Q7. **Apply K-means algorithm on given data for k=3. Use $C_1(2)$, $C_2(16)$ and $C_3(38)$ as initial cluster centres.**

Data: 2, 4, 6, 3, 31, 12, 15, 16, 38, 35, 14, 21, 23, 25, 30

Ans: **[10M | May16 & Dec16]**

Number of clusters k = 3

Initial cluster centre for $C_1 = 2$, $C_2 = 16$, $C_3 = 38$

We will check distance between data points and all cluster centres. We will use Euclidean Distance formula for finding distance.

Distance [x, a] = $\sqrt{(x-a)^2}$

OR

Distance [(x, y), (a, b)] = $\sqrt{(x-a)^2 + (y-b)^2}$

As given data is not in pair, we will use first formula of Euclidean Distance.

Finding Distance between data points and cluster centres.

We will use following notations for calculating Distance:

D_1 = Distance from cluster C_1 centre

D_2 = Distance from cluster C_2 centre

D_3 = Distance from cluster C_3 centre

<u>2:</u>

$D_1(2, 2) = \sqrt{(x-a)^2} = \sqrt{(2-2)^2} = 0$

$D_2(2, 16) = \sqrt{(x-a)^2} = \sqrt{(2-16)^2} = 14$

$D_3(2, 38) = \sqrt{(x-a)^2} = \sqrt{(2-38)^2} = 34$

Here 0 is smallest distance so Data point 2 belongs to C_1

<u>4:</u>

$D_1(4, 2) = \sqrt{(x-a)^2} = \sqrt{(4-2)^2} = 2$

$D_2(4, 16) = \sqrt{(x-a)^2} = \sqrt{(4-16)^2} = 12$

$D_3(4, 38) = \sqrt{(x-a)^2} = \sqrt{(4-38)^2} = 34$

Here 2 is smallest distance so Data point 4 belongs to C_1

<u>6:</u>

$D_1(6, 2) = \sqrt{(x-a)^2} = \sqrt{(6-2)^2} = 4$

$D_2(6, 16) = \sqrt{(x-a)^2} = \sqrt{(6-16)^2} = 10$

$D_3(6, 38) = \sqrt{(x-a)^2} = \sqrt{(6-38)^2} = 32$

Here 4 is smallest distance so Data point 6 belongs to C_1

<u>3:</u>

$D_1(3, 2) = \sqrt{(x-a)^2} = \sqrt{(3-2)^2} = 1$

$D_2(3, 16) = \sqrt{(x-a)^2} = \sqrt{(3-16)^2} = 13$

$D_3(3, 38) = \sqrt{(x-a)^2} = \sqrt{(3-38)^2} = 35$

Here 1 is smallest distance so Data point 3 belongs to C_1

31:

$D_1(31, 2) = \sqrt{(x-a)^2} = \sqrt{(31-2)^2} = 29$

$D_2(31, 16) = \sqrt{(x-a)^2} = \sqrt{(31-16)^2} = 15$

$D_3(31, 38) = \sqrt{(x-a)^2} = \sqrt{(31-38)^2} = 7$

Here 0 is smallest distance so Data point 31 belongs to C_3

12:

$D_1(12, 2) = \sqrt{(x-a)^2} = \sqrt{(12-2)^2} = 10$

$D_2(12, 16) = \sqrt{(x-a)^2} = \sqrt{(12-16)^2} = 4$

$D_3(12, 38) = \sqrt{(x-a)^2} = \sqrt{(12-38)^2} = 26$

Here 0 is smallest distance so Data point 12 belongs to C_2

15:

$D_1(15, 2) = \sqrt{(x-a)^2} = \sqrt{(15-2)^2} = 13$

$D_2(15, 16) = \sqrt{(x-a)^2} = \sqrt{(15-16)^2} = 1$

$D_3(15, 38) = \sqrt{(x-a)^2} = \sqrt{(15-38)^2} = 23$

Here 1 is smallest distance so Data point 15 belongs to C_2

16:

$D_1(16, 2) = \sqrt{(x-a)^2} = \sqrt{(16-2)^2} = 14$

$D_2(16, 16) = \sqrt{(x-a)^2} = \sqrt{(16-16)^2} = 0$

$D_3(16, 38) = \sqrt{(x-a)^2} = \sqrt{(16-38)^2} = 22$

Here 0 is smallest distance so Data point 4 belongs to C_2

38:

$D_1(38, 2) = \sqrt{(x-a)^2} = \sqrt{(38-2)^2} = 36$

$D_2(38, 16) = \sqrt{(x-a)^2} = \sqrt{(38-16)^2} = 22$

$D_3(38, 38) = \sqrt{(x-a)^2} = \sqrt{(38-38)^2} = 0$

Here 0 is smallest distance so Data point 38 belongs to C_3

35:

