

Exercise 1. Partition clustering

Use the C-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters:

$A1 = (2,10)$, $A2 = (2,5)$, $A3 = (8,4)$, $A4 = (5,8)$,
 $A5 = (7,5)$, $A6 = (6,4)$, $A7 = (1,2)$, $A8 = (4,9)$.

The distance matrix based on the Euclidean distance is given below:

	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

Suppose that the initial seeds (centres of each cluster) are A1, A4 and A7. Run the C-means algorithm for 1 epoch only. At the end of this epoch show:

- The new clusters (i.e. the examples belonging to each cluster)
- The centres of the new clusters
- Draw a 10 by 10 space with all the 8 points and show the clusters after the first epoch and the new centroids.
- How many more iterations are needed to converge? Draw the result for each epoch.

Exercise 2. Hierarchical clustering (I)

Use single-linkage and complete-linkage agglomerative clustering using the Euclidean distance to group the data described by the following distance matrix. Show the distance matrices and dendrograms.

	A	B	C	D
A	0	1	4	5
B		0	2	6
C			0	3
D				0

Exercise 3. Hierarchical clustering (II)

Use complete-linkage agglomerative clustering using the Euclidean distance to group the data described by the following distance matrix. Show the distance matrices.

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	5.00	6.00	3.61	7.07	7.21	8.06	2.24
A2		0	6.08	4.24	5.00	4.12	3.16	4.47
A3			0	5.00	1.41	1.41	7.28	6.40
A4				0	3.61	4.12	7.21	1.41
A5					0	1.41	6.71	5.00
A6						0	5.39	5.39
A7							0	7.62
A8								0