Agenda:

- ✓ Get the data & pre-process it
- ✓ Apply k-NN model for classification and regression
- ✓ Understand why kNN is an instance-based learning technique
- ✓ Understand the effect of standardization of data on kNN predictions
- ✓ Condensing the data & comparing the predictions
- ✓ Applying FNN model
- ✓ Building Recommendation Model

#k-NN Classification:

- ➢ The goal is to predict prospective personal-loan takers for a bank.

  - The credit and general information of a general customer base is used to predict whether a given new customer can be given a personal loan

  - This enables the bank to mitigate the risks and maximize the profits, by targeting high prospective customers and at the same by being conservative to non-potential customers, while giving personal loans

Steps:

1. Clear the workspace and set the Working directory to the input files.
2. Read the given ”UniversalBank.csv”, and check the structure, dimensions of data.
3. Drop the duplicate records if you have any. (optional)
4. Check for missing values and impute them.
5. Drop the attributes which are not necessary
6. Convert the categorical variables to factor if you have any.(Hint: Education)
7. Dummify the factor variable.
8. Split the data in to train and test. (check the ratio of target variable for both the classes).
9. Get the train and test sets excluding the target variable

   ```
   bankdata_trainwithoutclass = subset(bankdata_train,select=-c(Personal.Loan))
   bankdata_testwithoutclass = subset(bankdata_test,select=-c(Personal.Loan))
   ```

10. Applying k-NN Algorithm by taking k=1 (Note: Un-standardize data).

    ```
    noOfNeigh <- 1
    ```

```
pred=knn(bankdata_trainwithoutclass, bankdata_testwithoutclass,
bankdata_train$Personal.Loan, k = noOfNeigh)
a=table(pred,bankdata_test$Personal.Loan)
a
accu= sum(diag(a))/nrow(bankdata_testwithoutclass)
accu
```

11. Repeat step 10 for different values of k(k=2,3,4,5, etc..) and observe the change in accuracy. And decide k value when there is not much change in accuracy or decrease in accuracy
12. Now repeat the steps8 to Steps11 by standardizing the independent numerical attributes.
13. Now compare your results with unstandardized data.

#Activity-2
Regression on simulated data using knn.

Assignment1:
Predict the Total revenue Generated by each Customer: Regression using KNN(Customer data.csv)

1. Read Customer data.
2. Do necessary preprocessing
   (such as changing type, missing value imputation, standardization, dummifying categorical)
3. Split in to train and test and separate the target variable
4. Build knn regression model with k=1 and check the metrics

```
Library(FNN)
noOfNeigh <- 1
prediction <- knn.reg(train = trainData, test = testData, y = train.tgt, k =
noOfNeigh)

#Evaluate the predictions
actual <- test.tgt
result2 <- rmse(actual = actual, predicted = prediction$pred)
result2
```

5. Repeat step5 for various k values(k=3,5,6,7).

Activity 3:

Objective:

- Introducing RecommenderLab
- UserBased Collbarative Filtering
- ItemBased Collbarative Filtering

Building Recommendation System on Simulated data.

Assignment 2:

Build recommendation engine on movie-ratings data set.

1. read movie-rating.csv
2. Check the structure and make the data set suitable(user-movie rating matrix) to build recommendation engine.(Hint:rehape2 library)
3. Convert to realRatingMatrix
4. Split in to train and test(known and unknown)(Hint: evaluationScheme)
5. Build UBCF and IBCF using Recommender.
6. Predict the ratings on test(known) data.
7. Compute the error metrics on unknown data.