$D_1(35, 2) = \sqrt{(x-a)^2} = \sqrt{(35-2)^2} = 33$

$D_2(35, 16) = \sqrt{(x-a)^2} = \sqrt{(35-16)^2} = 21$

$D_3(35, 38) = \sqrt{(x-a)^2} = \sqrt{(35-38)^2} = 3$

Here 3 is smallest distance so Data point 35 belongs to C_3

14:

$D_1(14, 2) = \sqrt{(x-a)^2} = \sqrt{(14-2)^2} = 12$

$D_2(14, 16) = \sqrt{(x-a)^2} = \sqrt{(14-16)^2} = 2$

$D_3(14, 38) = \sqrt{(x-a)^2} = \sqrt{(14-38)^2} = 24$

Here 2 is smallest distance so Data point 14 belongs to C_2

21:

$D_1(21, 2) = \sqrt{(x-a)^2} = \sqrt{(21-2)^2} = 19$

$D_2(21, 16) = \sqrt{(x-a)^2} = \sqrt{(21-16)^2} = 5$

$D_3(21, 38) = \sqrt{(x-a)^2} = \sqrt{(21-38)^2} = 17$

Here 5 is smallest distance so Data point 21 belongs to C_2

23:

$D_1(23, 2) = \sqrt{(x-a)^2} = \sqrt{(23-2)^2} = 21$

$D_2(23, 16) = \sqrt{(x-a)^2} = \sqrt{(23-16)^2} = 7$

$D_3(23, 38) = \sqrt{(x-a)^2} = \sqrt{(23-38)^2} = 15$

Here 7 is smallest distance so Data point 23 belongs to C_2

25:

$D_1(25, 2) = \sqrt{(x-a)^2} = \sqrt{(25-2)^2} = 23$

$D_2(25, 16) = \sqrt{(x-a)^2} = \sqrt{(25-16)^2} = 9$

$D_3(25, 38) = \sqrt{(x-a)^2} = \sqrt{(25-38)^2} = 13$

Here 9 is smallest distance so Data point 25 belongs to C_2

30:

$D_1(30, 2) = \sqrt{(x-a)^2} = \sqrt{(30-2)^2} = 28$

$D_2(30, 16) = \sqrt{(x-a)^2} = \sqrt{(30-16)^2} = 14$

$D_3(30, 38) = \sqrt{(x-a)^2} = \sqrt{(30-38)^2} = 8$

Here 8 is smallest distance so Data point 30 belongs to C_3

The clusters will be,

$C_1 = \{2, 4, 6, 3\}$,

$C_2 = \{12, 15, 16, 14, 21, 23, 25\}$,

$C_3 = \{31, 38, 35, 30\}$

Now we have to recalculate the centre of these clusters as following

$C_1 = \frac{2+4+6+3}{4} = \frac{15}{4} = 3.75$ (we can round off this value to 4 also)

$C_2 = \frac{12+15+16+14+21+23+25}{7} = \frac{126}{7} = 18$

$C_3 = \frac{31+38+35+30}{4} = \frac{134}{4} = 33.5$ (we can round of this value to 34 also)

Now we will again calculate distance from each data point to all new cluster centres,

2:

$D_1(2, 4) = \sqrt{(x-a)^2} = \sqrt{(2-4)^2} = 2$

$D_2(2, 18) = \sqrt{(x-a)^2} = \sqrt{(2-18)^2} = 16$

$D_3(2, 34) = \sqrt{(x-a)^2} = \sqrt{(2-34)^2} = 32$

Here 2 is smallest distance so Data point 2 belongs to C_1

4:

$D_1(4, 4) = \sqrt{(x-a)^2} = \sqrt{(4-4)^2} = 0$

$D_2(4, 18) = \sqrt{(x-a)^2} = \sqrt{(4-18)^2} = 14$

$D_3(4, 34) = \sqrt{(x-a)^2} = \sqrt{(4-34)^2} = 30$

Here 0 is smallest distance so Data point 4 belongs to C_1

6:

$D_1(6, 4) = \sqrt{(x-a)^2} = \sqrt{(6-4)^2} = 2$

$D_2(6, 18) = \sqrt{(x-a)^2} = \sqrt{(6-18)^2} = 12$

$D_3(6, 34) = \sqrt{(x-a)^2} = \sqrt{(6-34)^2} = 28$

Here 2 is smallest distance so Data point 6 belongs to C_1

3:

$D_1(3, 4) = \sqrt{(x-a)^2} = \sqrt{(3-4)^2} = 1$

$D_2(3, 18) = \sqrt{(x-a)^2} = \sqrt{(3-18)^2} = 15$

$D_3(3, 34) = \sqrt{(x-a)^2} = \sqrt{(3-34)^2} = 31$

Here 1 is smallest distance so Data point 3 belongs to C_1

31:

$D_1(31, 4) = \sqrt{(x-a)^2} = \sqrt{(31-4)^2} = 27$

$D_2(31, 18) = \sqrt{(x-a)^2} = \sqrt{(31-18)^2} = 13$

$D_3(31, 34) = \sqrt{(x-a)^2} = \sqrt{(31-34)^2} = 3$

Here 3 is smallest distance so Data point 31 belongs to C_3

12:

$D_1(12, 4) = \sqrt{(x-a)^2} = \sqrt{(12-4)^2} = 8$

$D_2(12, 18) = \sqrt{(x-a)^2} = \sqrt{(12-18)^2} = 6$

$D_3(12, 34) = \sqrt{(x-a)^2} = \sqrt{(12-34)^2} = 22$

Here 6 is smallest distance so Data point 12 belongs to C_2

15:

$D_1(15, 4) = \sqrt{(x-a)^2} = \sqrt{(15-4)^2} = 11$

$D_2(15, 18) = \sqrt{(x-a)^2} = \sqrt{(15-18)^2} = 3$

$D_3(15, 34) = \sqrt{(x-a)^2} = \sqrt{(15-34)^2} = 19$

Here 3 is smallest distance so Data point 15 belongs to C_2

16:

$D_1(16, 4) = \sqrt{(x-a)^2} = \sqrt{(16-4)^2} = 12$

$D_2(16, 18) = \sqrt{(x-a)^2} = \sqrt{(16-18)^2} = 2$

$D_3(16, 34) = \sqrt{(x-a)^2} = \sqrt{(16-34)^2} = 18$

Here 2 is smallest distance so Data point 4 belongs to C_2

38:

$D_1(38, 4) = \sqrt{(x-a)^2} = \sqrt{(38-4)^2} = 34$

$D_2(38, 18) = \sqrt{(x-a)^2} = \sqrt{(38-18)^2} = 20$

$D_3(38, 34) = \sqrt{(x-a)^2} = \sqrt{(38-34)^2} = 4$

Here 4 is smallest distance so Data point 38 belongs to C_3

35:

$D_1(35, 4) = \sqrt{(x-a)^2} = \sqrt{(35-4)^2} = 31$

$D_2(35, 18) = \sqrt{(x-a)^2} = \sqrt{(35-18)^2} = 17$

$D_3(35, 34) = \sqrt{(x-a)^2} = \sqrt{(35-34)^2} = 1$

Here 1 is smallest distance so Data point 35 belongs to C_3

14:

$D_1(14, 4) = \sqrt{(x-a)^2} = \sqrt{(14-4)^2} = 10$

$D_2(14, 18) = \sqrt{(x-a)^2} = \sqrt{(14-18)^2} = 4$

$D_3(14, 34) = \sqrt{(x-a)^2} = \sqrt{(14-34)^2} = 20$

Here 4 is smallest distance so Data point 14 belongs to C_2

21:

$D_1(21, 4) = \sqrt{(x-a)^2} = \sqrt{(21-4)^2} = 17$

$D_2(21, 18) = \sqrt{(x-a)^2} = \sqrt{(21-18)^2} = 3$

$D_3(21, 34) = \sqrt{(x-a)^2} = \sqrt{(21-34)^2} = 13$

Here 3 is smallest distance so Data point 21 belongs to C_2

23:

$D_1(23, 4) = \sqrt{(x-a)^2} = \sqrt{(23-4)^2} = 19$

$D_2(23, 18) = \sqrt{(x-a)^2} = \sqrt{(23-18)^2} = 5$

$D_3(23, 34) = \sqrt{(x-a)^2} = \sqrt{(23-34)^2} = 11$

Here 5 is smallest distance so Data point 23 belongs to C_2

25:

$D_1(25, 4) = \sqrt{(x-a)^2} = \sqrt{(25-4)^2} = 21$

$D_2(25, 18) = \sqrt{(x-a)^2} = \sqrt{(25-18)^2} = 7$

$D_3(25, 34) = \sqrt{(x-a)^2} = \sqrt{(25-34)^2} = 9$

Here 7 is smallest distance so Data point 25 belongs to C_2

30:

$D_1(30, 4) = \sqrt{(x-a)^2} = \sqrt{(30-4)^2} = 26$

$D_2(30, 18) = \sqrt{(x-a)^2} = \sqrt{(30-18)^2} = 12$

$D_3(30, 34) = \sqrt{(x-a)^2} = \sqrt{(30-34)^2} = 4$

Here 4 is smallest distance so Data point 30 belongs to C_3

The updated clusters will be,

$C_1 = \{2, 4, 6, 3\}$,

$C_2 = \{12, 15, 16, 14, 21, 23, 25\}$,

$C_3 = \{31, 38, 35, 30\}$

We can see that there is no difference between previous clusters and these updated clusters, so we will stop the process here.

