DAY-2

What is 'K' in K Nearest Neighbour ?

In the **k-Nearest Neighbours (k-NN)** algorithm **k** is just a number that tells the algorithm how many nearby points (neighbours) to look at when it makes a decision.

Example:

Imagine you're deciding which fruit it is based on its shape and size. You compare it to fruits you already know.

- If **k = 3**, the algorithm looks at the 3 closest fruits to the new one.
- If 2 of those 3 fruits are apples and 1 is a banana, the algorithm says the new fruit is an apple because most of its neighbours are apples.

✅ *How to Choose the Value of K in the KNN Algorithm?*

1. Cross-Validation

- This is a reliable method to choose the best K.
- The idea is to split the dataset into multiple parts (folds).
- Train the model on some parts and test it on others.
- Repeat the process and find the value of K that gives the best average performance (accuracy).

2. Elbow Method

- In this method, we plot error rate (or accuracy) for different values of K.
- As K increases, the error usually decreases at first.
- After a certain point, the improvement becomes smaller and slower.
- The point where the graph makes a clear **"elbow" shape** is considered the best K.

3. Use Odd Numbers for K (in classification)

- It's a good practice to choose **odd values of K** to avoid **ties** in voting.
- This helps the algorithm to clearly choose the majority class.

Distance Metrics Used in KNN Algorithm

KNN uses distance metrics to identify nearest neighbour, these neighbours are used for classification and regression task. To identify nearest neighbour we use below distance metrics:

1. Euclidean Distance (Most Common)

- It is the **straight-line distance** between two points — like using a ruler.
- Formula (for two points A and B):

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

2. Manhattan Distance (a.k.a. Taxicab Distance)

- Measures distance **along grid lines**, like how a taxi drives through city blocks.
- Formula:

$$d = |x_1 - x_2| + |y_1 - y_2| + \ldots$$

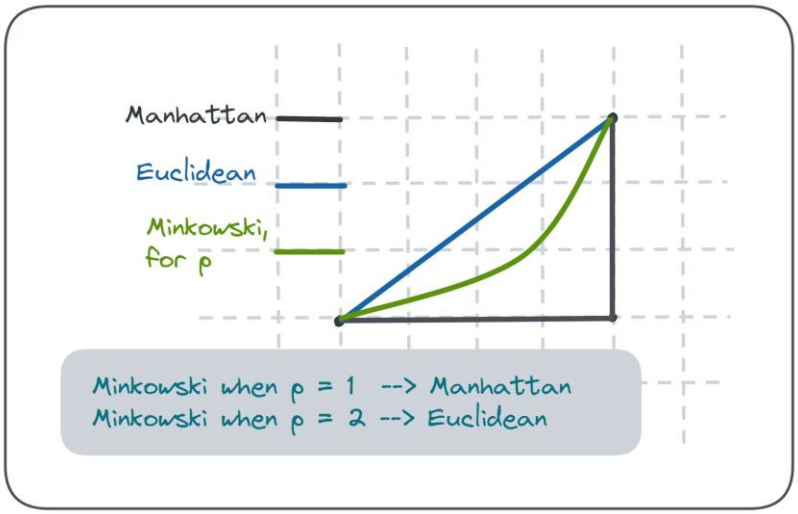
- This is useful when movement is **restricted to vertical and horizontal paths**.

3. Minkowski Distance (General Form)

- This is a **generalized distance formula** that can become either Euclidean or Manhattan distance based on a parameter **p**.
- Formula:

$$d = \left(\sum |x_i - y_i|^p\right)^{1/p}$$

  - If **p = 1**, it becomes **Manhattan Distance**
  - If **p = 2**, it becomes **Euclidean Distance**
- Think of it as a **flexible distance formula** that adapts based on the problem.



Worked on KNN Model – "HEART DISEASE DETECTION SYSTEM"

OUTPUT:

```
Age: 58
Sex (1=male, 0=female): 1
Chest Pain Type (0-3): 2
Resting Blood Pressure: 130
Cholestrol: 230
Fasting Blood Sugar > 120mg/dl (1=True, 0=False): 0
Resting ECG (0-2): 1
Max Heart Rate Achieved: 150
Exercise Induces Angina (1=yes, 0=no): 0
ST Depression: 1.2
Slope of ST Segment (0-2): 1
Number of Major Vessels (0-3): 0
Thalassemia (1=normal, 2=fixed defect, 3=reversible defect): 2
```

```
Prediction Result: ✅ No Heart Disease
```