**K-means Algorithm Numerical**

**Apply K-means clustering for the following data sets:**

Suppose we have the following data points (2D coordinates):

(1, 2), (2, 3), (3, 3), (6, 5), (7, 8), (8, 8)

We want to cluster these points into **2 clusters** using the K-means algorithm.

**Steps:**

1. **Initialization:** Choose the number of clusters **K=2**. Randomly initialize the cluster centroids. Let's assume we randomly pick the points **(1, 2)** and **(8, 8)** as the initial centroids.
2. **Assignment Step:** Assign each point to the nearest centroid. We calculate the Euclidean distance between each point and both centroids. The point is assigned to the cluster whose centroid is closer.
   * For centroid (1, 2):
     + Distance to (1, 2) = 0
     + Distance to (2, 3) = √((2-1)² + (3-2)²) = √2 ≈ 1.414
     + Distance to (3, 3) = √((3-1)² + (3-2)²) = √5 ≈ 2.236
     + Distance to (6, 5) = √((6-1)² + (5-2)²) = √34 ≈ 5.831
     + Distance to (7, 8) = √((7-1)² + (8-2)²) = √72 ≈ 8.485
     + Distance to (8, 8) = √((8-1)² + (8-2)²) = √98 ≈ 9.899
   * For centroid (8, 8):
     + Distance to (1, 2) = √((8-1)² + (8-2)²) = √98 ≈ 9.899
     + Distance to (2, 3) = √((8-2)² + (8-3)²) = √72 ≈ 8.485
     + Distance to (3, 3) = √((8-3)² + (8-3)²) = √50 ≈ 7.071
     + Distance to (6, 5) = √((8-6)² + (8-5)²) = √13 ≈ 3.606
     + Distance to (7, 8) = √((8-7)² + (8-8)²) = √1 = 1
     + Distance to (8, 8) = √((8-8)² + (8-8)²) = 0

**Cluster assignments:**

* + Points (1, 2), (2, 3), and (3, 3) are closer to centroid (1, 2), so they belong to Cluster 1.
  + Points (6, 5), (7, 8), and (8, 8) are closer to centroid (8, 8), so they belong to Cluster 2.

1. **Update Step:** Calculate the new centroids by finding the mean of all points in each cluster.
   * For Cluster 1 (points: (1, 2), (2, 3), (3, 3)):
     + New centroid = (2, 2.67)
   * For Cluster 2 (points: (6, 5), (7, 8), (8, 8)):
     + New centroid = (7, 7)
2. **Repeat:** Repeat steps 2 and 3 until the centroids no longer change significantly (convergence).

**Second Assignment Step:** Recalculate the distance between the points and the new centroids:

* + For centroid (2, 2.67):
    - Distance to (1, 2) = √((2-1)² + (2.67-2)²) = √1 + 0.4489 = 1.104
    - Distance to (2, 3) = √((2-2)² + (3-2.67)²) = √0 + 0.1089 = 0.33
    - Distance to (3, 3) = √((2-3)² + (2.67-3)²) = √1 + 0.1089 = 1.051
    - Distance to (6, 5) = √((2-6)² + (2.67-5)²) = √16 + 5.4489 = 5.426
    - Distance to (7, 8) = √((2-7)² + (2.67-8)²) = √25 + 28.4489 = 10.157
    - Distance to (8, 8) = √((2-8)² + (2.67-8)²) = √36 + 28.4489 = 11.111
  + For centroid (7, 7):
    - Distance to (1, 2) = √((7-1)² + (7-2)²) = √36 + 25 = 7.810
    - Distance to (2, 3) = √((7-2)² + (7-3)²) = √25 + 16 = 6.403
    - Distance to (3, 3) = √((7-3)² + (7-3)²) = √16 + 16 = 5.657
    - Distance to (6, 5) = √((7-6)² + (7-5)²) = √1 + 4 = 2.236
    - Distance to (7, 8) = √((7-7)² + (7-8)²) = √0 + 1 = 1
    - Distance to (8, 8) = √((7-8)² + (7-8)²) = √2 = 1.414

**New cluster assignments:**

* + Points (1, 2), (2, 3), and (3, 3) are still closer to centroid (2, 2.67), so they belong to Cluster 1.
  + Points (6, 5), (7, 8), and (8, 8) are closer to centroid (7, 7), so they belong to Cluster 2.

The centroids have not changed, so the algorithm has converged.

**Final Result:**

* **Cluster 1**: Points (1, 2), (2, 3), (3, 3) with centroid (2, 2.67)
* **Cluster 2**: Points (6, 5), (7, 8), (8, 8) with centroid (7, 7)

This is a simple example of how K-means clustering works.