

Both K-Nearest Neighbors (KNN) and K-Means clustering use distance as a metric, but they use it in different ways and for different purposes. Here's a summary of how distance is utilized in each algorithm:

K-Nearest Neighbors (KNN)

- **Purpose:** KNN is a supervised learning algorithm used for classification and regression.
- **Use of Distance:**
 - To determine the proximity of data points to the query point.
 - The algorithm calculates the distance between the query point and all points in the training set.
 - Common distance metrics include Euclidean distance, Manhattan distance, and others, depending on the problem and data characteristics.
- **Steps Involving Distance:**
 1. **Distance Calculation:** Compute the distance between the query point and each point in the training set.
 2. **Nearest Neighbors:** Identify the k nearest neighbors based on the calculated distances.
 3. **Prediction:** Make a prediction based on the majority class (for classification) or average value (for regression) of the nearest neighbors.

K-Means Clustering

- **Purpose:** K-Means is an unsupervised learning algorithm used for clustering data into k groups.
- **Use of Distance:**
 - To assign data points to the nearest cluster (represented by its centroid).
 - The algorithm iteratively refines the positions of the centroids to minimize the total within-cluster variance.
 - The primary distance metric used is typically Euclidean distance.
- **Steps Involving Distance:**

1. **Assignment Step:** Compute the distance between each data point and each centroid, then assign each point to the nearest centroid.
2. **Update Step:** Recalculate the centroids as the mean of all points assigned to each centroid.
3. **Repeat:** Iterate the assignment and update steps until the centroids converge (i.e., they no longer change significantly).