Assignment 2

Nourah

2/21/2021

```r
### First I Will Install Packages
library(caret)
```

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

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

```r
library(class)
library(gmodels)
library(ggplot2)
library(FNN)
```

```
##
## Attaching package: 'FNN'
```

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

```r
library(dummies)
```

```
## dummies-1.5.6 provided by Decision Patterns
```

```r
library(fastDummies)
library(e1071)
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
```

```r
###Second I will Import the CSV File
df <-read.csv(file="~/Desktop/spring 2021/ML/ML2/UniversalBank.csv")
```

```r
###Perform a k-NN classification with all predictors except ID and ZIP code using k = 1
df<-df[,c(-1,-5)]
str(df)
```

```
## 'data.frame':    5000 obs. of  12 variables:
##  $ Age               : int  25 45 39 35 35 37 53 50 35 34 ...
##  $ Experience        : int  1 19 15 9 8 13 27 24 10 9 ...
##  $ Income            : int  49 34 11 100 45 29 72 22 81 180 ...
##  $ Family            : int  4 3 1 1 4 4 2 1 3 1 ...
##  $ CCAvg             : num  1.6 1.5 1 2.7 1 0.4 1.5 0.3 0.6 8.9 ...
##  $ Education         : int  1 1 1 2 2 2 2 3 2 3 ...
##  $ Mortgage          : int  0 0 0 0 0 155 0 0 104 0 ...
##  $ Personal.Loan     : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ Securities.Account: int  1 1 0 0 0 0 0 0 0 0 ...
##  $ CD.Account        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Online            : int  0 0 0 0 0 1 1 0 1 0 ...
##  $ CreditCard        : int  0 0 0 0 1 0 0 1 0 0 ...
```

```r
###Transform categorical predictors with more than two categories into dummy variables
dummymodel <- dummyVars(~Education,data=df)
head(predict(dummymodel,df))
```

```
##   Education
## 1         1
## 2         1
## 3         1
## 4         2
## 5         2
## 6         2
```

```r
dummymodel <- dummyVars(~Family, data=df)
head(predict(dummymodel,df))
```

```
##   Family
## 1      4
## 2      3
## 3      1
## 4      1
## 5      4
## 6      4
```

```r
###transform categorical predictors with more than two categories into dummy variables
df$Education<-as.factor(df$Education)
dummy_model<-dummyVars(~.,data=df)
head(predict(dummy_model,df))
```

```
##   Age Experience Income Family CCAvg Education.1 Education.2 Education.3
## 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           0           1           0
## 5  35          8     45      4   1.0           0           1           0
## 6  37         13     29      4   0.4           0           1           0
##   Mortgage 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      155             0                  0          0      1          0
```

```r
df1<-data.frame(predict(dummy_model, newdata=df))

df$Family<-as.factor(df$Family)
dummy_model<-dummyVars(~.,data=df)
head(predict(dummy_model,df))
```

