196225
## Education_2        -0.029592103 -0.0297969524 -0.065182564  0.083646664
## Education_3         0.025326519  0.0136870487 -0.051102538 -0.013450438
```

```
## Personal.Loan      -0.004410734 -0.0042105014  0.143714275  0.011536203
## Securities.Account -0.001096636 -0.0006843524 -0.008737713  0.009998509
## CD.Account         -0.011432003 -0.0108758362  0.030222974  0.001610375
## Online              0.000260105 -0.0027967482  0.010150478  0.011908366
## CreditCard          0.007560885  0.0085149336  0.018089754 -0.001728591
##                           CCAvg     Mortgage  Education_1  Education_2
## Age                -0.026003048 -0.034099036  0.004265583 -0.029592103
## Experience         -0.023610880 -0.033561535  0.016109904 -0.029796952
## Income              0.664809181  0.203209930  0.116285103 -0.065182564
## Family             -0.108370325 -0.044291970 -0.070196225  0.083646664
## CCAvg               1.008232075  0.090834670  0.089414851 -0.065293510
## Mortgage            0.090834670  0.924135362  0.027665553 -0.012436150
## Education_1         0.089414851  0.027665553  0.244768769 -0.125369369
## Education_2        -0.065293510 -0.012436150 -0.125369369  0.207771772
## Education_3        -0.024121340 -0.015229403 -0.119399399 -0.082402402
## Personal.Loan       0.111981101  0.039596981 -0.022770771  0.007195195
## Securities.Account -0.001350108  0.009242814  0.006370370 -0.003287287
## CD.Account          0.031797252  0.028005302  0.002424424 -0.003585586
## Online              0.008211281  0.018332563  0.007909910  0.002540541
## CreditCard          0.010519947  0.013994809  0.015363363 -0.007795796
##                    Education_3 Personal.Loan Securities.Account   CD.Account
## Age                 0.025326519  -0.004410734      -0.0010966355 -0.011432003
## Experience          0.013687049  -0.004210501      -0.0006843524 -0.010875836
## Income             -0.051102538   0.143714275      -0.0087377125  0.030222974
## Family             -0.013450438   0.011536203       0.0099985093  0.001610375
## CCAvg              -0.024121340   0.111981101      -0.0013501077  0.031797252
## Mortgage           -0.015229403   0.039596981       0.0092428141  0.028005302
## Education_1        -0.119399399  -0.022770771       0.0063703704  0.002424424
## Education_2        -0.082402402   0.007195195      -0.0032872873 -0.003585586
## Education_3         0.201801802   0.015575576      -0.0030830831  0.001161161
## Personal.Loan       0.015575576   0.088484484      -0.0044324324  0.017823824
## Securities.Account -0.003083083  -0.004432432       0.0786826827  0.018460460
## CD.Account          0.001161161   0.017823824       0.0184604605  0.050241241
## Online             -0.010450450  -0.002156156      -0.0022802803  0.016647648
## CreditCard         -0.007567568   0.005409409       0.0030310310  0.031025025
##                          Online   CreditCard
## Age                 0.000260105  0.007560885
## Experience         -0.002796748  0.008514934
## Income              0.010150478  0.018089754
## Family              0.011908366 -0.001728591
## CCAvg               0.008211281  0.010519947
## Mortgage            0.018332563  0.013994809
## Education_1         0.007909910  0.015363363
## Education_2         0.002540541 -0.007795796
## Education_3        -0.010450450 -0.007567568
## Personal.Loan      -0.002156156  0.005409409
## Securities.Account -0.002280280  0.003031031
## CD.Account          0.016647648  0.031025025
## Online              0.244915916 -0.004050050
## CreditCard         -0.004050050  0.211007007
```

#Perform KNN on train,validation and test sets with k=3 and compare confusion matrices (Question 5)

```
nnTrain <-knn(train=train.norm_1[, -10], test=train.norm_1[,-10], cl=train.norm_1[,10], k=3, prob=TRUE)
nnValidation <-knn(train=train.norm_1[, -10], test=valid.norm_1[,-10], cl=train.norm_1[,10], k=3, prob=
nnTest <-knn(train=train.norm_1[, -10], test=test.norm_1[,-10], cl=train.norm_1[,10], k=3, prob=TRUE)
confusionMatrix(nnTest, as.factor(test.norm_1[,10]))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 901  44
##          1   1  54
##
##                Accuracy : 0.955
##                  95% CI : (0.9402, 0.967)
##     No Information Rate : 0.902
##     P-Value [Acc > NIR] : 3.583e-10
##
##                   Kappa : 0.6836
##
##  Mcnemar's Test P-Value : 3.825e-10
##
##             Sensitivity : 0.9989
##             Specificity : 0.5510
##          Pos Pred Value : 0.9534
##          Neg Pred Value : 0.9818
##              Prevalence : 0.9020
##          Detection Rate : 0.9010
##    Detection Prevalence : 0.9450
##       Balanced Accuracy : 0.7750
##
##        'Positive' Class : 0
##
```

```
confusionMatrix(nnTrain, as.factor(train.norm_1[,10]))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 2261   60
##          1    3  176
##
##                Accuracy : 0.9748
##                  95% CI : (0.9679, 0.9806)
##     No Information Rate : 0.9056
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8347
##
##  Mcnemar's Test P-Value : 1.722e-12
##
##             Sensitivity : 0.9987
##             Specificity : 0.7458
```

```
##          Pos Pred Value : 0.9741
##          Neg Pred Value : 0.9832
##              Prevalence : 0.9056
##          Detection Rate : 0.9044
##    Detection Prevalence : 0.9284
##       Balanced Accuracy : 0.8722
##
##        'Positive' Class : 0
##
```

```
confusionMatrix(nnValidation, as.factor(valid.norm_1[,10]))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1349   49
##          1    5   97
##
##                Accuracy : 0.964
##                  95% CI : (0.9533, 0.9728)
##     No Information Rate : 0.9027
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.7633
##
##  Mcnemar's Test P-Value : 4.87e-09
##
##             Sensitivity : 0.9963
##             Specificity : 0.6644
##          Pos Pred Value : 0.9649
##          Neg Pred Value : 0.9510
##              Prevalence : 0.9027
##          Detection Rate : 0.8993
##    Detection Prevalence : 0.9320
##       Balanced Accuracy : 0.8303
##
##        'Positive' Class : 0
##
```

"‘ #Comments on Matrices The accuracy levels differ for Test (0.955), training (0.9748) and validation (0.964) data sets. The confusion matrix of the training set presents the highest accuracy. This is most likely because the model is performing on data it has already "seen", therefore it is more closely fitted to that data.