HIERARCHIAL AGGLOMERATIVE CLUSTERING

SINGLE LINK

Dataset:

Features	X	Y
D1	1	3
D2	2	4
D3	3	1
D4	4	2
D5	5	1
D6	1	4
D7	2	1
D8	3	2
D9	4	5
D10	5	4

Distance $(D1, D2) = \sqrt{(1-2)^2 + (3-4)^2}$

$= \sqrt{(-1)^2 + (-1)^2} = \sqrt{2} = 1.4$

Distance $(D_1, D_3) = \sqrt{(1-3)^2 + (3-1)^2} = \sqrt{(-2)^2 + (2)^2} = \sqrt{4+4}$

$= \sqrt{8} = 2.8$

Distance $(D_1, D4) = \sqrt{(1-4)^2 + (3-2)^2} = \sqrt{(-3)^2 + (1)^2} = \sqrt{9+1}$

$= \sqrt{10} = 3.2$

Distance $(D1, D5) = \sqrt{(1-5)^2 + (3-1)^2} = \sqrt{(-4)^2 + (2)^2} = \sqrt{16+4}$

$= \sqrt{20} = 4.5$

Distance (D1, D6) $= \sqrt{(1-1)^2 + (3-4)^2}$

$= \sqrt{(-1)^2} = 1$

Distance (D2, D3) $= \sqrt{(2-3)^2 + (4-1)^2}$

$= \sqrt{(-1)^2 + (3)^2} = \sqrt{1+9} = \sqrt{10} = 3.2$

Distance (D2, D4) $= \sqrt{(2-4)^2 + (4-2)^2}$

$= \sqrt{(-2)^2 + (2)^2} = \sqrt{4+4} = \sqrt{8}$

$= 2.8$

Distance (D2, D5) $= \sqrt{(2-5)^2 + (4-1)^2}$

$= \sqrt{(-3)^2 + (3)^2} = \sqrt{9+9} = \sqrt{18}$

$= 4.2$

Distance (D2, D6) $= \sqrt{(2-1)^2 + (4-4)^2}$

$= \sqrt{(1)^2 + (0)} = \sqrt{1} = 1$

Distance (D3, D4) $= \sqrt{(3-4)^2 + (1-2)^2} = \sqrt{(-1)^2 + (-1)^2}$

$= \sqrt{2} = 1.4$

Distance (D3, D5) $= \sqrt{(3-5)^2 + (1-1)^2}$

$= \sqrt{(-2)^2 + 0} = \sqrt{4} = 2$

Distance (D3, D6) $= \sqrt{(3-1)^2 + (1-4)^2} = \sqrt{(2)^2 + (-3)^2}$

$= \sqrt{4+9} = \sqrt{13} = 3.6$

Distance (D3, D7) $= \sqrt{(3-2)^2 + (1-1)^2}$

$= \sqrt{(1)^2 + (0)} = 1$

Distance (D3, D8) $= \sqrt{(3-3) + (1-2)^2} = \sqrt{0 + (-1)^2}$

$= 1$

Distance (D3, D9) $= \sqrt{(3-4)^2+(1-5)^2} = \sqrt{(-1)^2+(-4)^2}$

$= \sqrt{17} = 4.1$

Distance (D3, D10) $= \sqrt{(3-5)^2+(1-4)^2} = \sqrt{(-2)^2+(-3)^2}$

$= \sqrt{4+9} = \sqrt{13} = 3.6$

Distance (D2, D7) $= \sqrt{(2-2)^2+(4-1)^2}$

$= \sqrt{0+9} = 3$

Distance (D2, D8) $= \sqrt{(2-3)^2+(4-2)^2}$

$= \sqrt{(-1)^2+(2)^2} = \sqrt{5} = 2.2$

Distance (D2, D9) $= \sqrt{(2-4)^2+(4-5)^2}$

$= \sqrt{(-2)^2+(-1)^2} = \sqrt{5} = 2.2$

Distance (D2, D10) $= \sqrt{(2-5)^2+(4-4)^2}$

$= \sqrt{(-3)^2+0} = \sqrt{9} = 3$