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

```r
df1<-data.frame(predict(dummy_model, newdata=df))
```

```r
###Preparation for data splitting and validation
set.seed(2)
train.rows <-sample(rownames(df1), dim(df1)[1]*.6)
train.data <- df1[train.rows, ]
valid.rows <- setdiff(rownames(df1), train.rows)
valid.data <- df1[valid.rows, ]
summary(train.data)
```

```
##       Age          Experience        Income          Family.1
##  Min.   :23.00   Min.   :-3.00   Min.   :  8.00   Min.   :0.0000
##  1st Qu.:35.00   1st Qu.:10.00   1st Qu.: 39.00   1st Qu.:0.0000
##  Median :45.00   Median :20.00   Median : 64.00   Median :0.0000
##  Mean   :45.27   Mean   :20.03   Mean   : 74.27   Mean   :0.2937
##  3rd Qu.:56.00   3rd Qu.:30.00   3rd Qu.:101.00   3rd Qu.:1.0000
##  Max.   :67.00   Max.   :43.00   Max.   :224.00   Max.   :1.0000
##     Family.2        Family.3        Family.4          CCAvg
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000   Min.   : 0.000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.: 0.700
##  Median :0.000   Median :0.000   Median :0.0000   Median : 1.600
##  Mean   :0.259   Mean   :0.197   Mean   :0.2503   Mean   : 1.963
##  3rd Qu.:1.000   3rd Qu.:0.000   3rd Qu.:1.0000   3rd Qu.: 2.600
##  Max.   :1.000   Max.   :1.000   Max.   :1.0000   Max.   :10.000
##   Education.1      Education.2      Education.3       Mortgage
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :  0.00
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:  0.00
##  Median :0.0000   Median :0.0000   Median :0.000   Median :  0.00
##  Mean   :0.4183   Mean   :0.2767   Mean   :0.305   Mean   : 56.67
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:102.00
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :601.00
##  Personal.Loan     Securities.Account   CD.Account          Online
##  Min.   :0.00000   Min.   :0.0000     Min.   :0.00000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.0000     1st Qu.:0.00000   1st Qu.:0.000
##  Median :0.00000   Median :0.0000     Median :0.00000   Median :1.000
##  Mean   :0.09733   Mean   :0.1093     Mean   :0.06267   Mean   :0.604
##  3rd Qu.:0.00000   3rd Qu.:0.0000     3rd Qu.:0.00000   3rd Qu.:1.000
##  Max.   :1.00000   Max.   :1.0000     Max.   :1.00000   Max.   :1.000
##    CreditCard
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.3023
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```r
summary(valid.data)
```

```
##       Age          Experience        Income          Family.1
##  Min.   :23.00   Min.   :-2.00   Min.   :  8.00   Min.   :0.0000
##  1st Qu.:36.00   1st Qu.:11.00   1st Qu.: 38.00   1st Qu.:0.0000
##  Median :46.00   Median :20.00   Median : 63.00   Median :0.0000
##  Mean   :45.44   Mean   :20.22   Mean   : 73.04   Mean   :0.2955
##  3rd Qu.:55.00   3rd Qu.:29.00   3rd Qu.: 94.25   3rd Qu.:1.0000
##  Max.   :67.00   Max.   :43.00   Max.   :205.00   Max.   :1.0000
##     Family.2         Family.3         Family.4          CCAvg
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   : 0.0
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 0.7
##  Median :0.0000   Median :0.0000   Median :0.0000   Median : 1.5
##  Mean   :0.2595   Mean   :0.2095   Mean   :0.2355   Mean   : 1.9
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.: 2.5
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :10.0
##   Education.1      Education.2      Education.3       Mortgage
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :  0.00
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:  0.00
##  Median :0.0000   Median :0.0000   Median :0.000   Median :  0.00
##  Mean   :0.4205   Mean   :0.2865   Mean   :0.293   Mean   : 56.23
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:100.00
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :635.00
##  Personal.Loan   Securities.Account   CD.Account        Online
##  Min.   :0.000   Min.   :0.000      Min.   :0.000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:0.000      1st Qu.:0.000   1st Qu.:0.000
##  Median :0.000   Median :0.000      Median :0.000   Median :1.000
##  Mean   :0.094   Mean   :0.097      Mean   :0.057   Mean   :0.586
##  3rd Qu.:0.000   3rd Qu.:0.000      3rd Qu.:0.000   3rd Qu.:1.000
##  Max.   :1.000   Max.   :1.000      Max.   :1.000   Max.   :1.000
##    CreditCard
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2815
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```r
###Normalizing
train_normalization <-train.data
valid_normalization <-valid.data
normalization.values <- preProcess(train.data[, c("Age","Experience","Income","CCAvg","Mortgage"
