6434476923 พัทธ์ อาคาปินธรรม

10.2 Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters, where the points are

$$A_1(2,10), A_2(2,5), A_3(8,4), B_1(5,8), B_2(7,5), B_3(6,4), C_1(1,2), C_2(4,9).$$

The distance function is Euclidean distance. Suppose initially we assign A_1, B_1, and C_1 as the center of each cluster, respectively. Use the *k-means* algorithm to show *only*

(a) The three cluster centers after the first round of execution.

(b) The final three clusters.

1) เลือก k

ให้ $C_1 = A_1 = (2, 10)$

$C_2 = B_1 = (5, 8)$

$C_3 = C_1 = (1, 2)$

จะได้ centroid 3 จุด คือ A_1 , B_1 , C_1

2) คำนวณระยะห่างแต่ละจุดกับ centroid ทั้งสาม

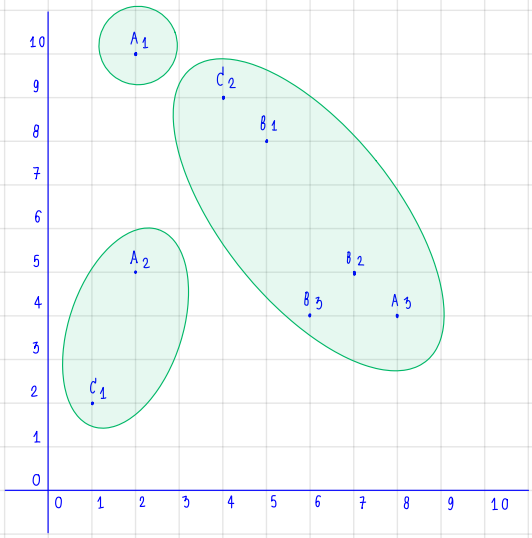
$$\text{distance} = (x_1 - x_2)^2 + (y_1 - y_2)^2$$

จุด	พิกัด	ระยะห่างจาก C_1 (2, 10)	ระยะห่างจาก C_2 (5, 8)	ระยะห่างจาก C_3 (1, 2)	ระยะห่างที่น้อยที่สุด
A_1	(2, 10)	$d = \sqrt{(2-2)^2+(10-10)^2} = 0$	$d = \sqrt{(2-5)^2+(10-8)^2} = \sqrt{13}$	$d = \sqrt{(2-1)^2+(10-2)^2} = \sqrt{65}$	C_1
A_2	(2, 5)	$d = \sqrt{(2-2)^2+(5-10)^2} = 5$	$d = \sqrt{(2-5)^2+(5-8)^2} = \sqrt{18}$	$d = \sqrt{(2-1)^2+(5-2)^2} = \sqrt{10}$	C_3
A_3	(8, 4)	$d = \sqrt{(8-2)^2+(4-10)^2} = \sqrt{72}$	$d = \sqrt{(8-5)^2+(4-8)^2} = 5$	$d = \sqrt{(8-1)^2+(4-2)^2} = \sqrt{53}$	C_2
B_1	(5, 8)	$d = \sqrt{(5-2)^2+(8-10)^2} = \sqrt{13}$	$d = \sqrt{(5-5)^2+(8-8)^2} = 0$	$d = \sqrt{(5-1)^2+(8-2)^2} = \sqrt{52}$	C_2
B_2	(7, 5)	$d = \sqrt{(7-2)^2+(5-10)^2} = \sqrt{50}$	$d = \sqrt{(7-5)^2+(5-8)^2} = \sqrt{13}$	$d = \sqrt{(7-1)^2+(5-2)^2} = \sqrt{45}$	C_2
B_3	(6, 4)	$d = \sqrt{(6-2)^2+(4-10)^2} = \sqrt{52}$	$d = \sqrt{(6-5)^2+(4-8)^2} = \sqrt{17}$	$d = \sqrt{(6-1)^2+(4-2)^2} = \sqrt{29}$	C_2
C_1	(1, 2)	$d = \sqrt{(1-2)^2+(2-10)^2} = \sqrt{65}$	$d = \sqrt{(1-5)^2+(2-8)^2} = \sqrt{52}$	$d = \sqrt{(1-1)^2+(2-2)^2} = 0$	C_3
C_2	(4, 9)	$d = \sqrt{(4-2)^2+(9-10)^2} = \sqrt{5}$	$d = \sqrt{(4-5)^2+(9-8)^2} = \sqrt{2}$	$d = \sqrt{(4-1)^2+(9-2)^2} = \sqrt{58}$	C_2