Distance (D1, D7) $= \sqrt{(1-2)^2+(3-1)^2}$

$= \sqrt{(-1)^2+(2)^2} = \sqrt{5} = 2.2$

Distance (D1, D8) $= \sqrt{(1-3)^2+(3-2)^2}$

$= \sqrt{(-2)^2+(1)^2} = \sqrt{5} = 2.2$

Distance (D1, D9) $= \sqrt{(1-4)^2+(3-5)^2}$

$= \sqrt{(-3)^2+(-2)^2} = \sqrt{13} = 3.6$

Distance (D1, D10) $= \sqrt{(1-5)^2+(3-4)^2}$

$= \sqrt{(-4)^2+(-1)^2} = \sqrt{17} = 4.1$

Distance (D4, D5) $= \sqrt{(4-5)^2+(2-1)^2} = \sqrt{(-1)^2+(1)^2}$

$= \sqrt{2} = 1.4$

Pg #03

Distance (D4, D6) $= \sqrt{(4-1)^2 + (2-4)^2}$

$= \sqrt{(3)^2 + (-2)^2} = \sqrt{13} = 3.6$

Distance (D4, D7) $= \sqrt{(4-2)^2 + (2-1)^2}$

$= \sqrt{(2)^2 + (1)^2} = \sqrt{4+1} = \sqrt{5} = 2.2$

Distance (D4, D8) $= \sqrt{(4-3)^2 + (2-2)^2}$

$= \sqrt{1} = 1$

Distance (D4, D9) $= \sqrt{(4-4)^2 + (2-5)^2}$

$= \sqrt{(-3)^2} = \sqrt{9} = 3$

Distance (D4, D10) $= \sqrt{(4-5)^2 + (2-4)^2}$

$= \sqrt{(-1)^2 + (-2)^2} = \sqrt{5}$

$= 2.2$

Distance (D5, D6) $= \sqrt{(5-1)^2 + (1-4)^2}$

$= \sqrt{(4)^2 + (-3)^2} = \sqrt{16+9} = \sqrt{25} = 5$

Distance (D5, D7) $= \sqrt{(5-2)^2 + (1-1)^2}$

$= \sqrt{(3)^2} = \sqrt{9} = 3$

Distance (D5, D8) $= \sqrt{(5-3)^2 + (1-2)^2}$

$= \sqrt{(2)^2 + (-1)^2} = \sqrt{4+1} = \sqrt{5} = 2.2$

Distance (D5, D9) $= \sqrt{(5-4)^2 + (1-5)^2}$

$= \sqrt{(1)^2 + (-4)^2} = \sqrt{17} = 4.1$

Distance (D5, D10) $= \sqrt{(5-5)^2 + (1-4)^2}$

$= \sqrt{0 + (-3)^2} = \sqrt{9}$

$= 3$

Distance(D6, D7) $= \sqrt{(1-2)^2 + (4-1)^2} = \sqrt{(-1)^2 + (3)^2}$

$= \sqrt{10} = 3.2$

Distance(D6, D8) $= \sqrt{(1-3)^2 + (4-2)^2} = \sqrt{(-2)^2 + (2)^2}$

$= \sqrt{8} = 2.8$

Distance(D6, D9) $= \sqrt{(1-4)^2 + (4-5)^2} = \sqrt{(-3)^2 + (-1)^2}$

$= \sqrt{10} = 3.2$

Distance(D6, D10) $= \sqrt{(1-5)^2 + (4-4)^2} = \sqrt{(-4)^2 + 0}$

$= \sqrt{16} = 4$

Distance(D7, D8) $= \sqrt{(2-3)^2 + (1-2)^2} = \sqrt{(-1)^2 + (-1)^2}$

$= \sqrt{2} = 1.4$

Distance(D7, D9) $= \sqrt{(2-4)^2 + (1-5)^2} = \sqrt{(2)^2 + (-4)^2}$

$= 4.5$

Distance(D7, D10) $= \sqrt{(2-5)^2 + (1-4)^2} = \sqrt{(-3)^2 + (-3)^2}$

$= \sqrt{9+9} = \sqrt{18} = 4.2$

