

# KNN ALGORITHM

K-Nearest neighbour is a supervised learning algorithm. assumes the similarity between the new data and the available data points. and classifies the new data into the category similar to the existing data points.

Stores the available data during training, and classifies the new data points based on the stored similarities.

It is a Non-parametric algorithm, because it does not make predictions about underlying data distribution.

Lazy learner algorithm because it doesn't learn during the training phase, it first stores the data and uses it when required.

In regression problems, KNN uses average (or weighted average) for predicting numerical values.

## **KNN-working**

1. Choose the value of  $k$

Select a positive integer value  $k$  that represents the number of nearest neighbors to consider when making a prediction. The choice of  $k$  is crucial and can impact the performance of the algorithm.

2. Calculate the Euclidean distance

Measure the distance between the new data point and all data points in the training set

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

3. Identify the K-Nearest Neighbours

Identify the  $k$  data points from the training set that are closest to the new data point based on the calculated distances. These are the  $k$ -nearest neighbors.

4. Among these  $k$  neighbors, count the number of the data points in each category.

For Regression: If predicting a numerical value, calculate the average of the target values of the  $k$ -nearest neighbors.

#### 5. Make Prediction:

The predicted class label or value is assigned to the new data point based on the result of the average calculation.

6. Evaluate the performance of the model using metrics such as accuracy (for classification) or mean squared error (for regression). If necessary, adjust the value of  $k$  and repeat the process to improve model performance.