จะได้ $C_1 = A_1$

$C_2 = A_3, B_1, B_2, B_3, C_2$

$C_3 = A_2, C_1$



6434476923 พัทธ์ อาคาปิบธรรม

3) หา mean point ใหม่ของแต่ละ cluster → เพื่อขยับค่า K

$$C_m = \frac{\sum_{i=1}^{N} (t_{ip})}{N}$$

Centroid of C_1 = A_1 = (2, 10) เพราะมีจุดเดียว

Centroid of C_2 → $x = \frac{8+5+7+6+4}{5} = 6$

$y = \frac{4+8+5+4+9}{5} = 6$

Centroid of C_2 = (6, 6)

Centroid of C_3 → $x = \frac{2+1}{2} = 1.5$

$y = \frac{5+2}{2} = 3.5$

Centroid of C_3 = (1.5, 3.5)

4) คำนวณระยะห่างแต่ละจุดกับ centroid ทั้งสาม

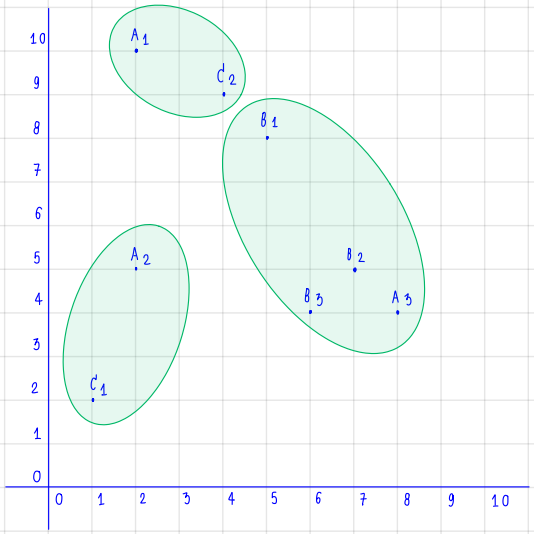
$$distance = (x_1 - x_2)^2 + (y_1 - y_2)^2$$

จุด	พิกัด	ระยะห่างจาก C_1 (2, 10)	ระยะห่างจาก C_2 (6, 6)	ระยะห่างจาก C_3 (1.5, 3.5)	ระยะห่างที่น้อยที่สุด
A_1	(2, 10)	$d = \sqrt{(2-2)^2+(10-10)^2} = 0$	$d = \sqrt{(2-6)^2+(10-6)^2} = \sqrt{32}$	$d = \sqrt{(2-1.5)^2+(10-3.5)^2} = \sqrt{72}$	C_1
A_2	(2, 5)	$d = \sqrt{(2-2)^2+(5-10)^2} = 5$	$d = \sqrt{(2-6)^2+(5-6)^2} = \sqrt{17}$	$d = \sqrt{(2-1.5)^2+(5-3.5)^2} = \sqrt{2.25}$	C_3
A_3	(8, 4)	$d = \sqrt{(8-2)^2+(4-10)^2} = \sqrt{72}$	$d = \sqrt{(8-6)^2+(4-6)^2} = \sqrt{8}$	$d = \sqrt{(8-1.5)^2+(4-3.5)^2} = \sqrt{36.25}$	C_2
B_1	(5, 8)	$d = \sqrt{(5-2)^2+(8-10)^2} = \sqrt{13}$	$d = \sqrt{(5-6)^2+(8-6)^2} = \sqrt{5}$	$d = \sqrt{(5-1.5)^2+(8-3.5)^2} = \sqrt{29.25}$	C_2
B_2	(7, 5)	$d = \sqrt{(7-2)^2+(5-10)^2} = \sqrt{50}$	$d = \sqrt{(7-6)^2+(5-6)^2} = \sqrt{2}$	$d = \sqrt{(7-1.5)^2+(5-3.5)^2} = \sqrt{27.25}$	C_2
B_3	(6, 4)	$d = \sqrt{(6-2)^2+(4-10)^2} = \sqrt{52}$	$d = \sqrt{(6-6)^2+(4-6)^2} = 2$	$d = \sqrt{(6-1.5)^2+(4-3.5)^2} = \sqrt{4.25}$	C_2
C_1	(1, 2)	$d = \sqrt{(1-2)^2+(2-10)^2} = \sqrt{65}$	$d = \sqrt{(1-6)^2+(2-6)^2} = \sqrt{41}$	$d = \sqrt{(1-1.5)^2+(2-3.5)^2} = \sqrt{3.25}$	C_3
C_2	(4, 9)	$d = \sqrt{(4-2)^2+(9-10)^2} = \sqrt{5}$	$d = \sqrt{(4-6)^2+(9-6)^2} = \sqrt{13}$	$d = \sqrt{(4-1.5)^2+(9-3.5)^2} = \sqrt{34.25}$	C_1