Distance(D8, D9) $= \sqrt{(3-4)^2 + (2-5)^2} = \sqrt{(-1)^2 + (-3)^2}$

$= \sqrt{10} = 3.2$

Distance(D8, D10) $= \sqrt{(3-5)^2 + (2-4)^2} = \sqrt{(-2)^2 + (-2)^2}$

$= \sqrt{8} = 2.8$

Distance(D9, D10) $= \sqrt{(4-5)^2 + (5-4)^2}$

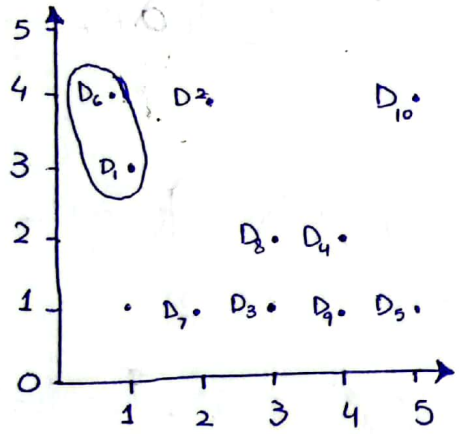
$= \sqrt{(-1)^2 + (1)^2}$

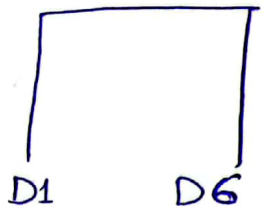
$= \sqrt{1+1} = \sqrt{2}$

$= 1.4$

Distance Matrix

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
D1	0									
D2	1.4	0								
D3	2.8	3.2	0							
D4	3.2	2.8	1.4	0						
D5	4.5	4.2	2	1.4	0					
D6	1	1	3.6	3.6	5	0				
D7	2.2	2	1	2.2	3	3.2	0			
D8	2.2	2.2	1	1	2.2	2.8	1.4	0		
D9	3.6	2.2	4.1	3	4.1	3.2	4.5	3.2	0	
D10	4.1	3	3.6	2.2	3	4	4.2	2.8	1.4	0





Let's find distance of (D1, D6) from ~~each~~ rest of the features.

Distance $[(D_1, D_6), D_2]$ = ~~Min~~ Min $[dis\ (D1, D2)(D6, D2)]$

$= \min(1.4, 1)$

$= 1$

Distance $[(D1, D6), D3]$ = Min $[dis\ (D1, D3)\ (D6, D3)]$

$= \min(2.8, 3.6)$

$= 2.8$

$$\text{Distance}[(D1,D6),D4] = \min(\text{dis}(D1,D4)(D6,D4))$$
$$= \min(3.4, 3.6)$$
$$= 3.4$$

$$\text{Distance}[(D1,D6),D5] = \min(\text{dis}(D1,D5)(D6,D5))$$
$$= \min[4.5, 5]$$
$$= 4.5$$

$$\text{Distance}[(D1,D6),D6] = \min[(D1,D6)(D6,D6)]$$
$$= \min(1, 0)$$
$$= 0$$

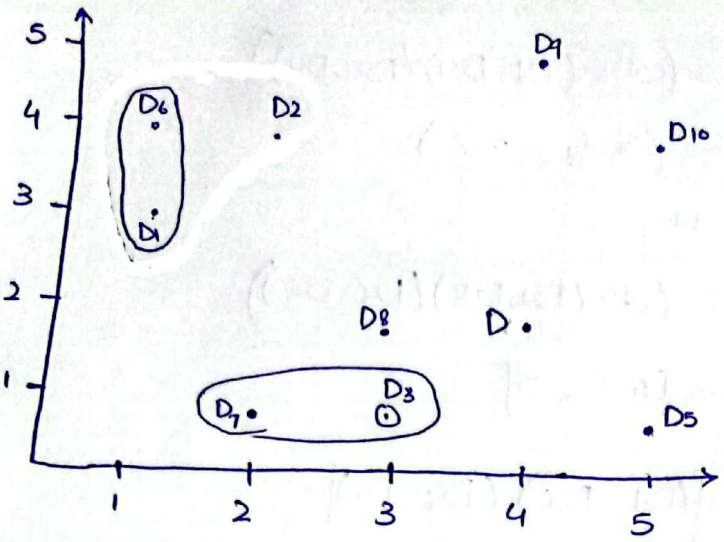
$$\text{Distance}[(D1,D6),D7] = \min[(D1,D7),(D6,D7)]$$
$$= \min(2.2, 3.2)$$
$$= 2.2$$

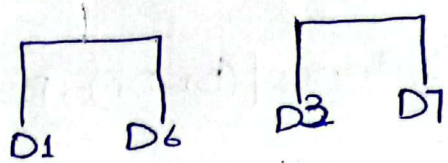
$$\text{Distance}[(D1,D6),D8] = \min[(D1,D8)(D6,D8)]$$
$$= \min[2.2, 2.8]$$
$$= 2.2$$

$$\text{Distance}[(D1,D6),D9] = \min[(D1,D9)(D6,D9)]$$
$$= \min[3.6, 3.2]$$
$$= 3.2$$

$$\text{Distance}[(D1,D6),D10] = \min[(D1,D10),(D6,D10)]$$
$$= \min(4.1, 4)$$
$$= 4$$

	D1,D6	D2	D3	D4	D5	D6	D7	D8	D9	D10
D1,D6	0									
D2	1	0								
D3	2.8	3.2	0							
D4	3.4	2.8	1.4	0						
D5	4.5	4.2	2	1.4	0					
~~D6~~	~~0~~	~~1~~	~~3.6~~	~~3.6~~	~~5~~	~~4~~				
D7	2.2	2	1	1	3	3.2	0			
D8	2.2	2.2	1	1	2.2	2.8	1.4	0		
D9	3.2	2.2	4.1	4.1	4.1	3.2	4.5	3.2	0	
D10	4	3	3.6	3.6	3	4	4.2	2.8	1.4	0





Let's find distance of (D3, D7) from rest of the features.

