20170204_Batch25_CSE7405c_KNN_ActivitySheet

Agenda:

- Classification and Regression using k-NN

Description:

On the universal bank data, apply k-NN algorithm for classifying prospective loan takers.

1. Read the universal bank dataset into R; call this actual data
2. Perform required data pre-processing steps
3. Split the actual data into train and test sets, without standardizing it
4. On this data, apply k-NN algorithm for classifying loan takers and non-takers; Find the error metrics on train and test sets; Remember to remove target from the train data, and give it separately in the model

   Reference:
   library(class)
   pred=knn(train[1:17],test[1:17], train$Personal.Loan,k=3)
   a=table(test$Personal.Loan,pred); a

5. Now, standardize the actual data from step (1) and then split it into train and test sets

   Reference:
   library(vegan)
   data2<-decostand(data_combined[1:17],method="standardize")
   train1<-data2[trainrows,]
   test1<-data2[-trainrows,]

6. Apply k-NN algorithm on standardized data for classifying loan takers and non-takers; Find the error metrics on train and test sets

   Reference:
   pred3=knn(train1,test1, train$Personal.Loan, k = 3)
   c=table(test$Personal.Loan,pred3)

7. Condensation of the data to reduce the number of computations and obtain minimum consistent set.
   - Consider standardized data (train1 and test 1 obtained in step 5)

     Reference:
     keep=condense(train1, train[,18])
     pred=knn(train1[keep, , drop=FALSE], test1, train$Personal.Loan[keep],k=10)
     a <- table(pred,test$Personal.Loan)
     a

20170204_Batch25_CSE7405c_KNN_ActivitySheet

#KNN- Regression

1. Clear the Environment and load the 'FNN' and 'Metrics' packages

2. Read the data (airfoil_self_noise.txt) and perform the pre-processing on it.

   #Data Description
   Data Set Information:

   The NASA data set comprises different size NACA 0012 airfoils at various wind tunnel speeds and angles of attack. The span of the airfoil and the observer position were the same in all of the experiments.

   Attribute Information:

   This problem has the following inputs:
   a. Frequency, in Hertzs.
   b. Angle of attack, in degrees.
   c. Chord length, in meters.
   d. Free-stream velocity, in meters per second.
   e. Suction side displacement thickness, in meters.

   The only output is:
   f. Scaled sound pressure level, in decibels.

   ---

3. Split the data into train and test.

4. Separate the dependent and independent attributes in train and test

5. Run the KNN algorithm to predict for test with different values of K and evaluate you regression prediction.

```
# Run the model
pred <- knn.reg(train = trainData, test = testData, y = train.tgt, k = 3 )
actual <- test.tgt
pred <- data.frame(pred$pred)
result2 <- rmse(actual = actual, predicted = pred)
result2
```