Finalised clusters –

$C_1 = \{2, 4, 6, 3\}$,

$C_2 = \{12, 15, 16, 14, 21, 23, 25\}$,

$C_3 = \{31, 38, 35, 30\}$

Q8. Apply K-means algorithm on given, data for k=2. Use $C_1(2,4)$ & $C_2(6, 3)$ as initial cluster centres Data : a(2,4), b(3,3), c(5,5), d(6,3), e(4,3), f(6,6)

Ans:

[10M – Dec17]

Number of clusters k = 2

Initial cluster centre for C_1 = (2, 4), C_2 = (6, 3)

We will check distance between data points and all cluster centres. We will use Euclidean Distance formula for finding distance.

Distance [x, a] = $\sqrt{(x-a)^2}$

OR

Distance [(x, y), (a, b)] = $\sqrt{(x-a)^2 + (y-b)^2}$

As given data is in pair, we will use second formula of Euclidean Distance.

Finding Distance between data points and cluster centres.

We will use following notations for calculating Distance:

D_1 = Distance from cluster C_1 centre

D_2 = Distance from cluster C_2 centre

(2, 4):

$D_1[(2, 4), (2, 4)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(2-2)^2 + (4-4)^2} = 0$

$D_2[(2, 4), (6, 3)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(2-6)^2 + (4-3)^2} = 4.13$

Here 0 is smallest distance so Data point (2, 4) belongs to cluster C_1.

As Data point belongs to cluster C_1, we will recalculate the centre of cluster C_1 as following-

Using following formula for finding new centres of cluster =

Centre [(x, y), (a, b)] = $\left(\frac{x+a}{2}, \frac{y+b}{2}\right)$

Here, (x, y) = current data point

(a, b) = old centre of cluster

Updated Centre of cluster C_1 = $\left(\frac{x+a}{2}, \frac{y+b}{2}\right) = \left(\frac{2+2}{2}, \frac{4+4}{2}\right)$ = (2, 4)

(3, 3):

$D_1[(3, 3), (2, 4)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(3-2)^2 + (3-4)^2} = 1.42$

$D_2[(3, 3), (6, 3)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(3-6)^2 + (3-3)^2} = 3$

Here 1.42 is smallest distance so Data point (3, 3) belongs to cluster C_1.

As Data point belongs to cluster C_1, we will recalculate the centre of cluster C_1 as following-

Updated Centre of cluster C_1 = $\left(\frac{x+a}{2}, \frac{y+b}{2}\right) = \left(\frac{3+2}{2}, \frac{3+4}{2}\right)$ = (2.5, 3.5)

(5, 5):

$D_1[(5, 5), (2.5, 3.5)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(5-2.5)^2 + (5-3.5)^2} = 2.92$

$D_2[(5, 5), (6, 3)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(5-6)^2 + (5-3)^2} = 2.45$

Here 2.45 is smallest distance so Data point (5, 5) belongs to cluster C_2

As Data point belongs to cluster C_2, we will recalculate the centre of cluster C_2 as following-

Updated Centre of cluster $C_2 = \left(\frac{x+a}{2}, \frac{y+b}{2}\right) = \left(\frac{5+6}{2}, \frac{5+3}{2}\right) = (5.5, 4)$

(6, 3):

$D_1[(6, 3), (2.5, 3.5)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(6-2.5)^2 + (3-3.5)^2} = 3.54$

$D_2[(6, 3), (5.5, 4)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(6-5.5)^2 + (3-4)^2} = 1.12$

Here 1.12 is smallest distance so Data point (6, 3) belongs to cluster C_2

As Data point belongs to cluster C_2, we will recalculate the centre of cluster C_2 as following-

Updated Centre of cluster $C_2 = \left(\frac{x+a}{2}, \frac{y+b}{2}\right) = \left(\frac{6+5.5}{2}, \frac{3+4}{2}\right) = (5.75, 3.5)$

(4, 3):

$D_1[(4, 3), (2.5, 3.5)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(4-2.5)^2 + (3-3.5)^2} = 1.59$

$D_2[(4, 3), (5.75, 3.5)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(4-5.75)^2 + (3-3.5)^2} = 1.83$

Here 1.59 is smallest distance so Data point (4, 3) belongs to cluster C_1

As Data point belongs to cluster C_1, we will recalculate the centre of cluster C_1 as following-

Updated Centre of cluster $C_1 = \left(\frac{x+a}{2}, \frac{y+b}{2}\right) = \left(\frac{4+2.5}{2}, \frac{3+3.5}{2}\right) = (3.25, 3.25)$

(6, 6):

$D_1[(6, 6), (3.25, 3.25)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(6-3.25)^2 + (6-3.25)^2} = 3.89$

$D_2[(6, 6), (5.75, 3.5)] = \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{(6-5.75)^2 + (6-3.5)^2} = 2.52$

Here 2.52 is smallest distance so Data point (6, 6) belongs to cluster C_1

The final clusters will be,

$C_1 = \{(2, 4), (3, 3), (4, 3), (6, 6)\}$,

$C_2 = \{(5, 5), (6, 3)\}$

Q9. Apply Agglomerative clustering algorithm on given data and draw dendrogram. Show three clusters with its allocated points. Use single link method.