Distance[(D3,D7)(D1,D6)] = min(D3(D1,D6), D7(D1,D6)] = min[2.8, 2.2] = 2.2

Distance [(D3, D7), D4] = min [dis (D3, D4)(D7, D4)]

= min (1.4, 2.2)

= 1.4

Distance [(D3, D7), D5] = min [(D3, D5) (D7, D5]

= min (3.6, 3)

= 3

~~Distance[(D3, D7) D6] = min[dis(D3, D6), (D7, D6)]~~

~~= min( )~~

Distance [(D3, D7), D8] = min [dis(D3, D8) (D7, D8)]

= min (1, 1.4)

= 1

Distance [(D3, D7), D9] = min [dis (D3, D9) (D7, D9)]

= min (4.1, 4.5)

= 4.1

Distance [(D3, D7), D10] = min [dis (D3, D10) (D7, D10]

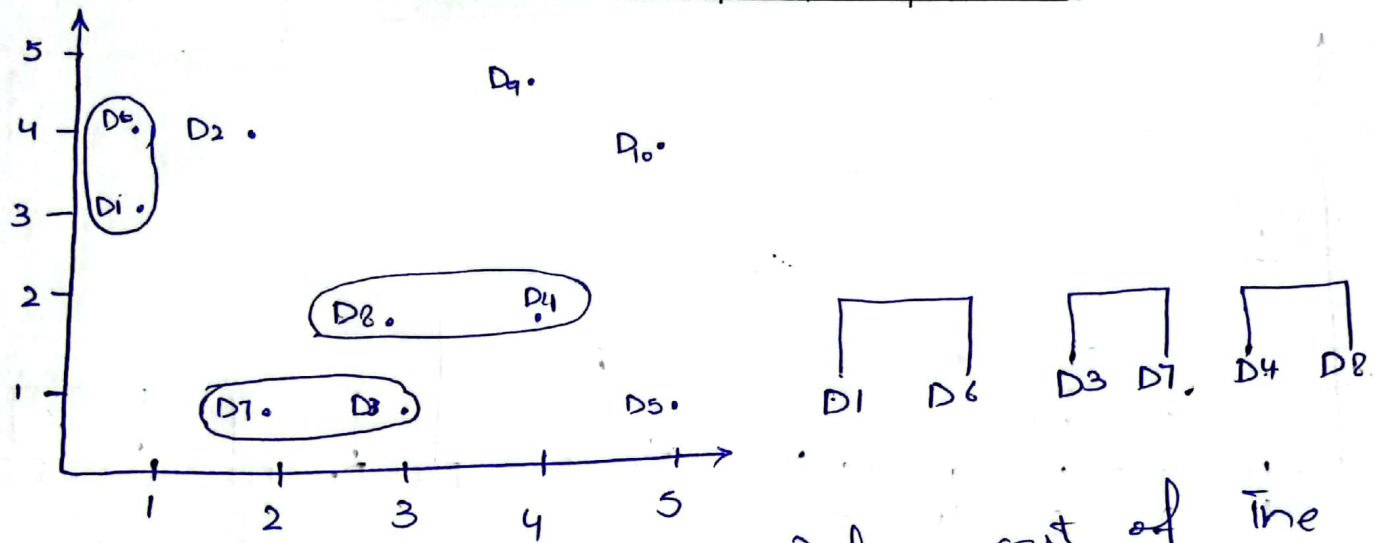
= min (3.6, 4.2)

= 3.6

Dis [(D3, D7), D2] = min (dis (D3, D2) (D7, D2)]

= min (3.2, 2) = 2

	D1,D6	D2	D3,D7	D4	D5	D8	D9	D10
D1,D6	0							
D2	1	0						
D3,D7	2.2	2	0					
D4	3.4	2.8	1.4	0				
D5	4.5	4.2	3	1.4	0			
D8	2.2	2.2	1	1	2.2	0		
D9	3.2	2.2	4.1	4.1	4.1	3.2	0	
D10	4	3	3.6	3.6	3	4	1.4	0



Let's find distance of (D4, D8) from rest of the features.

