Assignment 2

Alex Haffner

2/19/2022

```
## R Markdown

UniversalBank <- read.csv("~/Downloads/UniversalBank.csv")
summary(UniversalBank)
```

```
##        ID             Age          Experience       Income      
ZIP.Code    
##  Min.   :   1   Min.   :23.00   Min.   :-3.0   Min.   :  8.00   Min.   :
9307  
##  1st Qu.:1251   1st Qu.:35.00   1st Qu.:10.0   1st Qu.: 39.00   1st
Qu.:91911  
##  Median :2500   Median :45.00   Median :20.0   Median : 64.00   Median
:93437  
##  Mean   :2500   Mean   :45.34   Mean   :20.1   Mean   : 73.77   Mean
:93152  
##  3rd Qu.:3750   3rd Qu.:55.00   3rd Qu.:30.0   3rd Qu.: 98.00   3rd
Qu.:94608  
##  Max.   :5000   Max.   :67.00   Max.   :43.0   Max.   :224.00   Max.
:96651  
##      Family          CCAvg          Education        Mortgage    
##  Min.   :1.000   Min.   : 0.000   Min.   :1.000   Min.   :  0.0  
##  1st Qu.:1.000   1st Qu.: 0.700   1st Qu.:1.000   1st Qu.:  0.0  
##  Median :2.000   Median : 1.500   Median :2.000   Median :  0.0  
##  Mean   :2.396   Mean   : 1.938   Mean   :1.881   Mean   : 56.5  
##  3rd Qu.:3.000   3rd Qu.: 2.500   3rd Qu.:3.000   3rd Qu.:101.0  
##  Max.   :4.000   Max.   :10.000   Max.   :3.000   Max.   :635.0  
##  Personal.Loan   Securities.Account   CD.Account         Online      
##  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 :0.0000     Median :0.0000   Median :1.0000  
##  Mean   :0.096   Mean   :0.1044     Mean   :0.0604   Mean   :0.5968  
##  3rd Qu.:0.000   3rd Qu.:0.0000     3rd Qu.:0.0000   3rd Qu.:1.0000  
##  Max.   :1.000   Max.   :1.0000     Max.   :1.0000   Max.   :1.0000  
##    CreditCard   
##  Min.   :0.000  
##  1st Qu.:0.000  
##  Median :0.000  
##  Mean   :0.294  
##  3rd Qu.:1.000  
##  Max.   :1.000  
```

```
#Null Variables
UniversalBank$ID<-NULL
UniversalBank$ZIP.Code<-NULL


summary(UniversalBank)
```

```
##       Age          Experience       Income           Family
##  Min.   :23.00   Min.   :-3.0   Min.   :  8.00   Min.   :1.000
##  1st Qu.:35.00   1st Qu.:10.0   1st Qu.: 39.00   1st Qu.:1.000
##  Median :45.00   Median :20.0   Median : 64.00   Median :2.000
##  Mean   :45.34   Mean   :20.1   Mean   : 73.77   Mean   :2.396
##  3rd Qu.:55.00   3rd Qu.:30.0   3rd Qu.: 98.00   3rd Qu.:3.000
##  Max.   :67.00   Max.   :43.0   Max.   :224.00   Max.   :4.000
##      CCAvg          Education        Mortgage     Personal.Loan
##  Min.   : 0.000   Min.   :1.000   Min.   :  0.0   Min.   :0.000
##  1st Qu.: 0.700   1st Qu.:1.000   1st Qu.:  0.0   1st Qu.:0.000
##  Median : 1.500   Median :2.000   Median :  0.0   Median :0.000
##  Mean   : 1.938   Mean   :1.881   Mean   : 56.5   Mean   :0.096
##  3rd Qu.: 2.500   3rd Qu.:3.000   3rd Qu.:101.0   3rd Qu.:0.000
##  Max.   :10.000   Max.   :3.000   Max.   :635.0   Max.   :1.000
##  Securities.Account   CD.Account         Online         CreditCard
##  Min.   :0.0000     Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000     1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000     Median :0.0000   Median :1.0000   Median :0.000
##  Mean   :0.1044     Mean   :0.0604   Mean   :0.5968   Mean   :0.294
##  3rd Qu.:0.0000     3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000     Max.   :1.0000   Max.   :1.0000   Max.   :1.000
```

```
#CaretLibrary
library(caret)
```