จะได้ว่า $C_1 = A_1, C_2$

$C_2 = A_3, B_1, B_2, B_3$

$C_3 = A_2, C_1$



6434476923 พัทธ์ อาคาปิบธรรม

5) หา mean point ใหม่ของแต่ละ cluster → เพื่อขยับค่า K

$$C_m = \frac{\sum_{i=1}^{N} (t_{ip})}{N}$$

Centroid of C_1 → $x = \frac{2+4}{2} = 3$

$y = \frac{10+9}{2} = 9.5$

Centroid of C_1 = $(3, 9.5)$

Centroid of C_2 → $x = \frac{8+5+7+6}{4} = 6.5$

$y = \frac{4+8+5+4}{4} = 5.25$

Centroid of C_2 = $(6.5, 5.25)$

Centroid of C_3 → $x = \frac{2+1}{2} = 1.5$

$y = \frac{5+2}{2} = 3.5$

Centroid of C_3 = $(1.5, 3.5)$

6) คำนวณระยะห่างแต่ละจุดกับ centroid ทั้งสาม

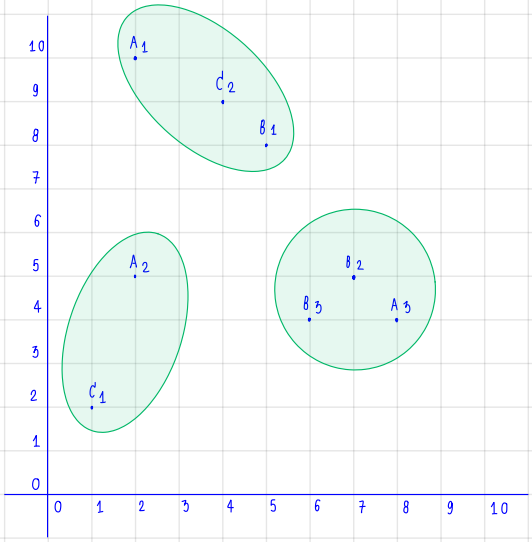
$$distance = (x_1 - x_2)^2 + (y_1 - y_2)^2$$

จุด	พิกัด	ระยะห่างจาก C_1 (3, 9.5)	ระยะห่างจาก C_2 (6.5, 5.25)	ระยะห่างจาก C_3 (1.5, 3.5)	ระยะห่างที่น้อยที่สุด
A_1	(2, 10)	$d = \sqrt{(2-3)^2+(10-9.5)^2} = \sqrt{1.25}$	$d = \sqrt{(2-6.5)^2+(10-5.25)^2} = \sqrt{42.8125}$	$d = \sqrt{(2-1.5)^2+(10-3.5)^2} = \sqrt{72}$	C_1
A_2	(2, 5)	$d = \sqrt{(2-3)^2+(5-9.5)^2} = \sqrt{21.25}$	$d = \sqrt{(2-6.5)^2+(5-5.25)^2} = \sqrt{20.3125}$	$d = \sqrt{(2-1.5)^2+(5-3.5)^2} = \sqrt{2.25}$	C_3
A_3	(8, 4)	$d = \sqrt{(8-3)^2+(4-9.5)^2} = \sqrt{55.25}$	$d = \sqrt{(8-6.5)^2+(4-5.25)^2} = \sqrt{3.8125}$	$d = \sqrt{(8-1.5)^2+(4-3.5)^2} = \sqrt{36.25}$	C_2
B_1	(5, 8)	$d = \sqrt{(5-3)^2+(8-9.5)^2} = 2.5$	$d = \sqrt{(5-6.5)^2+(8-5.25)^2} = \sqrt{9.8125}$	$d = \sqrt{(5-1.5)^2+(8-3.5)^2} = \sqrt{29.25}$	C_1
B_2	(7, 5)	$d = \sqrt{(7-3)^2+(5-9.5)^2} = \sqrt{36.25}$	$d = \sqrt{(7-6.5)^2+(5-5.25)^2} = \sqrt{0.3125}$	$d = \sqrt{(7-1.5)^2+(5-3.5)^2} = \sqrt{27.25}$	C_2
B_3	(6, 4)	$d = \sqrt{(6-3)^2+(4-9.5)^2} = \sqrt{39.25}$	$d = \sqrt{(6-6.5)^2+(4-5.25)^2} = \sqrt{1.8125}$	$d = \sqrt{(6-1.5)^2+(4-3.5)^2} = \sqrt{4.25}$	C_2
C_1	(1, 2)	$d = \sqrt{(1-3)^2+(2-9.5)^2} = \sqrt{60.25}$	$d = \sqrt{(1-6.5)^2+(2-5.25)^2} = \sqrt{40.8125}$	$d = \sqrt{(1-1.5)^2+(2-3.5)^2} = \sqrt{3.25}$	C_3
C_2	(4, 9)	$d = \sqrt{(4-3)^2+(9-9.5)^2} = \sqrt{1.25}$	$d = \sqrt{(4-6.5)^2+(9-5.25)^2} = \sqrt{20.3125}$	$d = \sqrt{(4-1.5)^2+(9-3.5)^2} = \sqrt{34.25}$	C_1