)], method=c("center", "scale"))
train_normalization[, c("Age","Experience","Income","CCAvg","Mortgage")] <- predict(normalizatio
n.values, train.data[, c("Age","Experience","Income","CCAvg","Mortgage")])
valid_normalization[, c("Age","Experience","Income","CCAvg","Mortgage")] <- predict(normalizatio
n.values, valid.data[, c("Age","Experience","Income","CCAvg","Mortgage")])
summary(train_normalization)
```

```
##       Age             Experience            Income           Family.1
##  Min.   :-1.91217   Min.   :-1.979430   Min.   :-1.4449   Min.   :0.0000
##  1st Qu.:-0.88170   1st Qu.:-0.861934   1st Qu.:-0.7689   1st Qu.:0.0000
##  Median :-0.02298   Median :-0.002321   Median :-0.2238   Median :0.0000
##  Mean   : 0.00000   Mean   : 0.000000   Mean   : 0.0000   Mean   :0.2937
##  3rd Qu.: 0.92161   3rd Qu.: 0.857292   3rd Qu.: 0.5829   3rd Qu.:1.0000
##  Max.   : 1.86620   Max.   : 1.974789   Max.   : 3.2647   Max.   :1.0000
##     Family.2        Family.3        Family.4          CCAvg
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000   Min.   :-1.1183
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:-0.7196
##  Median :0.000   Median :0.000   Median :0.0000   Median :-0.2069
##  Mean   :0.259   Mean   :0.197   Mean   :0.2503   Mean   : 0.0000
##  3rd Qu.:1.000   3rd Qu.:0.000   3rd Qu.:1.0000   3rd Qu.: 0.3628
##  Max.   :1.000   Max.   :1.000   Max.   :1.0000   Max.   : 4.5784
##   Education.1      Education.2      Education.3       Mortgage
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :-0.5568
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:-0.5568
##  Median :0.0000   Median :0.0000   Median :0.000   Median :-0.5568
##  Mean   :0.4183   Mean   :0.2767   Mean   :0.305   Mean   : 0.0000
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.: 0.4453
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   : 5.3481
##  Personal.Loan     Securities.Account   CD.Account          Online
##  Min.   :0.00000   Min.   :0.0000     Min.   :0.00000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.0000     1st Qu.:0.00000   1st Qu.:0.000
##  Median :0.00000   Median :0.0000     Median :0.00000   Median :1.000
##  Mean   :0.09733   Mean   :0.1093     Mean   :0.06267   Mean   :0.604
##  3rd Qu.:0.00000   3rd Qu.:0.0000     3rd Qu.:0.00000   3rd Qu.:1.000
##  Max.   :1.00000   Max.   :1.0000     Max.   :1.00000   Max.   :1.000
##    CreditCard
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.3023
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```r
###Q1 - Age = 40, Experience = 10, Income = 84, Family = 2, CCAvg = 2, Education_1 = 0, Educatio
n_2 = 1, Education_3 = 0, Mortgage = 0, Securities Account = 0, CD Account = 0, Online = 1, and
Credit Card = 1. Perform a k-NN classification with all predictors except ID and ZIP code using
k = 1. Remember to transform categorical predictors with more than two categories into dummy var
iables first. Specify the success class as 1 (loan acceptance), and use the default cutoff value
of 0.5. How would this customer be classified?

df2 <- data.frame(Age = 40, Experience = 10, Income = 84, CCAvg = 2,
                  Mortgage = 0, Securities.Account = 0, CD.Account = 0,
                  Online = 1, CreditCard =1, Family_1 = 0, Family_2 = 1,
                  Family_3 = 0, Family_4 = 0, Education_1 = 0,
                  Education_2 = 1, Education_3 = 0)
df3 <- predict(normalization.values, df2)
print(df3)
```

```
##          Age Experience    Income      CCAvg   Mortgage Securities.Account
## 1 -0.4523446 -0.8619338 0.2122216 0.02100374 -0.5568407                  0
##   CD.Account Online CreditCard Family_1 Family_2 Family_3 Family_4 Education_1
## 1          0      1          1        0        1        0        0           0
##   Education_2 Education_3
## 1           1           0
```

```r
###Predictors

predictors <- c(1:5,7:17)
nn <- knn(train=train_normalization[, predictors], test=df3, cl=train_normalization[, 6], k=1)
row.names(train_normalization)[attr(nn, "nn.index")]
```

```
## [1] "2736"
```

```r
print(train_normalization[attr(nn, "nn.index"),])
```

```
##             Age Experience      Income Family.1 Family.2 Family.3 Family.4
## 2736 -0.7958322 -0.6900112 -0.09302866        0        0        1        0
##          CCAvg Education.1 Education.2 Education.3 Mortgage Personal.Loan
## 2736 0.3628059           0           1           0  1.06432             0
##      Securities.Account CD.Account Online CreditCard
## 2736                  0          0      1          0
```

```r
nn[1]
```

