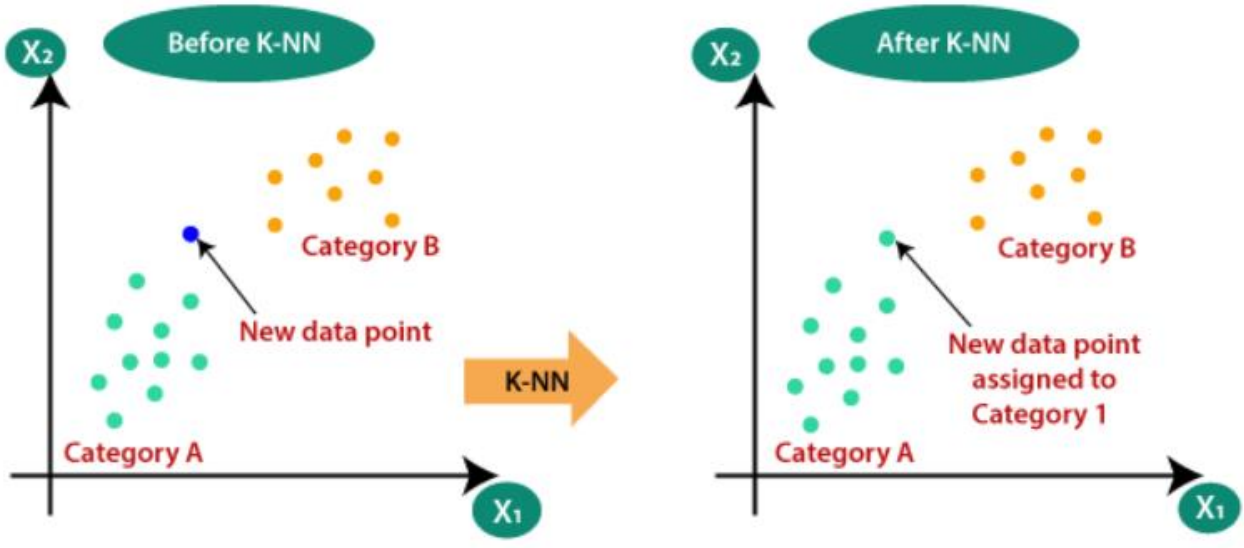
K-Nearest Neighbor(KNN)

By: Loga Aswin

- K-nearest neighbors (kNN) is a supervised machine learning technique that may be used to handle both classification and regression tasks.
- It is an algorithm that is used to classify a data point based on how its neighbors are classified.
- The "K" value refers to the number of nearest neighbor data points
- It's used in many different areas, such as handwriting detection, image recognition, and video recognition.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity.

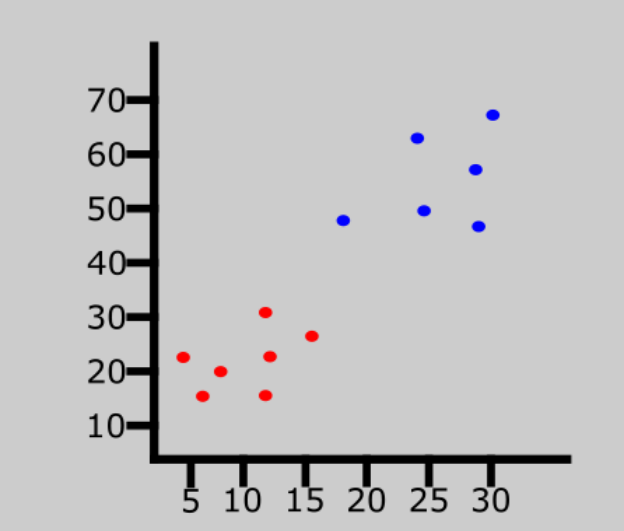
- K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
- It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.



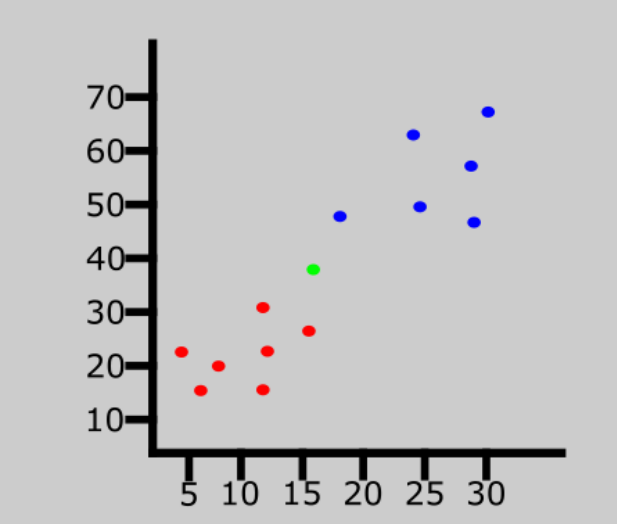
How does KNN Work?

