### **K-Means Algorithm**

The K-Means algorithm is a clustering method used to group data points into a specified number of clusters based on their similarity. Here’s how it works:

1. **Input**:
   * A dataset with multiple data points.
   * The desired number of clusters (k).
2. **Steps**:

**Step 1**: **Initialization**  
Randomly select k points (called centroids) from the dataset. These centroids represent the initial center of each cluster.

**Step 2**: **Assign Points to Clusters**  
For each data point in the dataset, assign it to the cluster whose centroid is the closest.

**Step 3**: **Update Centroids**  
After all points are assigned to clusters, calculate the new centroid for each cluster. The centroid is the "average" position of all points in the cluster.

**Step 4**: **Repeat Steps 2 and 3**  
Reassign all points to their nearest centroid and recalculate the centroids. Repeat this process until:

* + The centroids no longer change significantly.
  + A maximum number of iterations is reached.

1. **Stopping Conditions**:
   * The centroids remain stable (no further movement).
   * A predefined number of iterations is completed.
2. **Output**:
   * The k clusters, each with its assigned data points.
   * The final positions of the centroids for each cluster.

### **Conclusion :** K-Means clustering is a fundamental and efficient **unsupervised learning algorithm** used in artificial intelligence for grouping similar data points into clusters. By iteratively assigning data points to the nearest cluster centroid and updating centroids, K-Means effectively identifies hidden patterns in data. **Key Strengths of K-Means: Scalability** – Works well with large datasets due to its low computational complexity. **Simplicity** – Easy to implement and interpret. **Versatility** – Used in diverse fields such as image segmentation, customer segmentation, anomaly detection, and bioinformatics. However, K-Means has some limitations, including its sensitivity to **initial centroid selection** and the need to specify the number of clusters (**K**) in advance. Despite these challenges, enhancements like **K-Means++** and **Elbow Method** help improve performance and cluster quality. K-Means remains a **widely used and effective clustering algorithm** in artificial intelligence, providing a **fast, scalable, and intuitive** approach to data segmentation. Its ability to uncover hidden structures in large datasets makes it a valuable tool in machine learning and data mining applications.

**BOOK**

**Step 1:** Begin with a decision on the value of k = number of clusters.

**Step 2**:

Put any initial partition that classifies the data into k clusters. You may assign the training samples randomly, or systematically as the following:

1.Take the first k training sample as single-element clusters

2. Assign each of the remaining (N-k) training sample to the cluster with the nearest centroid. After each assignment, recompute the centroid of the gaining cluster.

**Step 3:** Take each sample in sequence and compute its distance from the centroid of each of the clusters.

If a sample is not currently in the cluster with the closest centroid, switch this sample to that cluster and update the centroid of the cluster gaining the new sample and the cluster losing the sample.

**Step 4 .** Repeat step 3 until convergence is achieved, that is until a pass through the training sample causes no new assignments

