

PATTERN RECOGNITION LAB  
CSE 4214

LAB EXPERIMENT 6

IMPLEMENTING K-NEAREST NEIGHBOR (K-NN) CLASSIFIER FOR A  
GIVEN DATASET AND MEASURE PERFORMANCE OF THE ALGORITHM BY  
GENERATING F-MEASURE SCORE.

SUBMITTED BY

MD. MUKITUL ISLAM 14.02.04.076

SECTION: B1



AHSANULLAH UNIVERSITY OF SCIENCE & TECHNOLOGY

# Implementing K-nearest neighbor (K-NN) classifier for a given dataset and measure performance of the algorithm by generating F-measure score.

Md. Mukitul Islam (140204076)  
Dept. of Computer Science & Engineering  
Ahsanullah University of Science & Technology

## I. OBJECTIVE

Objective of this experiment is to understand and implement K-nearest neighbor (K-NN) classifier for a given dataset and measure performance of the algorithm by generating F-measure score.

## II. PROBLEM DESCRIPTION

In this experiment, using Nearest Neighbor clustering algorithm and Euclidean distance we have to design a classifier to classify the feature point of IRIS dataset. We also have to determine the accuracy of the IRIS dataset using F-measure score for individual classes. Considering the threshold value 4 as the nearest neighbor. We will use lower 10% for the IRIS dataset as the test data.

## III. ALGORITHM DESCRIPTION

Steps of K-Nearest Neighbor algorithm-

1. Determine parameter K = number of nearest neighbor.
2. Determine the distance between the query-instance and all the training samples.
3. Sort the distance and determine nearest neighbors based on the K-th minimum distance.
4. Gather the category of the nearest neighbors.
5. Use majority of the category of nearest neighbors as the prediction value of the query-instance.

## IV. IMPLEMENTATION

In IRIS dataset there are 3 different classes and 150 rows of data. Among them 90% data is used as training data and rest 10% data is used as testing data. For calculating distance of each testing data from training data Euclidean distance calculation technique is used.

## V. MATLAB CODE OUTPUT

There are 150 data of 3 different classes in IRIS dataset. These data are divided into two parts for testing and training.

TABLE I  
TESTING & TRAINING DATA

Training Data	Testing Data
135	15

TABLE II  
CONFUSION MATRIX

	C1	C2	C3	
C1	7	0	0	7
C2	0	3	0	3
C3	1	0	4	5

TABLE III  
PRECISION, RECALL & F1-SCORE

Precision	Recall	F1-Score
93.33	95.83	94.55

Final Accuracy: 93.33%

## VI. CONCLUSION

This matlab implementation of K-Nearest Neighbor (K-NN) algorithm follows all the steps properly for predicting class of IRIS dataset So, the K-NN algorithm is successfully implemented.