Adjacency Matrix

	a	b	c	d	e	f
a	0	$\sqrt{2}$	$\sqrt{10}$	$\sqrt{17}$	$\sqrt{5}$	$\sqrt{20}$
b	$\sqrt{2}$	0	8	2	1	$\sqrt{18}$
c	$\sqrt{10}$	$\sqrt{8}$	0	$\sqrt{5}$	$\sqrt{5}$	2
d	$\sqrt{17}$	1	$\sqrt{5}$	0	2	3
e	$\sqrt{5}$	1	$\sqrt{5}$	2	0	$\sqrt{13}$
f	$\sqrt{20}$	$\sqrt{18}$	2	3	$\sqrt{13}$	0

Ans: **[10M – May16 & Dec17]**

We have to use single link method to solve this sum. We use following formula for Single link sum,

Min [dist (a), (b)] .... We will choose smallest distance value from two distance values.

We have given Adjacency matrix in which we can see that the upper bound part of diagonal identical to lower bound of diagonal so we can use any part of the matrix.

We will use Lower bound of diagonal as show below.

	a	b	c	d	e	f
a	0					
b	$\sqrt{2}$	0				
c	$\sqrt{10}$	$\sqrt{8}$	0			
d	$\sqrt{17}$	1	$\sqrt{5}$	0		
e	$\sqrt{5}$	1	$\sqrt{5}$	2	0	
f	$\sqrt{20}$	$\sqrt{18}$	2	3	$\sqrt{13}$	0

Now we have to find minimum distance value from above distance matrix. We can see that 1 is smallest distance value but it appears twice in matrix so we can choose any one value from it.

Taking distance value of (b, e) = 1

Now we will draw dendrogram for it.

Now we have to recalculate distance matrix.

We will find distance between clustered points I.e. (b, e) and other remaining points.

- **<u>Distance between (b, e) and a:</u>**

  Min[dist (b, e), a]

  = Min[dist (b, a), (e, a)]

  = Min[$\sqrt{2}, \sqrt{5}$]

  = $\sqrt{2}$ .... as we have to choose smallest value.

- **<u>Distance between (b, e) and c:</u>**

  Min[dist (b, e), c] = Min[dist (b, c), (e, c)] = Min[$\sqrt{8}, \sqrt{5}$] = $\sqrt{5}$

- **<u>Distance between (b, e) and d:</u>**

  Min[dist (b, e), d] = Min[dist (b, d), (e, d)] = Min[$1, 2$] = 1

- **<u>Distance between (b, e) and f:</u>**

  Min[dist (b, e), f] = Min[dist (b, f), (e, f)] = Min[$\sqrt{18}, \sqrt{13}$] = $\sqrt{13}$

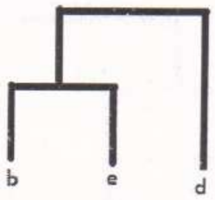
Now we have put these values to update distance matrix:

	A	(b, e)	c	d	f
a	0				
(b, e)	$\sqrt{2}$	0			
c	$\sqrt{10}$	$\sqrt{5}$	0		
d	$\sqrt{17}$	1	$\sqrt{5}$	0	
f	$\sqrt{20}$	$\sqrt{13}$	2	3	0

Now we have to find smallest distance value from updated distance matrix again

Here distance between [(b, e), d] = 1 is smallest distance value in distance matrix.

Now we have to draw dendrogram for these new clustered points.



Now recalculating distance matrix again.

Finding distance between clustered points and other remaining points.

- **Distance between [(b, e), d] and a:**

  Min {dist [(b, e), d], a} = Min {dist [(b, e), a], [d, a]} = Min $[\sqrt{2}, \sqrt{17}] = \sqrt{2}$

- **Distance between [(b, e), d] and c :**

  Min {dist [(b, e), d], c} = Min {dist [(b, e), c], [d, c]} = Min $[\sqrt{5}, \sqrt{5}] = \sqrt{5}$

- **Distance between [(b, e), d] and f :**

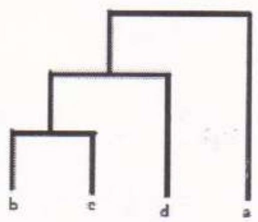
  Min {dist [(b, e), d], f} = Min {dist [(b, e), f], [d, f]} = Min $[\sqrt{13}, 3] = 3$

Now we have to put these distance values in distance matrix to update it.

	a	(b, e, d)	c	f
A	0			
(b, e, d)	$\sqrt{2}$	0		
C	$\sqrt{10}$	$\sqrt{5}$	0	
f	$\sqrt{20}$	3	2	0

Here distance between [(b, e, d), a] = $\sqrt{2}$ is smallest distance value in updated distance matrix.

