Exercise 1. Partition clustering

Solution:

a)

d(a,b) denotes the Eucledian distance between a and b. It is obtained directly from the distance matrix or calculated as follows: $d(a,b)=\text{sqrt}((x_b-x_a)^2+(y_b-y_a)^2))$

seed1=A1=(2,10), seed2=A4=(5,8), seed3=A7=(1,2)

epoch1 – start:

A1:
d(A1, seed1)=0 as A1 is seed1
d(A1, seed2)= $\sqrt{13}$ >0
d(A1, seed3)= $\sqrt{65}$ >0
➔A1 ∈ cluster1

A2:
d(A2,seed1)= $\sqrt{25}$ = 5
d(A2, seed2)= $\sqrt{18}$ = 4.24
d(A2, seed3)= $\sqrt{10}$ = 3.16 ← smaller
➔ A2 ∈ cluster3

A3:
d(A3, seed1)= $\sqrt{36}$ = 6
d(A3, seed2)= $\sqrt{25}$ = 5 ← smaller
d(A3, seed3)= $\sqrt{53}$ = 7.28
➔ A3 ∈ cluster2

A4:
d(A4, seed1)= $\sqrt{13}$
d(A4, seed2)=0 as A4 is seed2
d(A4, seed3)= $\sqrt{52}$ >0
➔ A4 ∈ cluster2

A5:
d(A5, seed1)= $\sqrt{50}$ = 7.07
d(A5, seed2)= $\sqrt{13}$ = 3.60 ← smaller
d(A5, seed3)= $\sqrt{45}$ = 6.70
➔ A5 ∈ cluster2

A6:
d(A6, seed1)= $\sqrt{52}$ = 7.21
d(A6, seed2)= $\sqrt{17}$ = 4.12 ← smaller
d(A6, seed3)= $\sqrt{29}$ = 5.38
➔ A6 ∈ cluster2

A7:
d(A7, seed1)= $\sqrt{65}$ >0
d(A7, seed2)= $\sqrt{52}$ >0
d(A7, seed3)=0 as A7 is seed3
➔ A7 ∈ cluster3

A8:
d(A8, seed1)= $\sqrt{5}$
d(A8, seed2)= $\sqrt{2}$ ← smaller
d(A8, seed3)= $\sqrt{58}$
➔ A8 ∈ cluster2

