 **GHARDA INSTITUTE OF TECHNOLOGY** 

***Department of Computer Engineering***

**Machine Learning Lab BE Computer (Semester-VII)**

**Experiment No.4 : k-Nearest Neighbor Classifier**

**Aim**- To study, understand and implement a kNN classification algorithm.

**Theory**-

K Nearest Neighbors Classification is one of the classification techniques based on instance-based learning. Models based on instance-based learning to generalize beyond the training examples. To do so, they store the training examples first. When it encounters a new instance (or test example), then they instantly build a relationship between stored training examples and this new instant to assign a target function value for this new instance. Instance-based methods are sometimes called lazy learning methods because they postponed learning until the new instance is encountered for prediction.

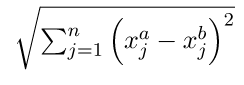
Instead of estimating the hypothetical function (or target function) once for the entire space, these methods will estimate it locally and differently for each new instance to be predicted.

#### **K-Nearest Neighbors Classifier Learning**

Basic Assumption:

1. All instances correspond to points in the n-dimensional space where n represents the number of features in any instance.
2. The nearest neighbors of an instance are defined in terms of the Euclidean distance.

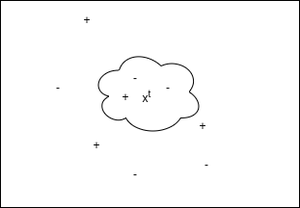
An instance can be represented by < x1, x2, .............., xn >. Euclidean distance between two instances xa and xb is given by d( xa, xb ) :



**How does it work?**

K-Nearest Neighbors Classifier first stores the training examples. During prediction, when it encounters a new instance (or test example) to predict, it finds the K number of training instances nearest to this new instance. Then assigns the most common class among the K-Nearest training instances to this test instance.

The optimal choice for K is by validating errors on test data. K can also be chosen by the square root of m, where m is the number of examples in the dataset.



*KNN Graphical Working Representation*

In the above figure, “+” denotes training instances labeled with 1. “-” denotes training instances with 0. Here we classified for the test instance xt as the most common class among K-Nearest training instances to it. Here we choose K = 3, so xt is classified as “-” or 0.

#### **Pseudocode:**

1. Store all training examples.
2. Repeat steps 3, 4, and 5 for each test example.
3. Find the K number of training examples nearest to the current test example.
4. *y\_pred* for current test example = most common class among K-Nearest training instances.
5. Go to step 2.

**Code -**

**Results & Discussion-**

**Disadvantage**

Instance Learning models are computationally very costly because all the computations are done during prediction. It also considers all the training examples for the prediction of every test example.

**Conclusion-**

The k-NN algorithm is used as a classifier and may also be used as regression.