Here are the key terms commonly used in **K-Means Clustering**:

---

1. Centroid

- The center of a cluster, represented as the mean of all the points in that cluster.
- It is updated iteratively during the clustering process to minimize the distance between the points and the centroid.

---

2. Cluster

- A group of data points that are similar to each other based on a given metric, such as Euclidean distance.
- Each cluster has its own centroid.

---

3. K (Number of Clusters)

- The predefined number of clusters you want to divide the data into.
- It is a user-specified parameter in K-Means.

---

4. Inertia (Within-Cluster Sum of Squares - WCSS)

- The measure of how tightly the data points are grouped around their respective centroids.
- Lower inertia means better clustering:

$$WCSS = \sum_{k=1}^{K} \sum_{i \in \text{cluster k}} ||x_i - c_k||^2$$

where x_i is a data point and c_k is the centroid of cluster k.

---

5. Euclidean Distance

- A metric used to calculate the distance between data points and centroids:

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

where x and y are points in n-dimensional space.

---

6. Iteration

- The process of updating centroids and reassigning data points to clusters.
- K-Means repeats iterations until convergence or a maximum number of iterations is reached.

---

7. Convergence

- The point at which centroids no longer change significantly or data points stop switching clusters.
- Indicates that the algorithm has stabilized.

---

8. Initial Centroid Selection

- The starting points for centroids.
- Poor initialization can lead to suboptimal clustering; hence, techniques like **K-Means++** are used for better initialization.

---

9. K-Means++

- A smarter initialization method to choose centroids that are far apart, improving the convergence speed and accuracy of clustering.

---

10. Elbow Method

- A technique to determine the optimal number of clusters (K) by plotting WCSS against the number of clusters and looking for an "elbow point" where the rate of decrease slows.

11. Silhouette Score

- A metric to evaluate how well each data point fits into its cluster:

$$\text{Silhouette Score} = \frac{b - a}{\max(a, b)}$$

where:

  - a = average intra-cluster distance.
  - b = average distance to the nearest cluster.

---

12. Hard Clustering

- Each data point is assigned to exactly one cluster.
- K-Means is an example of hard clustering.

---

13. Outliers

- Data points that are significantly distant from any cluster centroid, potentially affecting clustering performance.

---

14. High Dimensionality

- When the dataset has many features (dimensions), it can make clustering harder due to the **curse of dimensionality**.