To **recluster based on the mean**, we use the recalculated centroids (means of each cluster) to reassign data points to the closest centroid. Let's break this down step by step with an **example**:

Example

Dataset (2D points):

We have the following points in a 2D space:

$$P = \{(1,1),(2,1),(4,3),(5,4)\}$$

Assume we want to form **2 clusters** ($k = 2$).

Step 1: Initialize Centroids

Choose two random initial centroids:

$$C_1 = (1,1), \quad C_2 = (5,4)$$

Step 2: Assign Points to Closest Centroid

Calculate the **Euclidean distance** of each point to both centroids and assign each point to the closest centroid.

Distance Calculation:

1. Distance formula:
$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$
2. Calculate distances:
   - For point $(1,1)$:
     - To C_1: $d = \sqrt{(1-1)^2 + (1-1)^2} = 0$
     - To C_2: $d = \sqrt{(1-5)^2 + (1-4)^2} = \sqrt{16+9} = 5$
       Assign to C_1.
   - For point $(2,1)$:
     - To C_1: $d = \sqrt{(2-1)^2 + (1-1)^2} = 1$
     - To C_2: $d = \sqrt{(2-5)^2 + (1-4)^2} = \sqrt{9+9} = \sqrt{18} \approx 4.24$
       Assign to C_1.

- For point $(4, 3)$:
  - To C_1: $d = \sqrt{(4-1)^2 + (3-1)^2} = \sqrt{9+4} = \sqrt{13} \approx 3.61$
  - To C_2: $d = \sqrt{(4-5)^2 + (3-4)^2} = \sqrt{1+1} = \sqrt{2} \approx 1.41$
    Assign to C_2.
- For point $(5, 4)$:
  - To C_1: $d = \sqrt{(5-1)^2 + (4-1)^2} = \sqrt{16+9} = \sqrt{25} = 5$
  - To C_2: $d = \sqrt{(5-5)^2 + (4-4)^2} = 0$
    Assign to C_2.

Cluster Assignments:

- Cluster 1 (C_1): $(1, 1), (2, 1)$
- Cluster 2 (C_2): $(4, 3), (5, 4)$

Step 3: Recalculate Centroids (Cluster Means)

Compute the mean of each cluster:

For C_1:

Points: $(1, 1), (2, 1)$

$$C_1 = \left(\frac{1+2}{2}, \frac{1+1}{2}\right) = (1.5, 1)$$

For C_2:

Points: $(4, 3), (5, 4)$

$$C_2 = \left(\frac{4+5}{2}, \frac{3+4}{2}\right) = (4.5, 3.5)$$

New Centroids:

$$C_1 = (1.5, 1), \quad C_2 = (4.5, 3.5)$$

Step 4: Reassign Points to Closest Centroid

Recompute the distances between all points and the new centroids:

Distance Calculation with Updated Centroids:

- For point $(1, 1)$:
  - To C_1: $d = \sqrt{(1-1.5)^2 + (1-1)^2} = 0.5$
  - To C_2: $d = \sqrt{(1-4.5)^2 + (1-3.5)^2} = \sqrt{12.25 + 6.25} = \sqrt{18.5} \approx 4.3$
    Assign to C_1.
- For point $(2, 1)$:
  - To C_1: $d = \sqrt{(2-1.5)^2 + (1-1)^2} = 0.5$
  - To C_2: $d = \sqrt{(2-4.5)^2 + (1-3.5)^2} = \sqrt{6.25 + 6.25} = \sqrt{12.5} \approx 3.54$
    Assign to C_1.
- For point $(4, 3)$:
  - To C_1: $d = \sqrt{(4-1.5)^2 + (3-1)^2} = \sqrt{6.25 + 4} = \sqrt{10.25} \approx 3.2$
  - To C_2: $d = \sqrt{(4-4.5)^2 + (3-3.5)^2} = \sqrt{0.25 + 0.25} = \sqrt{0.5} \approx 0.71$
    Assign to C_2.
- For point $(5, 4)$:
  - To C_1: $d = \sqrt{(5-1.5)^2 + (4-1)^2} = \sqrt{12.25 + 9} = \sqrt{21.25} \approx 4.61$
  - To C_2: $d = \sqrt{(5-4.5)^2 + (4-3.5)^2} = \sqrt{0.25 + 0.25} = \sqrt{0.5} \approx 0.71$
    Assign to C_2.

Updated Cluster Assignments:

- Cluster 1 (C_1): $(1, 1), (2, 1)$
- Cluster 2 (C_2): $(4, 3), (5, 4)$

Repeat Until Convergence

Continue recalculating centroids and reassigning points until the centroids stop changing positions or the changes are below a predefined threshold.

This iterative process ensures that the clusters are formed such that intra-cluster variance is minimized!