Data Driven Computing and Networking (DDCN-2019)

Classification using K-NN Algorithm

A. Write a python script to perform the following tasks:-

1. Load all the required packages to implement KNN Classification algorithm
2. Load Iris flower data set using read_csv method and store it in a variable "df"
3. Create design matrix X and target vector Y from the variable "df". The "class" column represents the target vector.
4. Split the data stored in variable "df" into training sets (X_train, Y_train) and test sets(X_test, Y_test).
5. Instantiate KNeighborsClassifier method with three number of neighbors and store its output in variable knn, representing a classifier model.
6. Fit kmeans classifer model on training data set of design matrix and target vector X.
7. Predict the target vector using test data set of the design matrix X_test.
8. Evaluate accuracy of the predicted value by comparing the predicted target value and actual target value from the test data set Y_test.
9. Create odd list of knn neighbors in range 1 to 50 and store it in a variable myList.
10. Perform 10-fold cross validation for each value of the neighbors and store cross validation score in a list cv_score.
11. Compute MSE for each value of the list cv_score.
12. Determine best value of k and print optimal number of neighbors.
13. Plot Misclassification error and all values of k.
14. Assign the x axis label as " Number of Neighbors k" and y axis label as " Misclassification Error" and display the scatter chart.