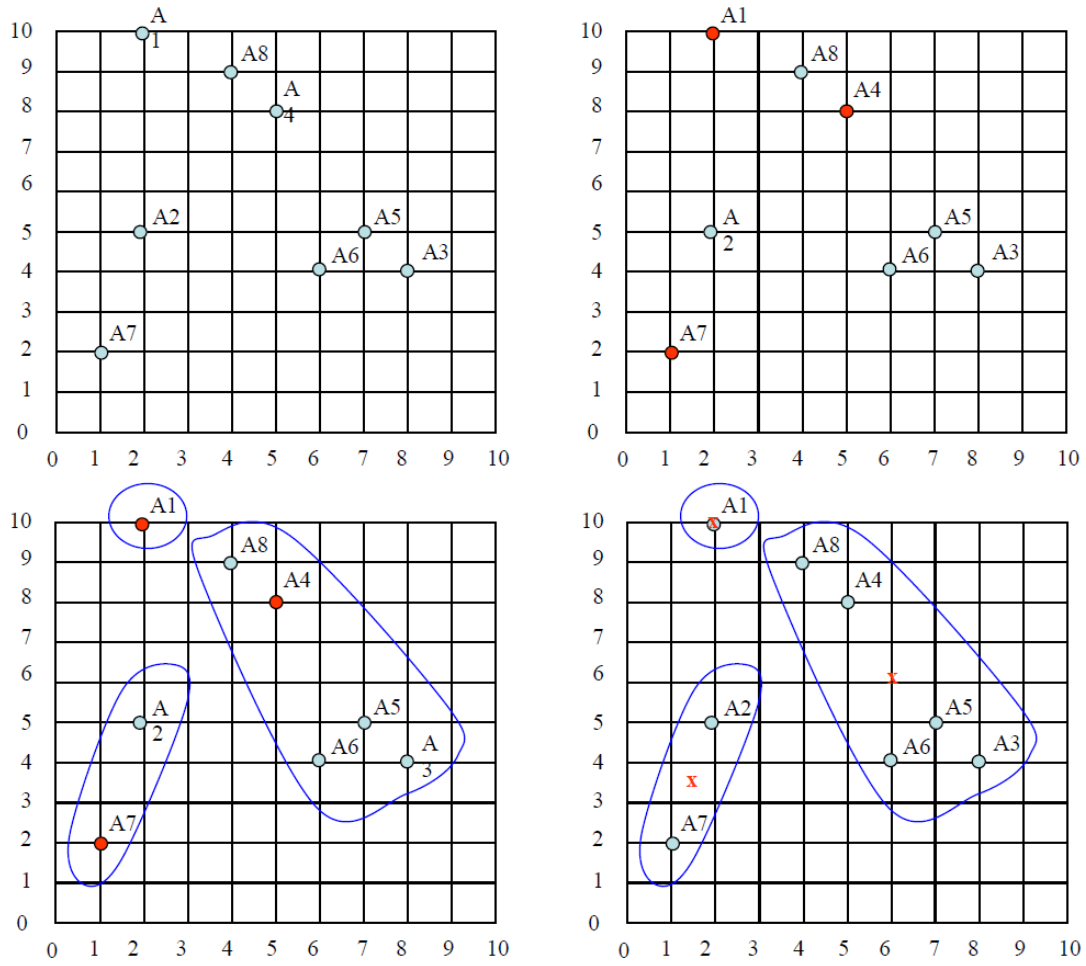
end of epoch1

new clusters: 1: {A1}, 2: {A3, A4, A5, A6, A8}, 3: {A2, A7}

b) centers of the new clusters:

C1= (2, 10), C2= ((8+5+7+6+4)/5, (4+8+5+4+9)/5) = (6, 6), C3= ((2+1)/2, (5+2)/2) = (1.5, 3.5)

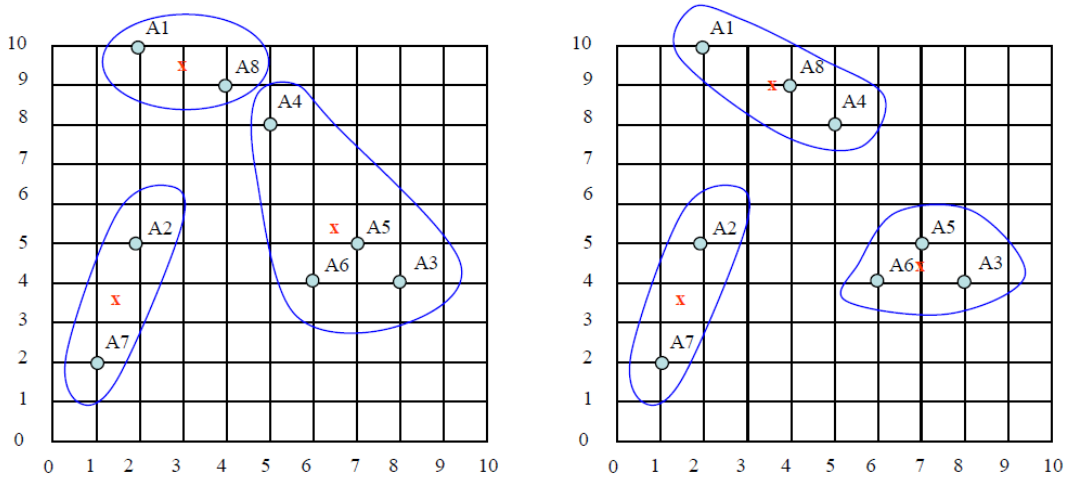
c)



---

d)

We would need two more epochs. After the 2nd epoch the results would be:
1: {A1, A8}, 2: {A3, A4, A5, A6}, 3: {A2, A7}
with centers C1=(3, 9.5), C2=(6.5, 5.25) and C3=(1.5, 3.5).
After the 3rd epoch, the results would be:
1: {A1, A4, A8}, 2: {A3, A5, A6}, 3: {A2, A7}
with centers C1=(3.66, 9), C2=(7, 4.33) and C3=(1.5, 3.5).



Exercise 2. Hierarchical clustering (I)

a) Single-linkage

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

Clusters: {A,B}, {C}, {D}

(A-B,C) = min (d(A,C),d(B,C)) = min(4,2) = 2

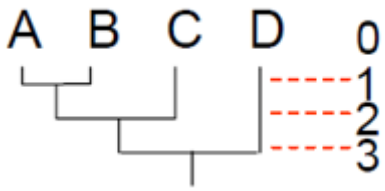
(A-B,D) = min(d(A,D),d(B,D)) = min(5,6) = 5

