

Customer Segmentation Using K-Means Clustering

Introduction to Customer Segmentation

Customer segmentation is the process of dividing customers into groups based on their shared characteristics. By categorizing customers into meaningful segments, businesses can target their marketing campaigns more effectively, improve customer service, and develop new products and services that cater to specific customer needs.

Customer segmentation allows companies to understand their customer base better, personalize their offerings, and optimize their business strategies.

Understanding K-Means Clustering

K-means clustering is a popular machine learning algorithm that can be used to segment customers based on their similarities. The algorithm aims to find k clusters in the data, where each cluster represents a group of customers that are similar to each other. K-means clustering relies on the concept of centroids, which are the center points of each cluster, and uses distance metrics to assign customers to the appropriate clusters.

K-MEANS CLUSTERING

1 Define Clusters

Specify the number k of cluster to assign.

2 Initialize Centroids

Randomly initialize k centroids.

3 Repeat

Expectation (4) & Maximization (5)
Until the Centroid positions do not change

4 Expectation

Assign each point to its closest centroid

5 Maximization

Compute the new centroid (mean) of each cluster

Equation

The diagram shows the K-means objective function formula with several annotations:

- number of clusters**: points to the variable k in the outer summation.
- number of cases**: points to the variable n in the inner summation.
- case i** : points to the index i in the inner summation.
- centroid for cluster j** : points to the variable c_j .
- Distance function**: points to the term $\|x_i^{(j)} - c_j\|^2$.
- objective function**: points to the variable J .

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

K-means Formula