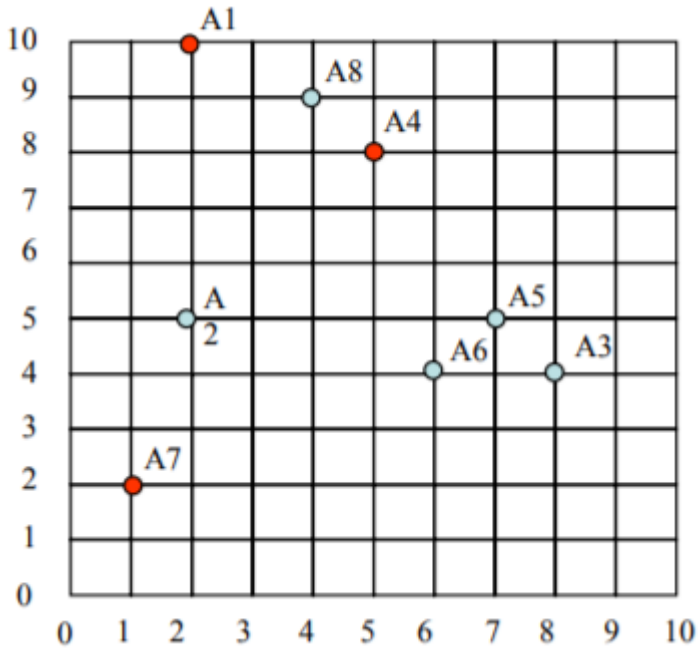
# K-means clustering

Use the *k*-means algorithm and Euclidean distance to clustering the following eight samples into three clusters. The distance matrix based on the Euclidean distance is given below.
Suppose that the initial seeds (center of each cluster) are A1, A4 and A7 (as shown in the figure). Run the k-means algorithm for 1 epoch, at the end of this epoch show

- The new clusters (i.e. the samples belonging to each cluster)
- The center of the new clusters




| ID | $x_1$ | $x_2$ |
|---|---|---|
| A1 | 2 | 10 |
| A2 | 2 | 5 |
| A3 | 8 | 4 |
| A4 | 5 | 8 |
| A5 | 7 | 5 |
| A6 | 6 | 4 |
| A7 | 1 | 2 |
| A8 | 4 | 9 |

| | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|---|---|---|---|---|---|---|---|---|
| A1 | 0 | $\sqrt{25}$ | $\sqrt{36}$ | $\sqrt{13}$ | $\sqrt{50}$ | $\sqrt{52}$ | $\sqrt{65}$ | $\sqrt{5}$ |
| A2 | | 0 | $\sqrt{37}$ | $\sqrt{18}$ | $\sqrt{25}$ | $\sqrt{17}$ | $\sqrt{10}$ | $\sqrt{20}$ |
| A3 | | | 0 | $\sqrt{25}$ | $\sqrt{2}$ | $\sqrt{2}$ | $\sqrt{53}$ | $\sqrt{41}$ |
| A4 | | | | 0 | $\sqrt{13}$ | $\sqrt{17}$ | $\sqrt{52}$ | $\sqrt{2}$ |
| A5 | | | | | 0 | $\sqrt{2}$ | $\sqrt{45}$ | $\sqrt{25}$ |
| A6 | | | | | | 0 | $\sqrt{29}$ | $\sqrt{29}$ |
| A7 | | | | | | | 0 | $\sqrt{58}$ |
| A8 | | | | | | | | 0 |