	A-B	C	D
A-B		2	5
C			3
D			

Clusters: {A,B,C}, {D}

(A-B-C, D) = min (d(A-B,D), d(C,D)) = min(5,3) = 3

Clusters: {A,B,C,D}



b) Complete-linkage

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

Clusters: {A,B}, {C}, {D}

(A-B,C) = max (d(A,C),d(B,C)) = max(4,2) = 4

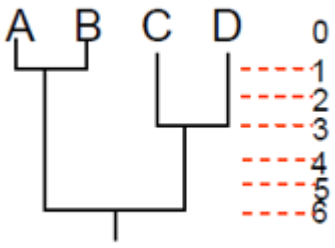
(A-B,D) = max(d(A,D),d(B,D)) = max(5,6) = 6

	A-B	C	D
A-B		4	6
C			3
D			

Clusters: {A,B}, {C,D}

(A-B,C-D) = max (d(A,C),d(B,C),d(A,D),d(B,D)) = max(4,2,5,6) = 6

Clusters: {A,B,C,D}



Exercise 3. Hierarchical clustering (II)

	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

The first clusters are between A4 and A8 and between A3, A5 and A6 because the distance between A4 and A8 and the distances between A3 and A5, A3 and A6, A5 and A6 are the smallest (1.41).

Clusters: {A1}, {A2}, {A7}, {A4,A8}, {A3,A5,A6}

(A1,A4-A8) = MAX(d(A1,A4),d(A1,A8)) = MAX(3.61,2.24) = 3.61
(A2,A4-A8) = MAX(d(A2,A4),d(A2,A8)) = MAX(4.24,4.47) = 4.47
(A7,A4-A8) = MAX(d(A7,A4),d(A7,A8)) = MAX(7.21,7.62) = 7.62
(A1, A3-A5-A6 = MAX(d(A1,A3),d(A1,A5),d(A1,A6)) = MAX(6.00,7.07,7.21) = 7.21
(A2, A3-A5-A6 = MAX(d(A2,A3),d(A2,A5),d(A2,A6)) = MAX(6.08,5.00,4.12) = 6.08
(A7, A3-A5-A6 = MAX(d(A7,A3),d(A7,A5),d(A7,A6)) = MAX(7.28,6.71,5.39) = 7.28
(A3-A5-A6, A4-A8) = MAX(d(A3,A4),d(A3,A8),d(A5,A4),d(A5,A8),d(A6,A4),d(A6,A8))
= MAX(5.00,6.40,3.61,5.00,4.12,1.41) = 6.40

	A1	A2	A3-A5-A6	A4-A8	A7
A1	0	5.00	7.21	3.61	8.06
A2		0	6.08	4.47	3.16
A3-A5-A6			0	6.40	7.28
A4-A8				0	7.62
A7					0

Clusters: {A1}, {A4,A8}, {A3,A5,A6}, {A2,A7}

(A1,A2-A7) = MAX(d(A1,A2),d(A1,A7)) = MAX(5.00,8.06) = 8.06
(A3-A5-A6, A2-A7) = MAX(d(A3,A2),d(A3,A7),d(A5,A2),d(A5,A7),d(A6,A2),d(A6,A7))
= MAX(6.08,7.28,5.00,6.71,4.12,5.39) = 7.28
(A2-A7,A4-A8) = MAX(d(A2,A4),d(A2,A8),d(A7,A4),d(A7,A8)) =
MAX(4.24,4.47,7.21,7.62) = 7.62

	A1	A2-A7	A3-A5-A6	A4-A8
A1	0	8.06	7.21	3.61
A2-A7		0	7.28	7.62
A3-A5-A6			0	6.40
A4-A8				0

Clusters: {A4,A8,A1}, {A3,A5,A6}, {A2,A7}

(A2-A7,A1-A4-A8) = MAX(d(A2,A1),d(A2,A4),d(A2,A8),d(A7,A1),d(A7,A4),d(A7,A8))
= MAX(5.00,4.24,4.47,8.06,7.21,7.62) = 8.06
(A3-A5-A6, A1-A4-A8) = MAX(d(A3,A1),d(A3,A4),d(A3,A8),
d(A5,A1),d(A5,A4),d(A5,A8),d(A6,A1),d(A6,A4),d(A6,A8)) =
MAX(6.00,5.00,6.40,7.07,3.61,5.00,7.21,4.12,1.41) = 7.21

	A1-A4-A8	A2-A7	A3-A5-A6
A1-A4-A8	0	8.06	7.21
A2-A7		0	7.28
A3-A5-A6			0

Clusters: {A4,A8,A1,A3,A5,A6}, {A2,A7}

In the next iteration, it would result in a single cluster:

Clusters: {A4,A8,A1,A3,A5,A6,A2,A7}