จะได้ว่า $C_1 = A_1, B_1, C_2$

$C_2 = A_3, B_2, B_3$

$C_3 = A_2, C_1$



6434476923 พัทธ์ อาคาปิยธรรม

7) หา mean point ใหม่ของแต่ละ cluster → เพื่อขยับค่า K

$$C_m = \frac{\sum_{i=1}^{N} (t_{ip})}{N}$$

Centroid of C_1 → $x = \frac{2+4+5}{3} = 3.67$

$y = \frac{10+9+8}{3} = 9$

Centroid of C_1 = (3.67, 9)

Centroid of C_2 → $x = \frac{8+7+6}{3} = 7$

$y = \frac{4+5+4}{3} = 4.33$

Centroid of C_2 = (7, 4.33)

Centroid of C_3 → $x = \frac{2+1}{2} = 1.5$

$y = \frac{5+2}{2} = 3.5$

Centroid of C_3 = (1.5, 3.5)

8) คำนวณระยะห่างแต่ละจุดกับ centroid ทั้งสาม

$$distance = (x_1 - x_2)^2 + (y_1 - y_2)^2$$

จุด	พิกัด	ระยะห่างจาก C_1 (3.67, 9)	ระยะห่างจาก C_2 (7, 4.33)	ระยะห่างจาก C_3 (1.5, 3.5)	ระยะห่างที่น้อยที่สุด
A_1	(2, 10)	$d = \sqrt{(2-3.67)^2+(10-9)^2} = \sqrt{3.7889}$	$d = \sqrt{(2-7)^2+(10-4.33)^2} = \sqrt{57.1489}$	$d = \sqrt{(2-1.5)^2+(10-3.5)^2} = \sqrt{72}$	C_1
A_2	(2, 5)	$d = \sqrt{(2-3.67)^2+(5-9)^2} = \sqrt{18.7889}$	$d = \sqrt{(2-7)^2+(5-4.33)^2} = \sqrt{25.4489}$	$d = \sqrt{(2-1.5)^2+(5-3.5)^2} = \sqrt{2.25}$	C_3
A_3	(8, 4)	$d = \sqrt{(8-3.67)^2+(4-9)^2} = \sqrt{43.7489}$	$d = \sqrt{(8-7)^2+(4-4.33)^2} = \sqrt{1.1089}$	$d = \sqrt{(8-1.5)^2+(4-3.5)^2} = \sqrt{36.25}$	C_2
B_1	(5, 8)	$d = \sqrt{(5-3.67)^2+(8-9)^2} = \sqrt{2.7689}$	$d = \sqrt{(5-7)^2+(8-4.33)^2} = \sqrt{17.4689}$	$d = \sqrt{(5-1.5)^2+(8-3.5)^2} = \sqrt{29.25}$	C_1
B_2	(7, 5)	$d = \sqrt{(7-3.67)^2+(5-9)^2} = \sqrt{27.0889}$	$d = \sqrt{(7-7)^2+(5-4.33)^2} = 0.67$	$d = \sqrt{(7-1.5)^2+(5-3.5)^2} = \sqrt{27.25}$	C_2
B_3	(6, 4)	$d = \sqrt{(6-3.67)^2+(4-9)^2} = \sqrt{30.4289}$	$d = \sqrt{(6-7)^2+(4-4.33)^2} = \sqrt{1.1089}$	$d = \sqrt{(6-1.5)^2+(4-3.5)^2} = \sqrt{4.25}$	C_2
C_1	(1, 2)	$d = \sqrt{(1-3.67)^2+(2-9)^2} = \sqrt{56.1289}$	$d = \sqrt{(1-7)^2+(2-4.33)^2} = \sqrt{41.4289}$	$d = \sqrt{(1-1.5)^2+(2-3.5)^2} = \sqrt{3.25}$	C_3
C_2	(4, 9)	$d = \sqrt{(4-3.67)^2+(9-9)^2} = 0.33$	$d = \sqrt{(4-7)^2+(9-4.33)^2} = \sqrt{30.8089}$	$d = \sqrt{(4-1.5)^2+(9-3.5)^2} = \sqrt{34.25}$	C_1