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

```
## Warning in register(): Can't find generic `scale_type` in package ggplot2
to
## register S3 method.
```

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

```
library(class)
summary(UniversalBank)
```

```
##       Age          Experience       Income           Family
##  Min.   :23.00   Min.   :-3.0   Min.   :  8.00   Min.   :1.000
##  1st Qu.:35.00   1st Qu.:10.0   1st Qu.: 39.00   1st Qu.:1.000
##  Median :45.00   Median :20.0   Median : 64.00   Median :2.000
##  Mean   :45.34   Mean   :20.1   Mean   : 73.77   Mean   :2.396
##  3rd Qu.:55.00   3rd Qu.:30.0   3rd Qu.: 98.00   3rd Qu.:3.000
##  Max.   :67.00   Max.   :43.0   Max.   :224.00   Max.   :4.000
##      CCAvg          Education        Mortgage     Personal.Loan
##  Min.   : 0.000   Min.   :1.000   Min.   :  0.0   Min.   :0.000
##  1st Qu.: 0.700   1st Qu.:1.000   1st Qu.:  0.0   1st Qu.:0.000
##  Median : 1.500   Median :2.000   Median :  0.0   Median :0.000
##  Mean   : 1.938   Mean   :1.881   Mean   : 56.5   Mean   :0.096
##  3rd Qu.: 2.500   3rd Qu.:3.000   3rd Qu.:101.0   3rd Qu.:0.000
##  Max.   :10.000   Max.   :3.000   Max.   :635.0   Max.   :1.000
##  Securities.Account   CD.Account         Online          CreditCard
##  Min.   :0.0000     Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000     1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000     Median :0.0000   Median :1.0000   Median :0.000
##  Mean   :0.1044     Mean   :0.0604   Mean   :0.5968   Mean   :0.294
##  3rd Qu.:0.0000     3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000     Max.   :1.0000   Max.   :1.0000   Max.   :1.000
```

```
#DummyVariable
UniversalBank$Personal.Loan=as.factor(UniversalBank$Personal.Loan)


summary(UniversalBank)
```

```
##       Age          Experience       Income           Family     
##  Min.   :23.00   Min.   :-3.0   Min.   :  8.00   Min.   :1.000
##  1st Qu.:35.00   1st Qu.:10.0   1st Qu.: 39.00   1st Qu.:1.000
##  Median :45.00   Median :20.0   Median : 64.00   Median :2.000
##  Mean   :45.34   Mean   :20.1   Mean   : 73.77   Mean   :2.396
##  3rd Qu.:55.00   3rd Qu.:30.0   3rd Qu.: 98.00   3rd Qu.:3.000
##  Max.   :67.00   Max.   :43.0   Max.   :224.00   Max.   :4.000
##      CCAvg          Education        Mortgage     Personal.Loan
##  Min.   : 0.000   Min.   :1.000   Min.   :  0.0   0:4520
##  1st Qu.: 0.700   1st Qu.:1.000   1st Qu.:  0.0   1: 480
##  Median : 1.500   Median :2.000   Median :  0.0
##  Mean   : 1.938   Mean   :1.881   Mean   : 56.5
##  3rd Qu.: 2.500   3rd Qu.:3.000   3rd Qu.:101.0
##  Max.   :10.000   Max.   :3.000   Max.   :635.0
##  Securities.Account   CD.Account         Online          CreditCard
##  Min.   :0.0000     Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000     1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000     Median :0.0000   Median :1.0000   Median :0.000
##  Mean   :0.1044     Mean   :0.0604   Mean   :0.5968   Mean   :0.294
##  3rd Qu.:0.0000     3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000     Max.   :1.0000   Max.   :1.0000   Max.   :1.000
```

```
Bank_Norm<-UniversalBank


#Normalize

Norm_model<-preProcess(UniversalBank[,-8],method = c("center","scale"))
Bank_Norm[,-8]=predict(Norm_model,UniversalBank[,-8])
summary(Bank_Norm)
```

```
##       Age             Experience            Income            Family      
##  Min.   :-1.94871   Min.   :-2.014710   Min.   :-1.4288   Min.   :-1.2167  
##  1st Qu.:-0.90188   1st Qu.:-0.881116   1st Qu.:-0.7554   1st Qu.:-1.2167  
##  Median :-0.02952   Median :-0.009121   Median :-0.2123   Median :-0.3454  
##  Mean   : 0.00000   Mean   : 0.000000   Mean   : 0.0000   Mean   : 0.0000  
##  3rd Qu.: 0.84284   3rd Qu.: 0.862874   3rd Qu.: 0.5263   3rd Qu.: 0.5259  
##  Max.   : 1.88967   Max.   : 1.996468   Max.   : 3.2634   Max.   : 1.3973  
##      CCAvg           Education          Mortgage       Personal.Loan
##  Min.   :-1.1089   Min.   :-1.0490   Min.   :-0.5555   0:4520       
##  1st Qu.:-0.7083   1st Qu.:-1.0490   1st Qu.:-0.5555   1: 480       
##  Median :-0.2506   Median : 0.1417   Median :-0.5555                
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000                
##  3rd Qu.: 0.3216   3rd Qu.: 1.3324   3rd Qu.: 0.4375                
##  Max.   : 4.6131   Max.   : 1.3324   Max.   : 5.6875                
##  Securities.Account   CD.Account          Online          CreditCard     
##  Min.   :-0.3414    Min.   :-0.2535   Min.   :-1.2165   Min.   :-0.6452  
##  1st Qu.:-0.3414    1st Qu.:-0.2535   1st Qu.:-1.2165   1st Qu.:-0.6452  
##  Median :-0.3414    Median :-0.2535   Median : 0.8219   Median :-0.6452  
##  Mean   : 0.0000    Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000  
##  3rd Qu.:-0.3414    3rd Qu.:-0.2535   3rd Qu.: 0.8219   3rd Qu.: 1.5495  
##  Max.   : 2.9286    Max.   : 3.9438   Max.   : 0.8219   Max.   : 1.5495  

#Train
Train_Index=createDataPartition(UniversalBank$Personal.Loan,p=0.6,list =
FALSE)
Train.df=Bank_Norm[Train_Index,]
Validation.df=Bank_Norm[-Train_Index,]
```