Step 1 - Assign a value to **K**.
Step 2 - Calculate the distance between the new data entry and all other existing data entries (you'll learn how to do this shortly). Arrange them in ascending order.
Step 3 - Find the **K** nearest neighbors to the new entry based on the calculated distances.
Step 4 - Assign the new data entry to the majority class in the nearest neighbors.

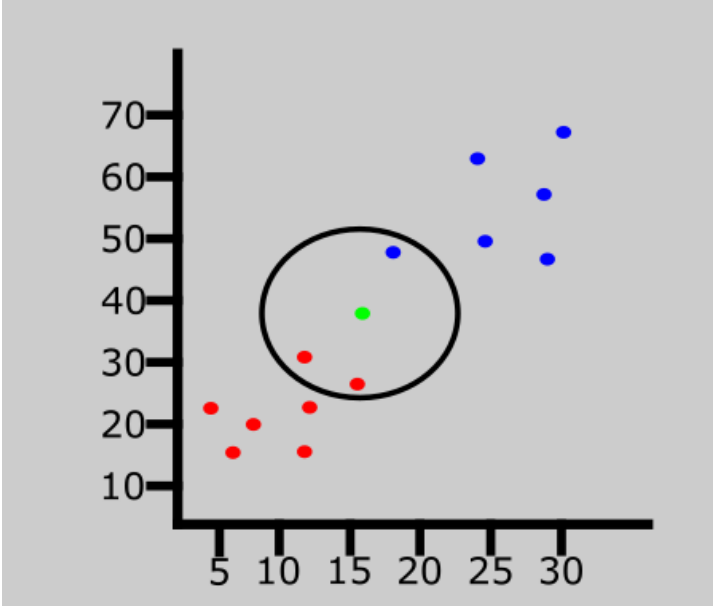
i. **data set consisting of two classes — red and blue.**



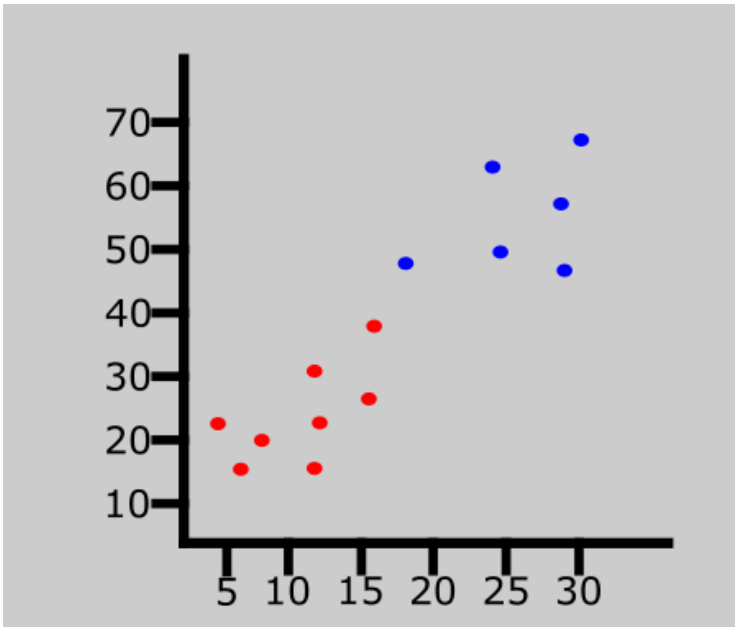
ii. **A new data has been introduced (Green point)**



iii. Let's assume the value of K is 3.



iv. Out of the 3 nearest neighbors, the majority class is red so the new entry will be assigned to that class by using Euclidean distance



Euclidean Distance A_1 and $B_2 = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$

Advantages of K-NN Algorithm

- It is simple to implement.
- No training is required before classification.

Disadvantages of K-NN Algorithm

- Can be cost-intensive when working with a large data set.
- A lot of memory is required for processing large data sets.
- Choosing the right value of **K** can be tricky.

K-Nearest Neighbor (Applications)

- RECOMMENDATION ENGINE
- CONCEPT SEARCH
- PATTERN RECOGNITION
- MISSING DATA IMPUTATION