Drawing dendrogram for new clustered points



Now recalculating distance matrix.

- **Finding distance between [(b, e, d), a] and c:**

  Min {dist [(b, e, d), a], [c]} = Min {dist [(b, e, d), c], [a, c]} = Min $[\sqrt{5}, \sqrt{10}] = \sqrt{5}$

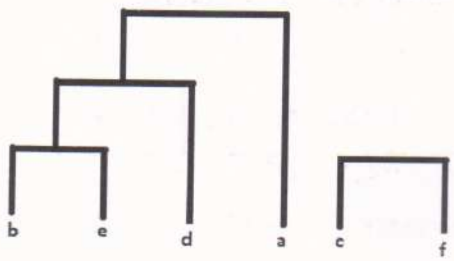
- **Finding distance between [(b, e, d), a] and f:**

  Min {dist [(b, e, d), a], [f]} = Min {dist [(b, e, d), f], [a, f]} = Min $[3, \sqrt{20}] = 3$

Putting these new distance values in distance matrix to update it.

	(b, e, d, a)	c	F
(b, e, d, a)	0		
c	$\sqrt{5}$	0	
f	3	2	0

Here, distance between (c, f) = 2 is smallest distance.



Recalculating distance matrix,

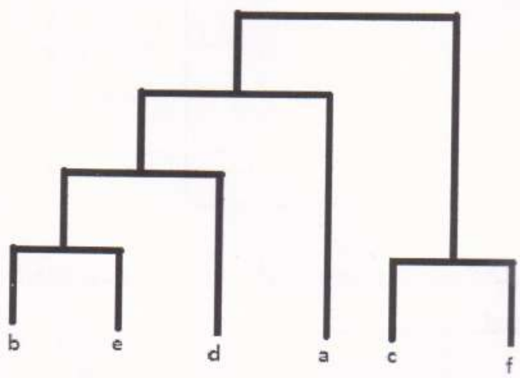
- **<u>Finding distance between (c, f) and (b, e, d, a)</u>**

  Min { dist (c, f), (b, e, d, a)} = Min { dist [c, (b, e, d, a)], [f, (b, e, d, a)]} = Min [$\sqrt{5}$, 3] = $\sqrt{5}$

Putting this value to update distance matrix

	(b, e, d, a)	(c, f)
(b, e, d, a)	0	
(c, f)	$\sqrt{5}$	0

Drawing dendrogram for new cluster,



Q10. For the given set of points identify clusters using complete link and average link using agglomerative clustering.

	A	B
P1	1	1
P2	1.5	1.5
P3	5	5
P4	3	4
P5	4	4
P6	3	3.5

Ans: **[10M – May18]**

We have to solve this sum using complete link method and Average link method.

We use following formula for complete link sum,

Max [dist (a), (b)] .... We will choose biggest distance value from two distance values.

We use following formula for complete link sum,

Avg [dist (a), (b)] = $\frac{1}{2}$ [a + b] .... We will choose biggest distance value from two distance values.

Here, given data is not in distance / adjacency matrix form. So we will convert it to distance / adjacency matrix using Euclidean distance formula which is as following,

Distance [(x, y), (a, b)] = $\sqrt{(x-a)^2 + (y-b)^2}$

Now finding distance between P1 and P2, Distance [P1, P2] =

distance[(1, 1), (1.5, 1.5)] = $\sqrt{(x-a)^2 + (y-b)^2}$ = $\sqrt{(1-1.5)^2 + (1-1.5)^2}$ = 0.71

As per above step we will find distance between other points as well.

Distance [P1, P3] = 5.66

Distance [P1, P4] = 3.61

Distance [P1, P5] = 4.25

Distance [P1, P6] = 3.21

Distance [P2, P3] = 4.95

Distance [P2, P4] = 2.92

Distance [P2, P5] = 3.54

Distance [P2, P6] = 2.5

Distance [P3, P4] = 2.24

Distance [P3, P5] = 1.42

Distance [P3, P6] = 2.5

Distance [P4, P5] = 1

Distance [P4, P6] = 0.5

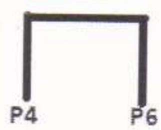
Distance [P5, P6] = 1.12

Putting these values in lower bound of diagonal of distance / adjacency matrix,

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.71	0				
P3	5.66	4.95	0			
P4	3.61	2.92	2.24	0		
P5	4.25	3.54	1.42	1	0	
P6	3.21	2.5	2.5	0.5	1.12	0

COMPLETE LINK METHOD:

Here, distance between P4 and P6 is 0.5 which is smallest distance in matrix.