#Predict

```
To_Predict=data.frame(Age=40, Experience=10, Income=84, Family=2,
                      CCAvg=2, Education=1,
                      Mortgage=0, Securities.Account=0,
                      CD.Account=0, Online=1, CreditCard=1)
print(To_Predict)

##   Age Experience Income Family CCAvg Education Mortgage Securities.Account
## 1  40         10     84      2     2         1        0                  0
##   CD.Account Online CreditCard
## 1          0      1          1

To_Predict_norm<-predict(Norm_model,To_Predict)
```

#Prediction

```
Prediction <-knn(train=Train.df[,1:7],
                 test=To_Predict_norm[,1:7],
                 cl=Train.df$Personal.Loan,
                 k=1)

print(Prediction)
```

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

#Task2

```
set.seed(123)
fitControl<-trainControl(method="repeatedcv",number = 3,repeats = 2)
searchGrid=expand.grid(k=1:10)

Knn.model=train(Personal.Loan~.,
                data = Train.df,
                method = 'knn',
                tuneGrid = searchGrid,
                trControl = fitControl)
Knn.model
```

```
## k-Nearest Neighbors
##
## 3000 samples
##   11 predictor
##    2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (3 fold, repeated 2 times)
## Summary of sample sizes: 2000, 2000, 2000, 2000, 2000, 2000, ...
## Resampling results across tuning parameters:
##
##   k   Accuracy   Kappa
##    1  0.9548333  0.7119388
##    2  0.9496667  0.6789197
##    3  0.9556667  0.7018490
##    4  0.9526667  0.6751525
##    5  0.9538333  0.6786610
##    6  0.9540000  0.6799036
##    7  0.9513333  0.6543835
##    8  0.9491667  0.6360260
##    9  0.9480000  0.6230283
##   10  0.9465000  0.6079954
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 3.
```

```
predictions<-predict(Knn.model,Validation.df)

confusionMatrix(predictions,Validation.df$Personal.Loan)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1804   76
##          1    4  116
##                                           
##                Accuracy : 0.96            
##                  95% CI : (0.9505, 0.9682)
##     No Information Rate : 0.904           
##     P-Value [Acc > NIR] : < 2.2e-16       
##                                           
##                   Kappa : 0.7231          
##                                           
##  Mcnemar's Test P-Value : 2.054e-15       
##                                           
##             Sensitivity : 0.9978          
##             Specificity : 0.6042          
##          Pos Pred Value : 0.9596          
##          Neg Pred Value : 0.9667          
##              Prevalence : 0.9040          
##          Detection Rate : 0.9020          
##    Detection Prevalence : 0.9400          
##       Balanced Accuracy : 0.8010          
##                                           
##        'Positive' Class : 0               
##
```

```
To_Predict_norm=data.frame(Age=40, Experience=10, Income=84, Family=2,
                           CCavg=2, Education=1, Mortgage=0,
Securities.Account=0,
                           CD.Account=0, Online=1, CreditCard=1)

To_Predict_norm=predict(Norm_model,To_Predict)
predict(Knn.model,To_Predict_norm)
```

