

COURSE NAME: ARTIFICIAL INTELLIGENCE
COURSE CODE: CIS 412

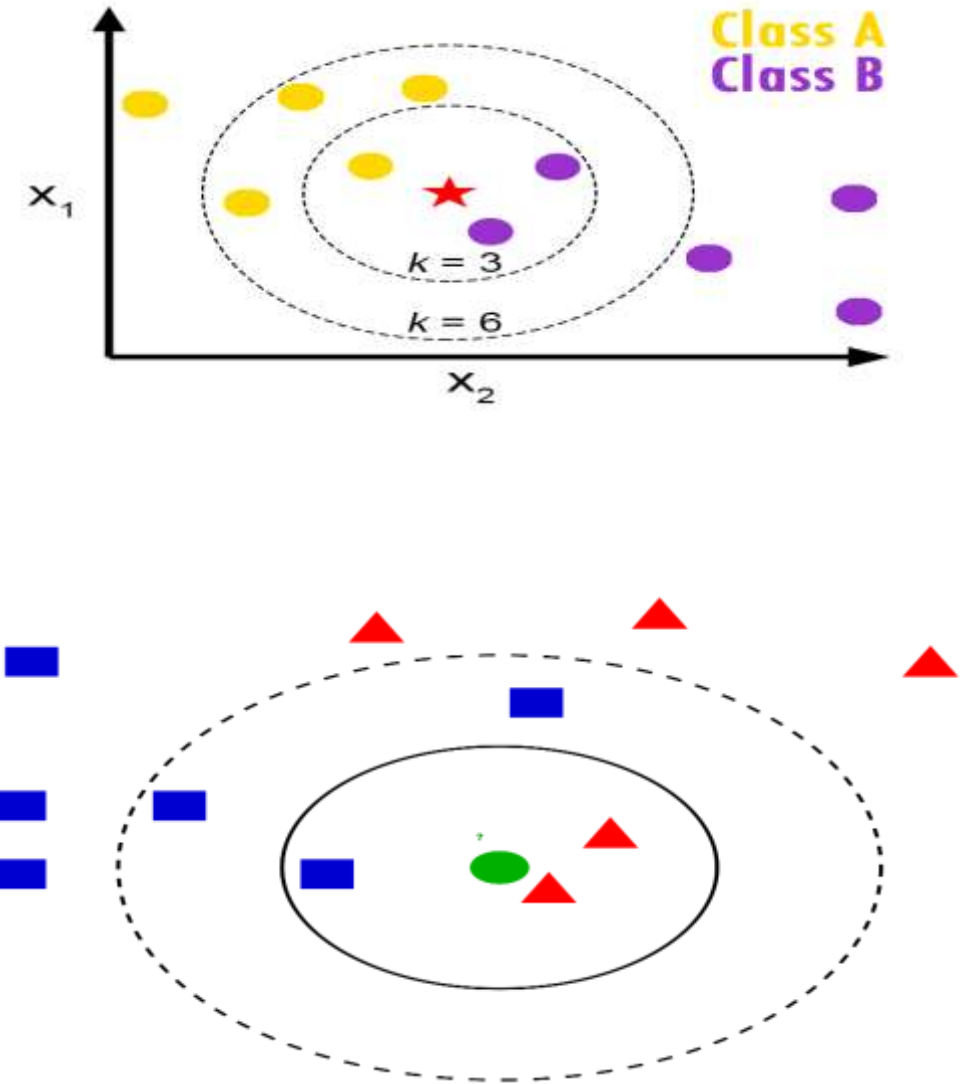
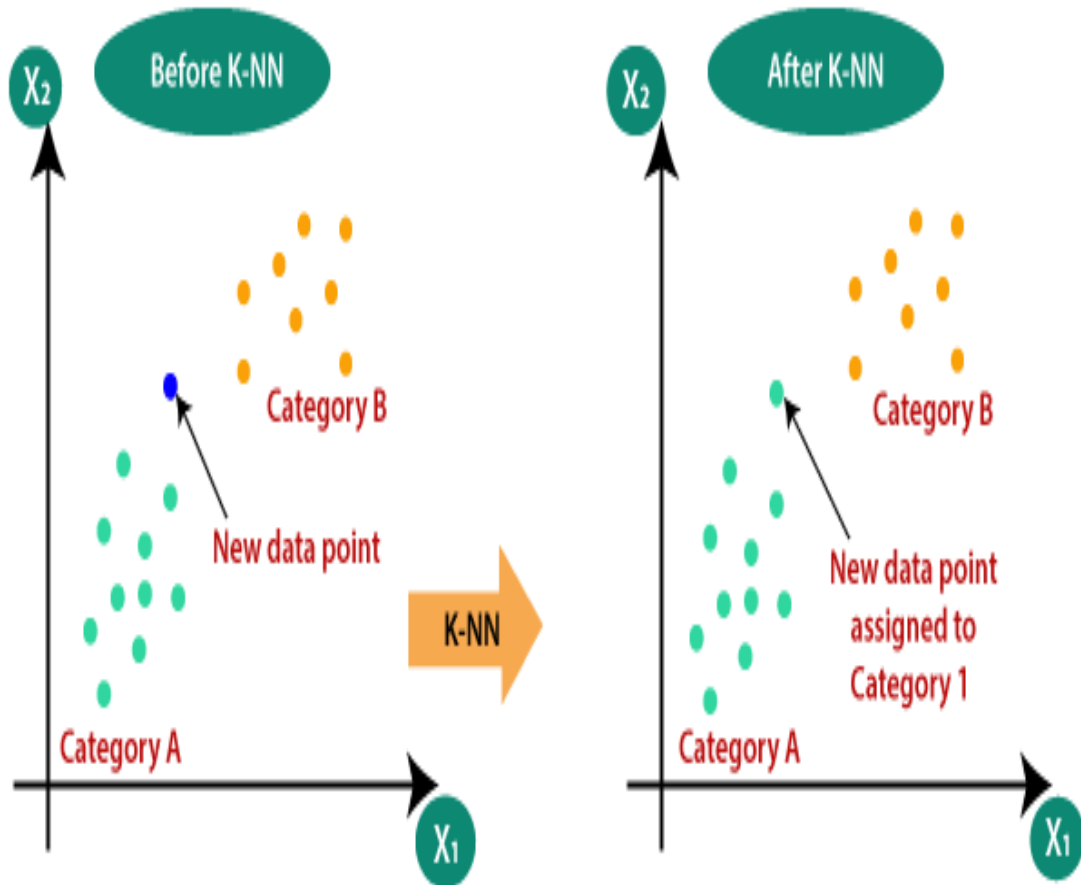
SYED TANGIM PASHA
LECTURER,