Recalculating distance matrix:

- Distance between (P4, P6) and P1

  = Max[dist (P4, P6), P1]

  = Max[dist (P4, P1), (P6, P1)]

  = Max[3.61, 3.21]

  = 3.61 .... We take biggest distance value as it is Complete link method

- Distance between (P4, P6) and P2

  = Max[dist (P4, P6), P2]

  = Max[dist (P4, P2), (P6, P2)]

  = Max[2.92, 2.5]

  = 2.92

- Distance between (P4, P6) and P3

  = Max[dist (P4, P6), P3]

  = Max[dist (P4, P3), (P6, P3)]

  = Max[2.24, 2.5]

  = 2.5

- Distance between (P4, P6) and P5

  = Max[dist (P4, P6), P5]

  = Max[dist (P4, P5), (P6, P5)]

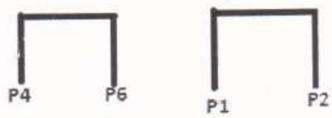
  = Max[1, 1.12]

  = 1.12

Updating distance matrix:

	P1	P2	P3	(P4, P6)	P5
P1	0				
P2	0.71	0			
P3	5.66	4.95	0		
(P4, P6)	3.61	2.92	2.5	0	
P5	4.25	3.54	1.42	1.12	0

Here (P1, P2) = 0.71 is smallest distance in matrix



Recalculating distance matrix:

- Distance between (P1, P2) and P3

  = Max[dist (P1, P2), P3]

  = Max[dist (P1, P3), (P2, P3)]

  = Max[5.66, 4.95]

  = 5.66

- Distance between (P1, P2) and (P4, P6)

  = Max[dist (P1, P2), (P4, P6)]

  = Max[dist {P1, (P4, P6)}, {P2, (P4, P6)}]

  = Max[3.61, 2.92]

  = 3.61

- Distance between (P1, P2) and P5

  = Max[dist (P1, P2), P5]

  = Max[dist (P1, P5), (P2, P5)]

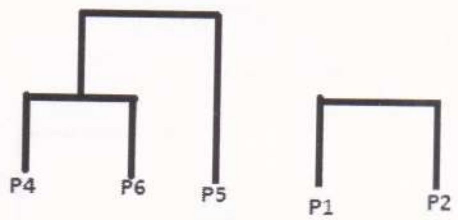
  = Max[4.25, 3.54]

  = 4.25

Updating distance matrix using above values,

	(P1, P2)	P3	(P4, P6)	P5
(P1, P2)	0			
P3	5.66	0		
(P4, P6)	3.61	2.5	0	
P5	4.25	1.42	1.12	0

Here [(P4, P6), P5] = 1.12 is smallest distance in matrix,



Recalculating distance matrix

- Distance between {(P4, P6), P5} and (P1, P2)

  = Max[dist {(P4, P6), P5}, (P1, P2)]

  = Max[dist {(P4, P6), (P1, P2)}, {P5, (P1, P2)}]

  = Max[3.61, 4.25]

  = 4.25

- Distance between {(P4, P6), P5} and P3

  = Max[dist {(P4, P6), P5}, P3]

  = Max[dist {(P4, P6), P3}, {P5, P3}]

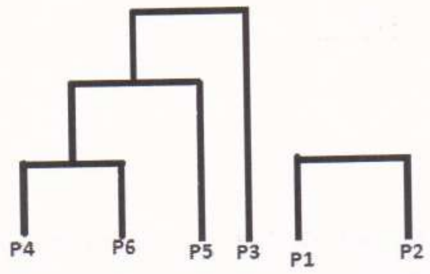
  = Max[2.5, 1.42]

  = 2.5

Updating distance matrix

	(P1, P2)	P3	(P4, P6, P5)
(P1, P2)	0		
P3	5.66	0	
(P4, P6, P5)	4.25	2.5	0

Here, [(P4, P6, P5), P3] = 2.5 is smallest distance in matrix,



Recalculating distance matrix:

- Distance between {(P4, P6, P5), P3} and {P1, P2}

  = Max[dist {(P4, P6, P5), P3}, {P1, P2}]

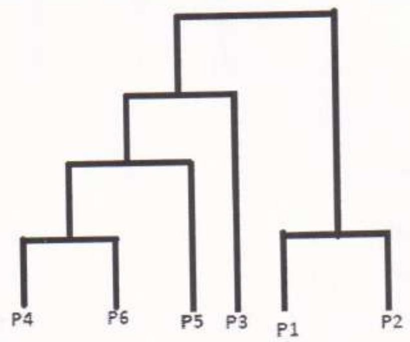
  = Max[dist{(P4, P6, P5), (P1, P2)}, {P3, (P1, P2)}]

  = Max[4.25, 5.66]

  = 5.66

Updating distance matrix

	(P1, P2)	(P4, P6, P5, P3)
(P1, P2)	0	
(P4, P6, P5, P3)	5.66	0



AVERAGE LINK METHOD:

Original Distance Matrix:

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.71	0				
P3	5.66	4.95	0			
P4	3.61	2.92	2.24	0		
P5	4.25	3.54	1.42	1	0	
P6	3.21	2.5	2.5	0.5	1.12	0

Here, distance between P4 and P6 is 0.5 which is smallest distance in matrix.