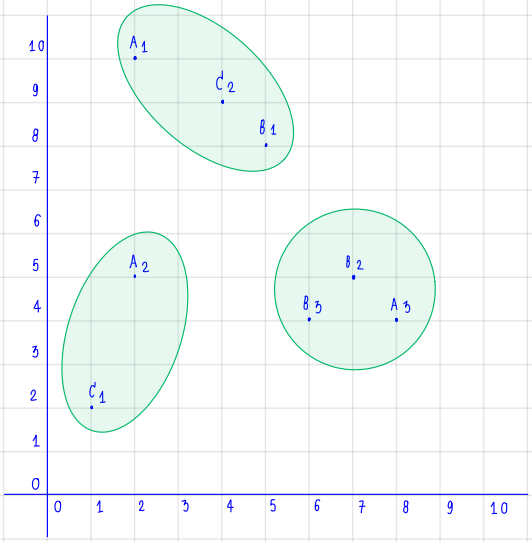
สรุปได้ว่า $C_1 = A_1, B_1, C_2$

$C_2 = A_3, B_2, B_3$

$C_3 = A_2, C_1$

ไม่มีการเปลี่ยนแปลงของ กลุ่ม cluster

↳ จบ



ตอบ cluster 1 = $A_1(2,10)$, $B_1(5,8)$, $C_2(4,9)$

cluster 2 = $A_3(8,4)$, $B_2(7,5)$, $B_3(6,4)$

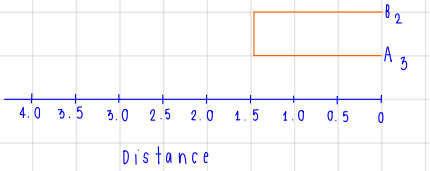
cluster 3 = $A_2(2,5)$, $C_1(1,2)$ #

Dendogram ด้วยวิธี single linkage

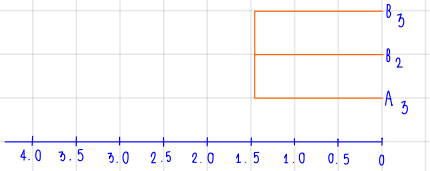
6434476923 พัทธ์ อาคาปิยธรรม

หา distance matrix distance = $(x_1 - x_2)^2 + (y_1 - y_2)^2$

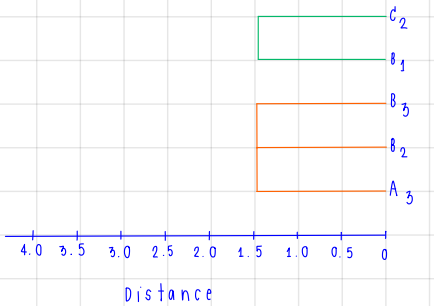
	A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2
A_1	0							
A_2	5	0						
A_3	$\sqrt{72}$	$\sqrt{37}$	0					
B_1	$\sqrt{13}$	$\sqrt{18}$	5	0				
B_2	$\sqrt{50}$	5	$\sqrt{2}$	$\sqrt{13}$	0			
B_3	$\sqrt{52}$	$\sqrt{17}$	2	$\sqrt{17}$	$\sqrt{2}$	0		
C_1	$\sqrt{65}$	$\sqrt{10}$	$\sqrt{53}$	$\sqrt{52}$	$\sqrt{45}$	$\sqrt{29}$	0	
C_2	$\sqrt{5}$	$\sqrt{20}$	$\sqrt{41}$	$\sqrt{2}$	5	$\sqrt{29}$	$\sqrt{58}$	0