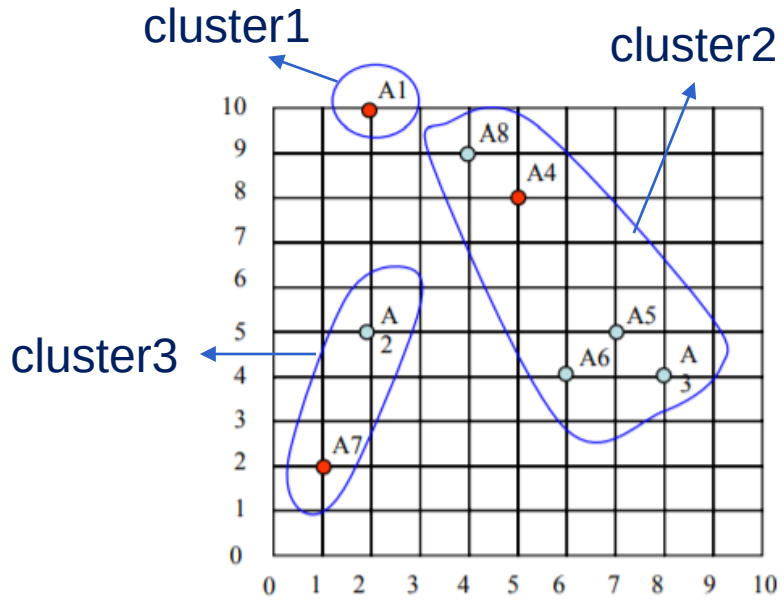
# Solutions: new clusters

| | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|---|---|---|---|---|---|---|---|---|
| A1 | 0 | $\sqrt{25}$ | $\sqrt{36}$ | $\sqrt{13}$ | $\sqrt{50}$ | $\sqrt{52}$ | $\sqrt{65}$ | $\sqrt{5}$ |
| A2 | | 0 | $\sqrt{37}$ | $\sqrt{18}$ | $\sqrt{25}$ | $\sqrt{17}$ | $\sqrt{10}$ | $\sqrt{20}$ |
| A3 | | | 0 | $\sqrt{25}$ | $\sqrt{2}$ | $\sqrt{2}$ | $\sqrt{53}$ | $\sqrt{41}$ |
| A4 | | | | 0 | $\sqrt{13}$ | $\sqrt{17}$ | $\sqrt{52}$ | $\sqrt{2}$ |
| A5 | | | | | 0 | $\sqrt{2}$ | $\sqrt{45}$ | $\sqrt{25}$ |
| A6 | | | | | | 0 | $\sqrt{29}$ | $\sqrt{29}$ |
| A7 | | | | | | | 0 | $\sqrt{58}$ |
| A8 | | | | | | | | 0 |

| Samples | Center of Cluster1<br>A1 | Center of cluster2<br>A4 | Center of Cluster3<br>A7 | |
|---|---|---|---|---|
| A1 | 0 | $\sqrt{13}$ | $\sqrt{65}$ | → A1 ∈ cluster1 |
| A2 | $\sqrt{25}$ | $\sqrt{18}$ | $\sqrt{10}$ | → A2 ∈ cluster3 |
| A3 | $\sqrt{36}$ | $\sqrt{25}$ | $\sqrt{53}$ | → A3 ∈ cluster2 |
| A4 | $\sqrt{13}$ | 0 | $\sqrt{52}$ | → A4 ∈ cluster2 |
| A5 | $\sqrt{50}$ | $\sqrt{13}$ | $\sqrt{45}$ | → A5 ∈ cluster2 |
| A6 | $\sqrt{52}$ | $\sqrt{17}$ | $\sqrt{29}$ | → A6 ∈ cluster2 |
| A7 | $\sqrt{65}$ | $\sqrt{52}$ | 0 | → A7 ∈ cluster3 |
| A8 | $\sqrt{5}$ | $\sqrt{2}$ | $\sqrt{58}$ | → A8 ∈ cluster2 |

# Solutions: new clusters

| | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|---|---|---|---|---|---|---|---|---|
| A1 | 0 | $\sqrt{25}$ | $\sqrt{36}$ | $\sqrt{13}$ | $\sqrt{50}$ | $\sqrt{52}$ | $\sqrt{65}$ | $\sqrt{5}$ |
| A2 | | 0 | $\sqrt{37}$ | $\sqrt{18}$ | $\sqrt{25}$ | $\sqrt{17}$ | $\sqrt{10}$ | $\sqrt{20}$ |
| A3 | | | 0 | $\sqrt{25}$ | $\sqrt{2}$ | $\sqrt{2}$ | $\sqrt{53}$ | $\sqrt{41}$ |
| A4 | | | | 0 | $\sqrt{13}$ | $\sqrt{17}$ | $\sqrt{52}$ | $\sqrt{2}$ |
| A5 | | | | | 0 | $\sqrt{2}$ | $\sqrt{45}$ | $\sqrt{25}$ |
| A6 | | | | | | 0 | $\sqrt{29}$ | $\sqrt{29}$ |
| A7 | | | | | | | 0 | $\sqrt{58}$ |
| A8 | | | | | | | | 0 |