$Dis[(D4, D8), D5] = min[(D4, D5)(D8, D5)] = min(1.4, 2.2) = 1.4$

$Dis[(D4, D8), D9] = min[(D4, D9), (D8, D9)] = min(4.1, 3.2) = 3.2$

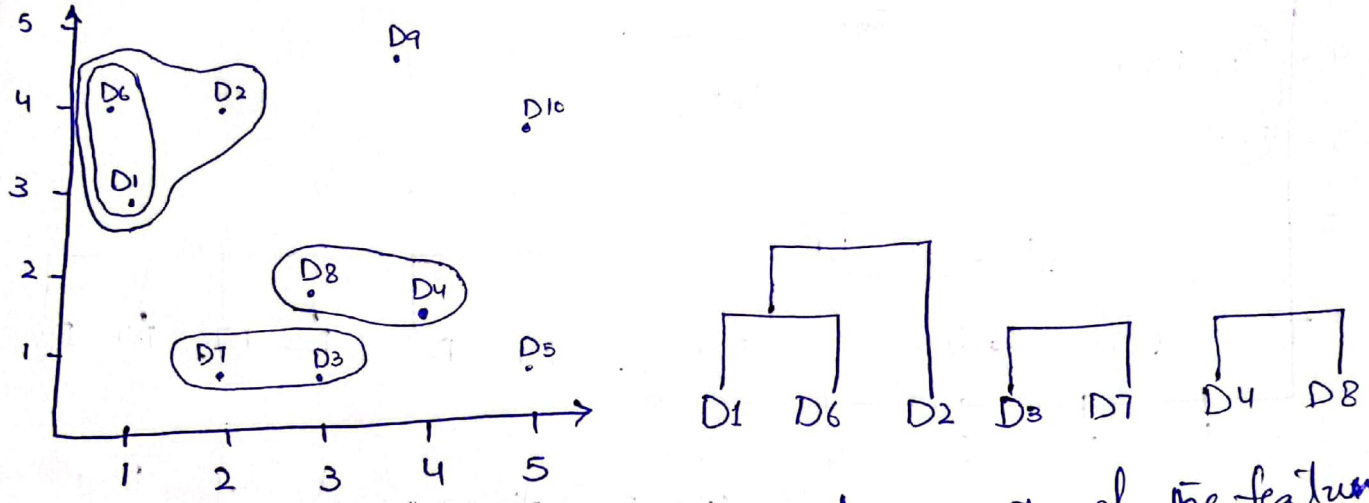
$Dis[(D4, D8), D10] = min[(D4, D10)(D8, D10] = min(3.6, 4) = 3.6$

$Dis[(D4, D8)(D1, D6)] = min[D4(D1, D6), D8(D1, D6)] = min(3.4, 2.2) = 2.2$

$Dis[(D4, D8)(D2)] = min[(D4, D2)(D8, D2)] = min(2.8, 2.2) = 2.2$

$Dis[(D4, D8)(D3, D7)] = min[D4(D3, D7), D8(D3, D7)] = min(1.4, 1) = 1$

	D1,D6	D2	D3,D7	D4,D8	D5	D9	D10
D1,D6	0						
D2	1	0					
D3,D7	2.2	2	0				
D4,D8	2.2	2.2	1	0			
D5	4.5	4.2	3	1.4	0		
D9	3.2	2.2	4.1	3.2	4.1	0	
D10	4	3	3.6	3.6	3	1.4	0



Let's find distance of [(D1, D6), D2] from resl of the features

→ Dis [(D1, D6, D2) (D ; D7)] =

= Dis [((D1, D6), (D3, D7)), (D2)(D3, D7)] = min (2.2, 2) = 2

→ Dis [((D1, D6, D2)(D4, D8)] = Min [((D1, D6)(D4, D8)), (D2)(D4, D8)]

= min (2.2, 2.2) = 2.2

→ Dis [((D1, D6), D2)(D5)] = Min [((D1, D6) D5), D2 (D5)] = min (4.5, 4.2) = 4.2

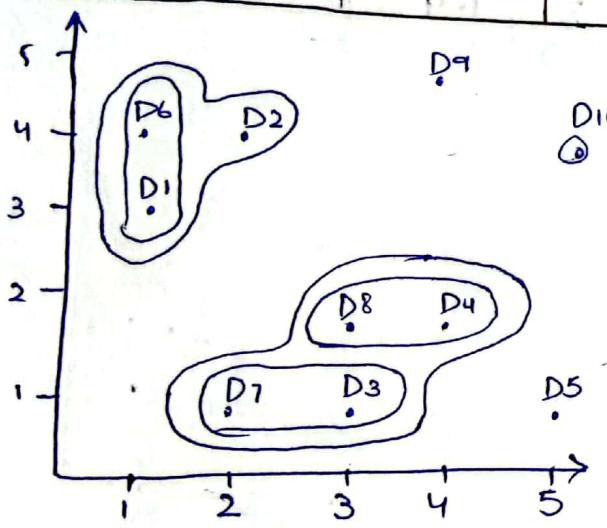
→ Dis [((D1, D6) D2) (D9)] = min [((D1, D6)(D9)), (D2)(D9)] = min (3.2, 2.2) = 2.2

