K-means clustering is a popular unsupervised machine learning algorithm used to partition a dataset into a predetermined number of clusters. The algorithm aims to minimize the within-cluster sum of squares, which measures the squared Euclidean distance between each data point and the centroid of its assigned cluster. Let's break down the mathematical steps involved in the K-means clustering algorithm:

1. Initialization:

- Select the number of clusters, K.

- Randomly initialize K centroids in the feature space or use a specific initialization strategy.

2. Assigning data points to clusters:

- Calculate the Euclidean distance between each data point and all centroids.

- Assign each data point to the cluster with the nearest centroid, forming K clusters.

3. Updating centroids:

- Calculate the new centroid for each cluster by taking the mean of all data points assigned to that cluster.

- The centroid represents the center of the cluster in the feature space.

4. Iterative process:

- Repeat steps 2 and 3 until convergence or a predefined stopping criterion is met. Convergence occurs when the centroids no longer change significantly or when the maximum number of iterations is reached.

The algorithm converges to a local minimum of the within-cluster sum of squares, but it does not guarantee convergence to the global minimum. Different random initializations may result in different clustering outcomes. The mathematical objective function for K-means clustering can be expressed as:

$J = \sum_i \sum_j ||x_i - \mu_j||^2$

where:

- J represents the objective function, which measures the sum of squared distances.

- x_i is a data point.

- μ_j is the centroid of cluster j.

- $||x_i - \mu_j||^2$ denotes the squared Euclidean distance between the data point x_i and the centroid μ_j.

The goal of the algorithm is to minimize this objective function by iteratively updating the cluster assignments and the centroids. When the algorithm converges, the centroids represent the final clustering solution.