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

```r
### Customer will accept the loan.
```

```r
###Q2 - What is a choice of k that balances between overfitting and ignoring the predictor infor
mation?

df4 <-data.frame(k=seq (1,14,1), accuracy= rep(0,14), sensitivity = rep(0,14), specificity = rep
(0,14), PPV=rep(0,14), NPV=rep(0,14), F1=rep(0,14))
for(i in 1:14)
{
  knn1 <- knn(train_normalization[,predictors], valid_normalization[,predictors], cl = as.factor
(train_normalization[, 6]), k = i)
  knn2 <- confusionMatrix(knn1, as.factor(valid_normalization[, 6]), positive="1")
  df4[i, 2] <- knn2$overall[1] #Accuracy
  df4[i, 3] <- knn2$byClass[c("Sensitivity")]
  df4[i, 4] <- knn2$byClass[c("Specificity")]
  df4[i, 5] <- knn2$byClass[c("Precision")]
  df4[i, 6] <- knn2$byClass[c("Positive Predected Value")]
  df4[i, 7] <- knn2$byClass[c("Negative Predected Value")]
  df4[i, 8] <- knn2$byClass[c("F")]
}
df4
```

k	accuracy	sensitivity	specificity	PPV	NPV	F1	V8
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.9635	0.9212411	0.9746996	0.9061033	NA	NA	NA
2	0.9570	0.8162291	0.9943074	0.9743590	NA	NA	NA
3	0.9675	0.8926014	0.9873498	0.9492386	NA	NA	NA
4	0.9575	0.8114558	0.9962049	0.9826590	NA	NA	NA
5	0.9670	0.8782816	0.9905123	0.9608355	NA	NA	NA
6	0.9560	0.8066826	0.9955724	0.9797101	NA	NA	NA
7	0.9595	0.8424821	0.9905123	0.9592391	NA	NA	NA
8	0.9480	0.7684964	0.9955724	0.9787234	NA	NA	NA
9	0.9555	0.8162291	0.9924099	0.9661017	NA	NA	NA
10	0.9440	0.7446301	0.9968374	0.9842271	NA	NA	NA

1-10 of 14 rows   Previous 1 2 Next

```r
### as the result show, we have the highest sensitivity from K=3

knn1<-knn(train_normalization[, predictors], valid_normalization [, predictors], cl=as.factor(tr
ain_normalization[,6]),k=3)
knn4<- CrossTable(x=valid_normalization[,6], y=knn1, prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  2000
##
##
##                          | knn1
## valid_normalization[, 6] |         0 |         1 | Row Total |
## -------------------------|-----------|-----------|-----------|
##                        0 |      1561 |        20 |      1581 |
##                          |     0.987 |     0.013 |     0.790 |
##                          |     0.972 |     0.051 |           |
##                          |     0.780 |     0.010 |           |
## -------------------------|-----------|-----------|-----------|
##                        1 |        45 |       374 |       419 |
##                          |     0.107 |     0.893 |     0.209 |
##                          |     0.028 |     0.949 |           |
##                          |     0.022 |     0.187 |           |
## -------------------------|-----------|-----------|-----------|
##             Column Total |      1606 |       394 |      2000 |
##                          |     0.803 |     0.197 |           |
## -------------------------|-----------|-----------|-----------|
##
##
```

```r
### Q4- Consider the following customer: Age = 40, Experience = 10, Income = 84,Family = 2, CCAv
g = 2, Education_1 = 0, Education_2 = 1, Education_3 = 0, Mortgage = 0, Securities Account = 0,
CD Account = 0, Online = 1 and Credit Card = 1. Classify the customer using the best k

knn5 <- knn(train=train_normalization[, predictors], test=df3, cl=train_normalization[, 6], k=3,
prob=TRUE)
print(train_normalization[attr(knn5, "knn5.index"),])
```

```
##  [1] Age                Experience         Income             Family.1
##  [5] Family.2           Family.3           Family.4           CCAvg
##  [9] Education.1        Education.2        Education.3        Mortgage
## [13] Personal.Loan      Securities.Account CD.Account         Online
## [17] CreditCard
## <0 rows> (or 0-length row.names)
```

```r
knn5[1]
```