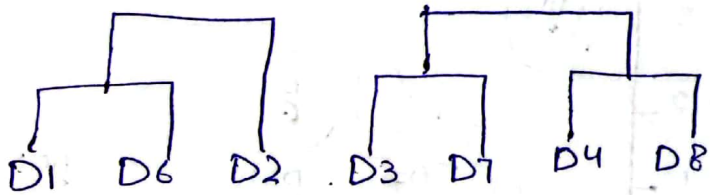
→ Dis [((D1, D6) D2), D10)] = min [((D1, D6)(D10)), D2 (D10)]

= min (4, 3)

= 3

	(D1,D6)D2	D3,D7	D4,D8	D5	D9	D10
(D1,D6)D2	0					
D3,D7	2	0				
D4,D8	2.2	1	0			
D5	4.2	3	1.4	0		
D9	2.2	4.1	3.2	4.1	0	
D10	3	3.6	3.6	3	1.4	0





Let's find the distance of [(D3,D7)(D4,D8)] from rest of the features.

$$Dis\left[((D3\ D7)(D4,D8)),((D1,D6),D2)\right] = \min\left[\left[(D3,D7)(D1,D6)\right]_{D2},\left[(D4,D8)^{(D1,D6)}D2\right]\right]$$

$$= \min(2, 2.2) = 2$$

$$Dis\left[((D3,D7)(D4,D8)),D5\right] = \min\left[(D3,D7)D5,(D4,D8)D5\right]$$

$$= \min(3, 1.4) = 1.4$$

$$Dis\left[((D3,D7)(D4,D8))D9\right] = \min\left[(D3,D7)D9,(D4,D8)D9\right]$$

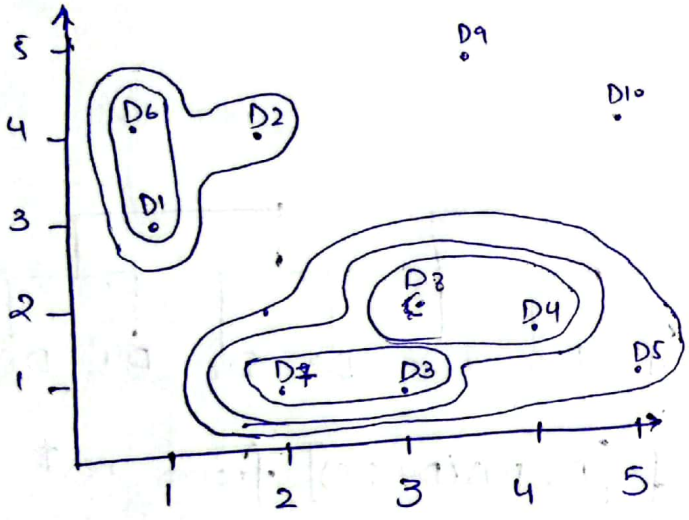
$$= \min(4.1, 3.2)$$

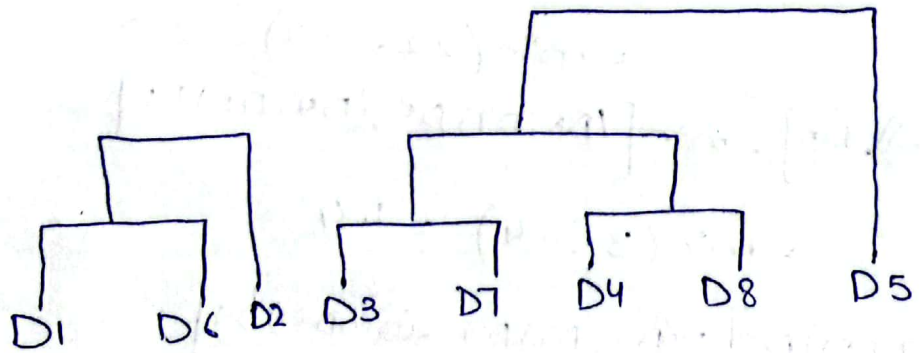
$$= 3.2$$

$$\rightarrow Dis[((D3,D7)(D4,D8)), D10] = \min[(D3,D7)D10 , (D4,D8)D10]$$

$$= \min[3.6, 3.6] = 3.6$$

	(D1, D6)D2	(D3, D7)(D4, D8)	D5	D9	D10
(D1, D6)D2	0				
(D3, D7)(D4, D8)	2.2	0			
D5	4.2	1.4	0		
D9	2.2	3.2	4.1	0	
D10	3.	3.6	3	1.4	0





