

K-Nearest Neighbors (KNN)

Description

K-Nearest Neighbors (KNN) is one of the simplest machine learning algorithms, primarily used for classification tasks. Given a set of labeled data points and a new, unlabeled point, KNN aims to assign a label to the new point based on the majority label of its 'k' nearest neighbors.

How it Works

1. **Compute Distance:** Calculate the distance between the unlabeled point and each of the labeled points. Various distance metrics like Euclidean, Manhattan, etc., can be used.
2. **Sort and Select:** Sort these distances, and select the 'k' nearest labeled points.
3. **Majority Voting:** Use majority voting to assign a label to the unlabeled point based on the most frequent label among the 'k' neighbors.

Code Example (C++)

Here's how the specific C++ code implements KNN:

1. A `struct Point` is defined to represent a point with coordinates `(x, y)` and a label.
2. The `distance` function computes the Euclidean distance between two points.
3. The `knn` function takes in the data, a test point, and 'k' as parameters. It calculates the distance between the test point and all labeled points, sorts them, and performs majority voting to determine the label of the test point.
4. The `main` function initializes some test data and a test point. It calls `knn` to predict the label for the test point and prints it out.

The code uses a simple Euclidean distance metric and majority voting with a specified 'k' (3 in this example).