K-Nearest Neighbors

Definition - It classifies new data points based on the classes of their nearest neighbors in the feature space. Essentially, KNN stores all the training data and makes decisions based on the proximity of data points.

How It Works

1 - **Choosing the Number of Neighbors (K)**: The first step in the KNN algorithm is to decide the number of neighbors to consider (denoted as K). This is a critical parameter that affects the performance of the model. A small K can lead to a model that is sensitive to noise in the data, while a large K can make the model too general.

2 - **Distance Measurement**: KNN relies on a distance metric to find the nearest neighbors. The most common distance metric is Euclidean distance, but other metrics like Manhattan distance, Minkowski distance, or cosine similarity can also be used depending on the problem and data type.

3 - **Finding Neighbors**: For a given data point that needs to be classified, KNN calculates the distance between this data point and all other points in the training dataset. It then identifies the K closest data points.

4 - **Classifying the Data Point**: Once the K nearest neighbors are identified, the algorithm assigns the class label to the data point based on a majority vote. In other words, the class that appears most frequently among the K neighbors is assigned to the new data point.

5 - **Output the Classification**: The data point is classified into the category that is most common among its K nearest neighbors.