Let's find distance $[((D3, D7)(D4, D8))D5]$ from rest of the features.

$$\rightarrow Dis\left[((D3, D7)(D4, D8))D5), (D1, D6)D2\right] = \min\left[\left[((D3, D7)(D4, D8))(D1, D6)D2\right], D5((D1, D6)D2)\right]$$

$$= \min[\ 2.2\ ,\ 4.2]$$

$$= 2.2$$

$$\rightarrow Dis\left[((D3, D7)(D4, D8))D5), D9\right] = \min\left[(D3, D7)(D4, D8)D9, D5(D9)\right]$$

$$= \min[\ 3.2\ ,\ 4.1\ ] =$$

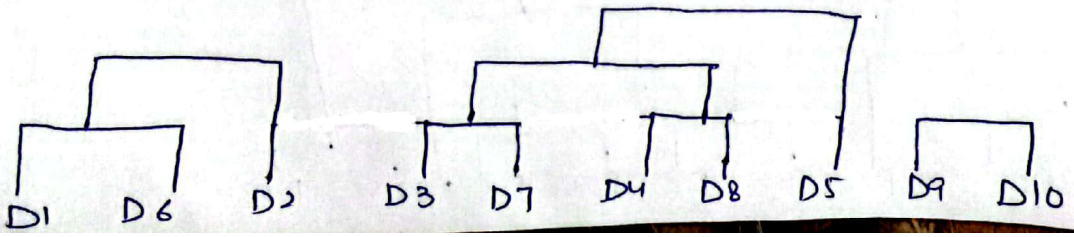
$$= 3.2$$

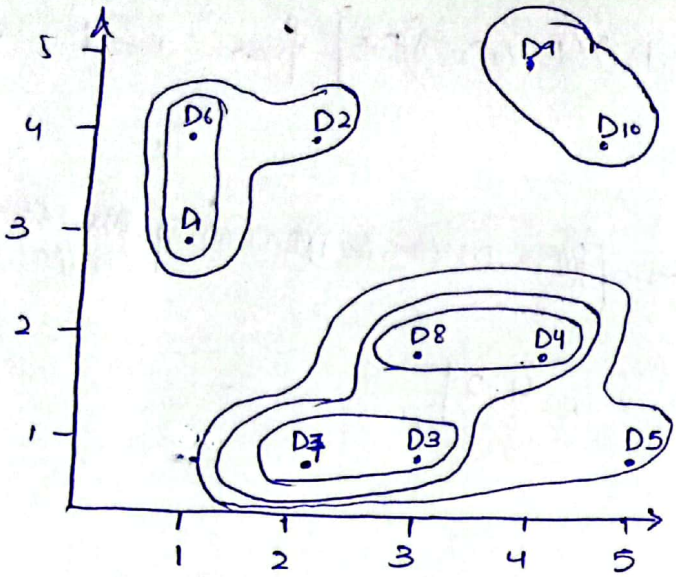
$$\rightarrow Dis\left[(D3, D7)(D4, D8)D5), D10\right] = \min\left[(D3, D7)(D4, D8)D10\ , D5(D10)\right]$$

$$= \min(\ 3.6\ ,\ 3\ )$$

$$= 3$$

	$(D1, D6)D2$	$[(D3, D7)(D4, D8)]D5$	$D9$	$D10$
$(D1, D6)D2$	0			
$(D3, D7)(D4, D8)D5$	2.2	0		
$D9$	2.2	3.2	0	
$D10$	3	3	1.4	0