DEPARTMENT OF COMPUTING AND INFORMATION SYSTEM (CIS)
DAFFODIL INTERNATIONAL UNIVERSITY (DIU)
DHAKA, BANGLADESH

KNN

- **KNN:** The abbreviation KNN stands for “**K-Nearest Neighbor**”. It is a supervised machine learning algorithm. The algorithm can be used to **solve both classification and regression problems**.
- KNN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well category by using KNN algorithm.
- KNN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
- It is also called a **lazy learner algorithm**, because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when It gets new data, then it classifies that data into a category that is much similar to the new data.

KNN



KNN

X_1	X_2	Class
4.2	2.8	1
4.0	2.0	1
3.8	0.5	1
2.0	1.5	1
2.7	2.5	1
1.7	3.2	0
2.7	4.0	0
1.2	5.2	0
2.2	6.2	0
0.3	6.2	0

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

EUCLIDIEAN DISTANCE

KNN

X_1	X_2	Distance Formula	Distance Value	Sort Rank
4.2	2.8	$\sqrt{(4.2 - 2.2)^2 + (2.8 - 3)^2}$	2.00998	5
4.0	2.0	$\sqrt{(4 - 2.2)^2 + (2 - 3)^2}$	2.05913	6
3.8	0.5	$\sqrt{(3.8 - 2.2)^2 + (0.5 - 3)^2}$	2.96816	8

১১২

X_1	X_2	Class
1.7	3.2	0
2.7	2.5	1
2.7	4.0	0
2.0	1.5	1
4.2	2.8	1
4.0	2.0	1
1.2	5.2	0
3.8	0.5	1
2.2	6.2	0
0.3	6.2	0

অধ্যায় ৬ : কে-নিয়ারেস্ট নেইবরস (K-Nearest Neighbors)

2.0	1.5	$\sqrt{(2 - 2.2)^2 + (1.5 - 3)^2}$	1.51327	4
2.7	2.5	$\sqrt{(2.7 - 2.2)^2 + (2.5 - 3)^2}$	0.70711	2
1.7	3.2	$\sqrt{(1.7 - 2.2)^2 + (3.2 - 3)^2}$	0.53852	1
2.7	4.0	$\sqrt{(2.7 - 2.2)^2 + (4 - 3)^2}$	1.11803	3
1.2	5.2	$\sqrt{(1.2 - 2.2)^2 + (5.2 - 3)^2}$	2.41661	7
2.2	6.2	$\sqrt{(2.2 - 2.2)^2 + (6.2 - 3)^2}$	3.2	9
0.3	6.2	$\sqrt{(0.3 - 2.2)^2 + (6.2 - 3)^2}$	3.72156	10

CHOOSE K VALUE

- For binary classification, it is helpful to choose **K** to be an **odd number** as this avoid ties votes.
- Derive a plot between error rate and **K** denoting values in a defined range. Then choose the K value as having a minimum error rate.
- Took the square root value of the dataset. Ex: if $n=150$, then $K = \sqrt{150} \approx 12.28 \approx 13$ (because it's better to choose odd value for K)

Minimum error:- 0.59 at K = 37