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

```r
###using 1 because 0 will mean not accept loan
```

```r
### Q5- Repartition the data, this time into training, validation, and test sets (50% : 30% : 20
%). Apply the k-NN method with the k chosen above. Compare the confusion matrix of the test set
with that of the training and validation sets. Comment on the differences and their reason.

set.seed(5)
test.index = createDataPartition(df1$Personal.Loan,p=0.2, list=FALSE)
test.data = df1[test.index,]
trainvalid.data = df1[-test.index,]
train.index = createDataPartition(trainvalid.data$Personal.Loan,p=0.625, list=FALSE)
train.data = trainvalid.data[train.index,]
valid.data = trainvalid.data[-train.index,]
Train.D <- colMeans(train.data)
Valid.D <- colMeans(valid.data)
sprintf("%.3f", Train.D)
```

```
##  [1] "45.521" "20.304" "74.028" "0.288"  "0.251"  "0.196"  "0.266"  "1.921"
##  [9] "0.420"  "0.283"  "0.296"  "57.450" "0.093"  "0.105"  "0.060"  "0.599"
## [17] "0.297"
```

```r
sprintf("%.3f", Valid.D)
```

```
##  [1] "45.182" "19.907" "73.820" "0.297"  "0.271"  "0.216"  "0.217"  "1.936"
##  [9] "0.434"  "0.256"  "0.310"  "52.753" "0.097"  "0.101"  "0.067"  "0.586"
## [17] "0.299"
```

```r
rm(Train.D)
rm(Valid.D)
```

```r
###Normalizing

train_normalization <- train.data
valid_normalization <- valid.data
normalization_values <- preProcess(train.data[, c("Age","Experience","Income","CCAvg","Mortgage"
)], method=c("center", "scale"))
train_normalization[, c("Age","Experience","Income","CCAvg","Mortgage")] <- predict(normalizatio
n_values, train.data[, c("Age","Experience","Income","CCAvg","Mortgage")])
valid_normalization[, c("Age","Experience","Income","CCAvg","Mortgage")] <- predict(normalizatio
n_values, valid.data[, c("Age","Experience","Income","CCAvg","Mortgage")])
summary(train_normalization)
```

```
##       Age            Experience           Income            Family.1
##  Min.   :-1.96834   Min.   :-2.03822   Min.   :-1.4421   Min.   :0.000
##  1st Qu.:-0.91955   1st Qu.:-0.81377   1st Qu.:-0.7650   1st Qu.:0.000
##  Median : 0.04185   Median : 0.06084   Median :-0.2190   Median :0.000
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000   Mean   :0.288
##  3rd Qu.: 0.82844   3rd Qu.: 0.84798   3rd Qu.: 0.5454   3rd Qu.:1.000
##  Max.   : 1.87723   Max.   : 1.98497   Max.   : 2.8604   Max.   :1.000
##     Family.2         Family.3         Family.4          CCAvg
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :-1.1183
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:-0.7109
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :-0.2162
##  Mean   :0.2508   Mean   :0.1956   Mean   :0.2656   Mean   : 0.0000
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.: 0.3367
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   : 4.7020
##   Education.1      Education.2      Education.3        Mortgage
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :-0.5582
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:-0.5582
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :-0.5582
##  Mean   :0.4204   Mean   :0.2832   Mean   :0.2964   Mean   : 0.0000
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.: 0.4523
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   : 5.6117
##  Personal.Loan    Securities.Account   CD.Account       Online
##  Min.   :0.0000   Min.   :0.0000     Min.   :0.00   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000     1st Qu.:0.00   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000     Median :0.00   Median :1.0000
##  Mean   :0.0928   Mean   :0.1048     Mean   :0.06   Mean   :0.5992
##  3rd Qu.:0.0000   3rd Qu.:0.0000     3rd Qu.:0.00   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000     Max.   :1.00   Max.   :1.0000
##    CreditCard
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2972
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```r
summary(valid_normalization)
```

```
##       Age            Experience            Income             Family.1
##  Min.   :-1.96834   Min.   :-2.03822   Min.   :-1.442057   Min.   :0.0000
##  1st Qu.:-0.91955   1st Qu.:-0.90123   1st Qu.:-0.786855   1st Qu.:0.0000
##  Median :-0.04555   Median :-0.02662   Median :-0.240852   Median :0.0000
##  Mean   :-0.02965   Mean   :-0.03479   Mean   :-0.004543   Mean   :0.2967
##  3rd Qu.: 0.82844   3rd Qu.: 0.84798   3rd Qu.: 0.529011   3rd Qu.:1.0000
##  Max.   : 1.87723   Max.   : 1.89751   Max.   : 3.144360   Max.   :1.0000
##     Family.2         Family.3        Family.4           CCAvg
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :-1.118332
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:-0.728372
##  Median :0.0000   Median :0.000   Median :0.0000   Median :-0.245286
##  Mean   :0.2707   Mean   :0.216   Mean   :0.2167   Mean   : 0.008522
##  3rd Qu.:1.0000   3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.: 0.336744
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   : 4.701974
##   Education.1    Education.2    Education.3      Mortgage
##  Min.   :0.000   Min.   :0.000   Min.   :0.00   Min.   :-0.55821
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.00   1st Qu.:-0.55821
##  Median :0.000   Median :0.000   Median :0.00   Median :-0.55821
##  Mean   :0.434   Mean   :0.256   Mean   :0.31   Mean   :-0.04564
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.00   3rd Qu.: 0.34785
##  Max.   :1.000   Max.   :1.000   Max.   :1.00   Max.   : 5.28136
##  Personal.Loan     Securities.Account   CD.Account          Online
##  Min.   :0.00000   Min.   :0.0000     Min.   :0.00000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.0000     1st Qu.:0.00000   1st Qu.:0.000
##  Median :0.00000   Median :0.0000     Median :0.00000   Median :1.000
##  Mean   :0.09667   Mean   :0.1013     Mean   :0.06667   Mean   :0.586
##  3rd Qu.:0.00000   3rd Qu.:0.0000     3rd Qu.:0.00000   3rd Qu.:1.000
##  Max.   :1.00000   Max.   :1.0000     Max.   :1.00000   Max.   :1.000
##    CreditCard
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2987
##  3rd Qu.:1.0000
##  Max.   :1.0000
```