Load needed packages and dataset

```
library('caret')
```

```
## Warning: package 'caret' was built under R version 4.1.2
```

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

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

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

```
library('ISLR')
```

```
## Warning: package 'ISLR' was built under R version 4.1.2
```

```
library('dplyr')
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
##
## 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('class')
UniversalBank <- read.csv('C:/Users/lmszr/Documents/School/Fundamentals of Machine Learning/UniversalBa
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   :93153
##   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
```

Q1

Choose only necessary columns

```
UniversalBank_s <- select(UniversalBank, Age, Experience, Income, Family, CCAvg, Education, Mortgage, Se
summary(UniversalBank_s)
```

```
##       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       Securities.Account
##  Min.   : 0.000   Min.   :1.000   Min.   :  0.0    Min.   :0.0000
##  1st Qu.: 0.700   1st Qu.:1.000   1st Qu.:  0.0    1st Qu.:0.0000
##  Median : 1.500   Median :2.000   Median :  0.0    Median :0.0000
##  Mean   : 1.938   Mean   :1.881   Mean   : 56.5    Mean   :0.1044
##  3rd Qu.: 2.500   3rd Qu.:3.000   3rd Qu.:101.0    3rd Qu.:0.0000
##  Max.   :10.000   Max.   :3.000   Max.   :635.0    Max.   :1.0000
##    CD.Account         Online         CreditCard
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000   Median :1.0000   Median :0.000
##  Mean   :0.0604   Mean   :0.5968   Mean   :0.294
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000
```

Normalize the data

```
Norm_model <- preProcess(UniversalBank_s,
                         method = c("center", "scale"))
UniversalBank_norm=predict(Norm_model,UniversalBank_s)
summary(UniversalBank_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        Securities.Account
## Min.   :-1.1089   Min.   :-1.0490   Min.   :-0.5555   Min.   :-0.3414
## 1st Qu.:-0.7083   1st Qu.:-1.0490   1st Qu.:-0.5555   1st Qu.:-0.3414
## Median :-0.2506   Median : 0.1417   Median :-0.5555   Median :-0.3414
## Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
## 3rd Qu.: 0.3216   3rd Qu.: 1.3324   3rd Qu.: 0.4375   3rd Qu.:-0.3414
## Max.   : 4.6131   Max.   : 1.3324   Max.   : 5.6875   Max.   : 2.9286
##    CD.Account          Online            CreditCard
## Min.   :-0.2535   Min.   :-1.2165   Min.   :-0.6452
## 1st Qu.:-0.2535   1st Qu.:-1.2165   1st Qu.:-0.6452
## Median :-0.2535   Median : 0.8219   Median :-0.6452
## Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
## 3rd Qu.:-0.2535   3rd Qu.: 0.8219   3rd Qu.: 1.5495
## Max.   : 3.9438   Max.   : 0.8219   Max.   : 1.5495
```

Add back in target Personal.Loan

```
UniversalBank$Personal.Loan <- factor(UniversalBank$Personal.Loan)
UniversalBank_norm$Personal.Loan <- UniversalBank$Personal.Loan
summary(UniversalBank_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        Securities.Account
## Min.   :-1.1089   Min.   :-1.0490   Min.   :-0.5555   Min.   :-0.3414
## 1st Qu.:-0.7083   1st Qu.:-1.0490   1st Qu.:-0.5555   1st Qu.:-0.3414
## Median :-0.2506   Median : 0.1417   Median :-0.5555   Median :-0.3414
## Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
## 3rd Qu.: 0.3216   3rd Qu.: 1.3324   3rd Qu.: 0.4375   3rd Qu.:-0.3414
## Max.   : 4.6131   Max.   : 1.3324   Max.   : 5.6875   Max.   : 2.9286
##    CD.Account          Online            CreditCard      Personal.Loan
## Min.   :-0.2535   Min.   :-1.2165   Min.   :-0.6452   0:4520
## 1st Qu.:-0.2535   1st Qu.:-1.2165   1st Qu.:-0.6452   1: 480
## Median :-0.2535   Median : 0.8219   Median :-0.6452
## Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
## 3rd Qu.:-0.2535   3rd Qu.: 0.8219   3rd Qu.: 1.5495
## Max.   : 3.9438   Max.   : 0.8219   Max.   : 1.5495
```

Partition data into test and training sets

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

Prediction data

```
To_Predict=data.frame(Age = 40, Experience = 10, Income = 84, Family = 2, CCAvg = 2, Education = 2, Mor
print(To_Predict)
```

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

Predict outcome

```
To_Predict_norm = predict(Norm_model, To_Predict)
print(To_Predict_norm)
```

```
##          Age Experience    Income     Family     CCAvg Education   Mortgage
## 1 -0.4657003 -0.8811162 0.2221371 -0.3453975 0.0355115 0.1416887 -0.5554684
##   Securities.Account CD.Account    Online CreditCard
## 1         -0.3413892 -0.2535149 0.8218687   1.549477
```

```
Prediction <-knn(train=Train.df[,1:11],
                 test=To_Predict_norm[,1:11],
                 cl=Train.df$Personal.Loan,
                 k=1)
print(Prediction)
```

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

Q2

Calculate best K

```
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  0.9560000  0.7203133
##    2  0.9515000  0.6925759
##    3  0.9560000  0.7087606
##    4  0.9523333  0.6811384
##    5  0.9523333  0.6750470
##    6  0.9510000  0.6671970
##    7  0.9481667  0.6388651
##    8  0.9486667  0.6421808
##    9  0.9470000  0.6248052
##   10  0.9470000  0.6204474
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 3.
```

Q3

Confusion matrix for best K

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

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

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1796   83
##          1   12  109
##
##              Accuracy : 0.9525
##                95% CI : (0.9422, 0.9614)
##   No Information Rate : 0.904
##   P-Value [Acc > NIR] : 4.861e-16
##
##                 Kappa : 0.6722
##
## Mcnemar's Test P-Value : 6.878e-13
##
##           Sensitivity : 0.9934
##           Specificity : 0.5677
##        Pos Pred Value : 0.9558
##        Neg Pred Value : 0.9008
##            Prevalence : 0.9040
##        Detection Rate : 0.8980
```

```
##     Detection Prevalence : 0.9395
##        Balanced Accuracy : 0.7805
##
##          'Positive' Class : 0
##
```

Q4

Predict

```
predict(Knn.model,To_Predict_norm)
```

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