Recalculating distance matrix,

- Distance between (P4, P6) and P1

  = Avg[dist (P4, P6), P1]

  = Avg[dist (P4, P1), (P6, P1)]

  = $\frac{1}{2}$[3.61 + 3.21]

  = 3.41 .... We take average distance value as it is Average link method

- Distance between (P4, P6) and P2

  = Avg[dist (P4, P6), P2]

  = Avg[dist (P4, P2), (P6, P2)]

  = $\frac{1}{2}$ [2.92 + 2.5]

  = 2.71

- Distance between (P4, P6) and P3

  = Avg[dist (P4, P6), P3]

  = Avg[dist (P4, P3), (P6, P3)]

  = $\frac{1}{2}$ [2.24 + 2.5]

  = 2.37

- Distance between (P4, P6) and P5

  = Avg[dist (P4, P6), P5]

  = Avg[dist (P4, P5), (P6, P5)]

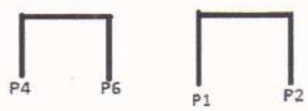
  = $\frac{1}{2}$ [1 + 1.12]

  = 1.06

Updating distance matrix:

	P1	P2	P3	(P4, P6)	P5
P1	0				
P2	0.71	0			
P3	5.66	4.95	0		
(P4, P6)	3.41	2.71	2.37	0	
P5	4.25	3.54	1.42	1.06	0

Here (P1, P2) = 0.71 is smallest distance in matrix



Recalculating distance matrix:

- Distance between (P1, P2) and P3

  = Avg[dist (P1, P2), P3]

  = Avg[dist (P1, P3), (P2, P3)]

  = $\frac{1}{2}$[5.66 + 4.95]

  = 5.31

- Distance between (P1, P2) and (P4, P6)

  = Avg[dist (P1, P2), (P4, P6)]

  = Avg[dist {P1, (P4, P6)}, {P2, (P4, P6)}]

  $= \frac{1}{2}[3.41 + 2.71]$

  = 3.06

- Distance between (P1, P2) and P5

  = Avg[dist (P1, P2), P5]

  = Avg[dist (P1, P5), (P2, P5)]

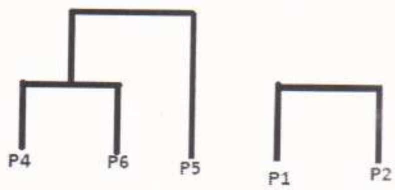
  $= \frac{1}{2}[4.25 + 3.54]$

  = 3.90

Updating distance matrix using above values,

	(P1, P2)	P3	(P4, P6)	P5
(P1, P2)	0			
P3	5.31	0		
(P4, P6)	3.06	2.5	0	
P5	3.90	1.42	1.12	0

Here [(P4, P6), P5] = 1.12 is smallest distance in matrix,



Recalculating distance matrix:

- Distance between {(P4, P6), P5} and (P1, P2)

  = Avg[dist {(P4, P6), P5}, (P1, P2)]

  = Avg[dist {(P4, P6), (P1, P2)}, {P5, (P1, P2)}]

  $= \frac{1}{2}[3.06 + 3.90]$

  = 3.48

- Distance between {(P4, P6), P5} and P3

  = Avg[dist {(P4, P6), P5}, P3]

  = Avg[dist {(P4, P6), P3}, {P5, P3}]

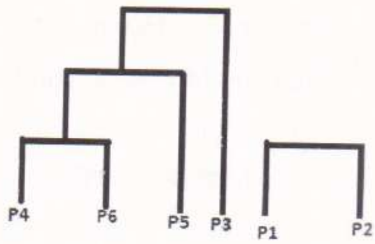
  $= \frac{1}{2}[2.5 + 1.42]$

  = 1.96

Updating distance matrix

	(P1, P2)	P3	(P4, P6, P5)
(P1, P2)	0		
P3	5.66	0	
(P4, P6, P5)	3.48	1.96	0

Here, [(P4, P6, P5), P3] = 1.96 is smallest distance in matrix,



Recalculating distance matrix:

- Distance between {(P4, P6, P5), P3} and {P1, P2}

  = Avg[dist {(P4, P6, P5), P3}, {P1, P2}]

  = Avg[dist{(P4, P6, P5), (P1, P2)}, {P3, (P1, P2)}]

  = $\frac{1}{2}$ [3.48 + 5.66]

  = 4.57

Updating distance matrix

	(P1, P2)	(P4, P6, P5, P3)
(P1, P2)	0	
(P4, P6, P5, P3)	4.57	0