

KNN

David Robinson

KNN

K-Nearest Neighbors (KNN) is a simple and effective algorithm that can be used for both classification and regression tasks in machine learning. The algorithm uses the training data to classify new data points based on how similar they are to their nearest neighbors, measured by distance.

1. Select the number of neighbors K
2. Calculate the distance between the query point and all data points in the dataset
3. Identify the K nearest neighbors based on the calculated distance
4. For classification, count the number of neighbors in each class
5. Assign the query point to the class with the majority vote

Euclidean Distance

Euclidean distance is the straight-line distance between two points in space.

$$d(x, X_i) = \sqrt{\sum_{j=1}^d (x_j - X_{ij})^2}$$

Manhattan Distance

Manhattan distance is the sum of the absolute differences between the coordinates of the points in each dimension.

$$d(x, X_i) = \sum_{j=1}^d |x_j - X_{ij}|$$

Minkowski Distance

Minkowski distance involves a parameter that adjusts the mix between Euclidean and Manhattan distance where the distance is simplified to Euclidean distance when $p = 2$ and the distance is simplified to Manhattan distance when $p = 1$.

$$d(x, X_i) = \left(\sum_{j=1}^d |x_j - X_{ij}|^p \right)^{\frac{1}{p}}$$