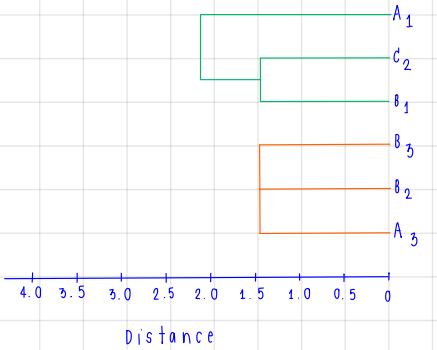
	A_1	A_2	A_3B_2	B_1	B_3	C_1	C_2
A_1	0						
A_2	5	0					
A_3B_2	$\sqrt{50}$	5	0				
B_1	$\sqrt{13}$	$\sqrt{18}$	5	0			
B_3	$\sqrt{52}$	$\sqrt{17}$	$\sqrt{2}$	$\sqrt{17}$	0		
C_1	$\sqrt{65}$	$\sqrt{10}$	$\sqrt{45}$	$\sqrt{52}$	$\sqrt{29}$	0	
C_2	$\sqrt{5}$	$\sqrt{20}$	5	$\sqrt{2}$	$\sqrt{29}$	$\sqrt{58}$	0



	A_1	A_2	$A_3B_2B_3$	B_1	C_1	C_2
A_1	0					
A_2	5	0				
$A_3B_2B_3$	$\sqrt{50}$	$\sqrt{17}$	0			
B_1	$\sqrt{13}$	$\sqrt{18}$	5	0		
C_1	$\sqrt{65}$	$\sqrt{10}$	$\sqrt{29}$	$\sqrt{52}$	0	
C_2	$\sqrt{5}$	$\sqrt{20}$	5	$\sqrt{2}$	$\sqrt{58}$	0

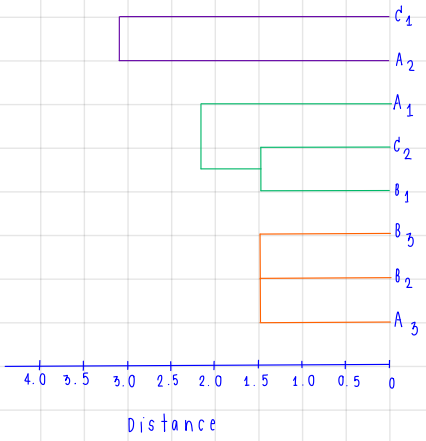


	A_1	A_2	$A_3B_2B_3$	B_1C_2	C_1
A_1	0				
A_2	5	0			
$A_3B_2B_3$	$\sqrt{50}$	$\sqrt{17}$	0		
B_1C_2	$\sqrt{5}$	$\sqrt{18}$	5	0	
C_1	$\sqrt{65}$	$\sqrt{10}$	$\sqrt{29}$	$\sqrt{52}$	0

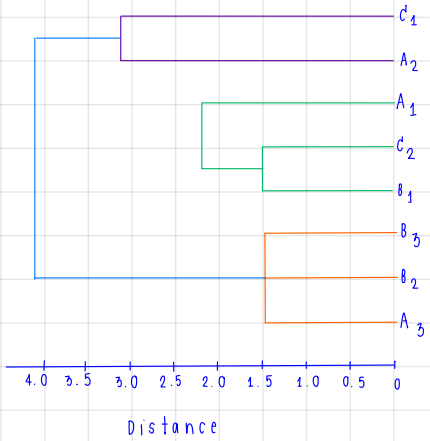


6434476923 พัทธ์ อาคาปิยธรรม

	$A_1B_1C_2$	A_2	$A_3B_2B_3$	C_1
$A_1B_1C_2$	0			
A_2	5	0		
$A_3B_2B_3$	$\sqrt{50}$	$\sqrt{17}$	0	
C_1	$\sqrt{52}$	$\sqrt{10}$	$\sqrt{29}$	0



	$A_1B_1C_2$	A_2C_1	$A_3B_2B_3$
$A_1B_1C_2$	0		
A_2C_1	5	0	
$A_3B_2B_3$	$\sqrt{50}$	$\sqrt{17}$	0



	$A_1B_1C_2$	$A_2C_1A_3B_2B_3$
$A_1B_1C_2$	0	
$A_2C_1A_3B_2B_3$	5	0