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

#Task 5

```
train_size=.5
Train_Index=createDataPartition(UniversalBank$Personal.Loan,p=0.5,list =
FALSE)
Train.df=Bank_Norm[Train_Index,]
valid_size=.30
Validation.df=Bank_Norm[-Train_Index,]

To_Predict=data.frame(Age=40, Experience=10, Income=84, Family=2,
                      CCAvg=2, Education=1,
                      Mortgage=0, Securities.Account=0,
                      CD.Account=0, Online=1, CreditCard=1)
print(To_Predict)
```

```
##   Age Experience Income Family CCAvg Education Mortgage Securities.Account
## 1  40         10     84      2     2         1        0                  0
##   CD.Account Online CreditCard
## 1          0      1          1
```

```
To_Predict_norm<-predict(Norm_model,To_Predict)

#Prediction
Prediction <-knn(train=Train.df[,1:7],
                 test=To_Predict_norm[,1:7],
                 cl=Train.df$Personal.Loan,
                 k=1)

print(Prediction)
```

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

```
set.seed(123)
fitControl<-trainControl(method="repeatedcv",number = 3,repeats = 2)
searchGrid=expand.grid(k=1:10)

Knn.model=train(Personal.Loan~.,
                data = Train.df,
                method = 'knn',
                tuneGrid = searchGrid,
                trControl = fitControl)
Knn.model
```

```
## k-Nearest Neighbors
##
## 2500 samples
##   11 predictor
##    2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (3 fold, repeated 2 times)
## Summary of sample sizes: 1667, 1667, 1666, 1667, 1666, 1667, ...
## Resampling results across tuning parameters:
##
##   k   Accuracy   Kappa
##    1  0.9544007  0.7026621
##    2  0.9455996  0.6436537
##    3  0.9530009  0.6745506
##    4  0.9514005  0.6552616
##    5  0.9514000  0.6501013
##    6  0.9478000  0.6214736
##    7  0.9443991  0.5838148
##    8  0.9443998  0.5823568
##    9  0.9406005  0.5430711
##   10  0.9404009  0.5414687
##
```

```
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 1.
```

```
predictions<-predict(Knn.model,Validation.df)
```

```
confusionMatrix(predictions,Validation.df$Personal.Loan)
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 2233   87
##          1   27  153
## 
##                Accuracy : 0.9544
##                  95% CI : (0.9455, 0.9622)
##     No Information Rate : 0.904
##     P-Value [Acc > NIR] : < 2.2e-16
## 
##                   Kappa : 0.7042
## 
##  Mcnemar's Test P-Value : 3.279e-08
## 
##             Sensitivity : 0.9881
##             Specificity : 0.6375
##          Pos Pred Value : 0.9625
##          Neg Pred Value : 0.8500
##              Prevalence : 0.9040
##          Detection Rate : 0.8932
##    Detection Prevalence : 0.9280
##       Balanced Accuracy : 0.8128
## 
##        'Positive' Class : 0
## 
```

#The Difference that I noticed was the increase in accuracy from 94.6% in the first set and up to 95% when calculating the conusion matrix. #This is because training and validating in smaller groups improves models and allows them to perform more accurately.