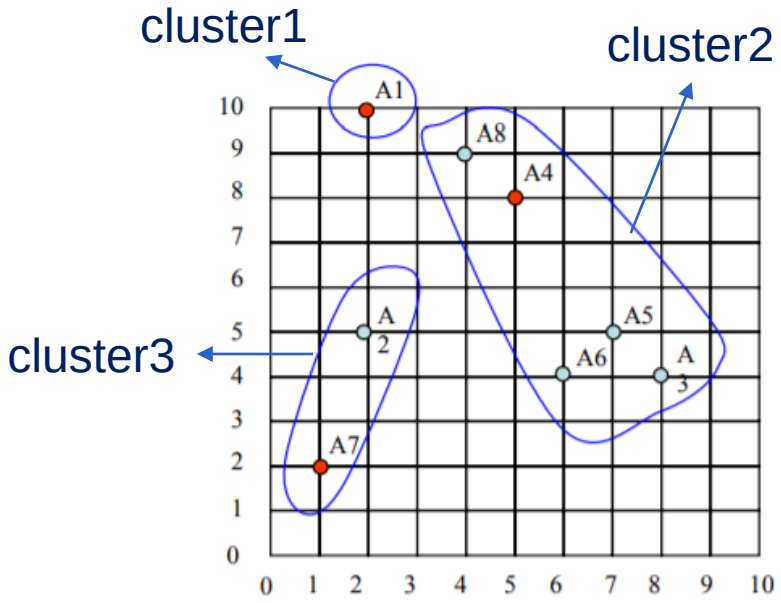

| Samples | Center of Cluster1 A1 | Center of cluster2 A4 | Center of Cluster3 A7 | |
|---|---|---|---|---|
| A1 | 0 | $\sqrt{13}$ | $\sqrt{65}$ | → A1 ∈ cluster1 |
| A2 | $\sqrt{25}$ | $\sqrt{18}$ | $\sqrt{10}$ | → A2 ∈ cluster3 |
| A3 | $\sqrt{36}$ | $\sqrt{25}$ | $\sqrt{53}$ | → A3 ∈ cluster2 |
| A4 | $\sqrt{13}$ | 0 | $\sqrt{52}$ | → A4 ∈ cluster2 |
| A5 | $\sqrt{50}$ | $\sqrt{13}$ | $\sqrt{45}$ | → A5 ∈ cluster2 |
| A6 | $\sqrt{52}$ | $\sqrt{17}$ | $\sqrt{29}$ | → A6 ∈ cluster2 |
| A7 | $\sqrt{65}$ | $\sqrt{52}$ | 0 | → A7 ∈ cluster3 |
| A8 | $\sqrt{5}$ | $\sqrt{2}$ | $\sqrt{58}$ | → A8 ∈ cluster2 |

# Solutions: new cluster centers

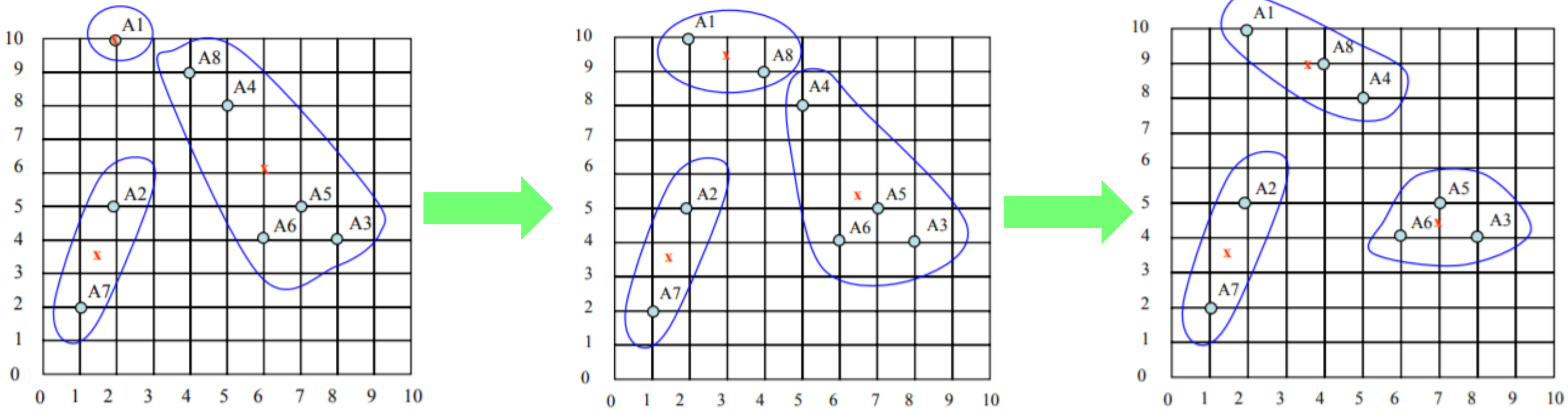
- New cluster centers:
  - Cluster 1: A1=(2,10)
    - New cluster center: C1 = (2, 10)
  - Cluster 2: A8=(4,9);A4=(5,8);A5=(7,5);A6=(6,4); A3=(8,4)
    - New cluster center:

      C2 = ((4+5+7+6+8)/5, (9+8+5+4+4)/5) = (6, 6)
  - Cluster3: A2(2,5); A7(1,2)
    - New cluster center:

      C2 = ((2+1)/2, (5+2)/2) = (1.5, 3.5)




| ID | $x_1$ | $x_2$ |
|---|---|---|
| A1 | 2 | 10 |
| A2 | 2 | 5 |
| A3 | 8 | 4 |
| A4 | 5 | 8 |
| A5 | 7 | 5 |
| A6 | 6 | 4 |
| A7 | 1 | 2 |
| A8 | 4 | 9 |


A1
A8
A4
A2
A5
A6
A3
A7
x
0 1 2 3 4 5 6 7 8 9 10