→ Distance $[(D9, D10), ((D1, D6)D2)]$

$= \min[[(D9)(D10)], ((D10)((D1, D6), D2))]$

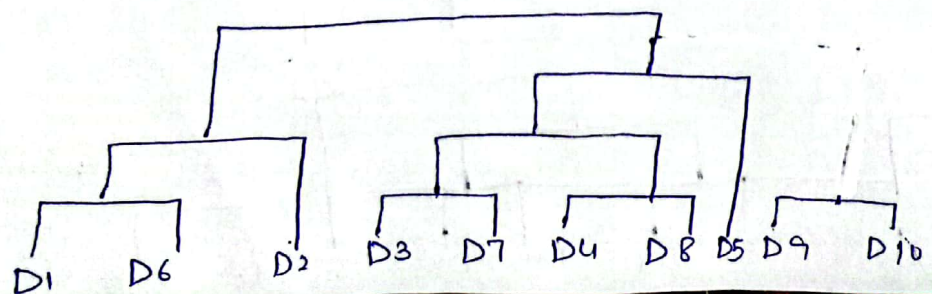
$= \min[2.2, 3] = 2.2$

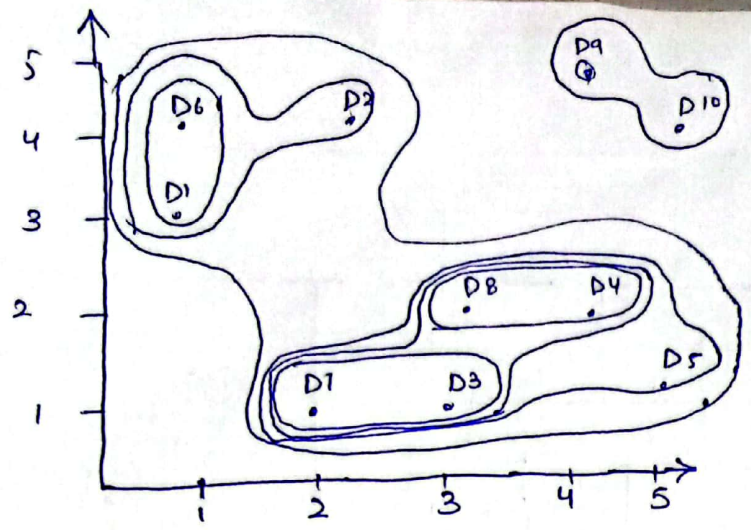
→ Distance: $[(D9, D10), [((D3, D7)(D4, D8))D5]]$

$= \min[D9[((D3, D7)(D4, D8))D5], D10[((D3, D7)(D4, D8))D5)]$

$= \min[3.2, 3] = 3$

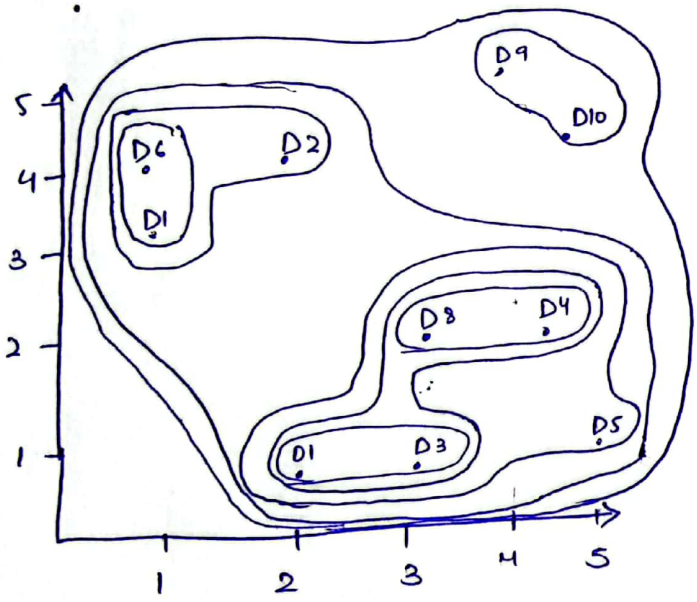
	$(D1, D6), D2$	$((D3, D7)(D4, D8))D5$	$D9, D10$
$(D1, D6)D2$	0		
$((D3, D7)(D4, D8))D5$	2.2	0	
$(D9, D10)$	2:2	3	0

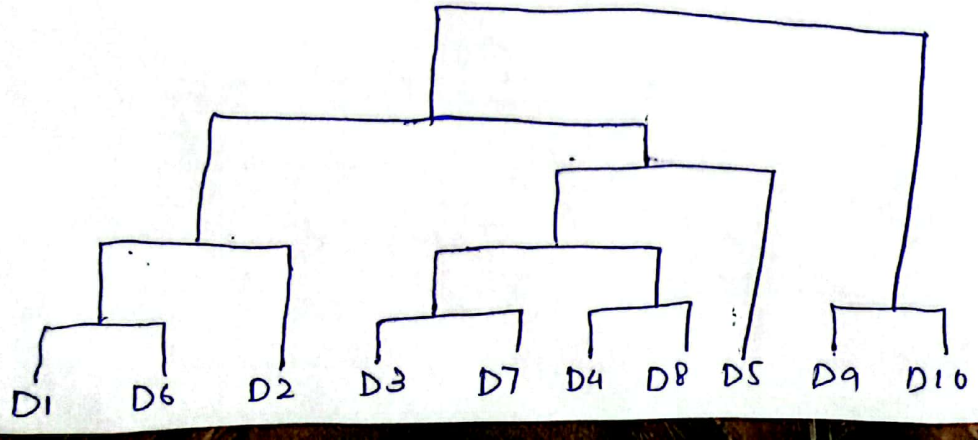




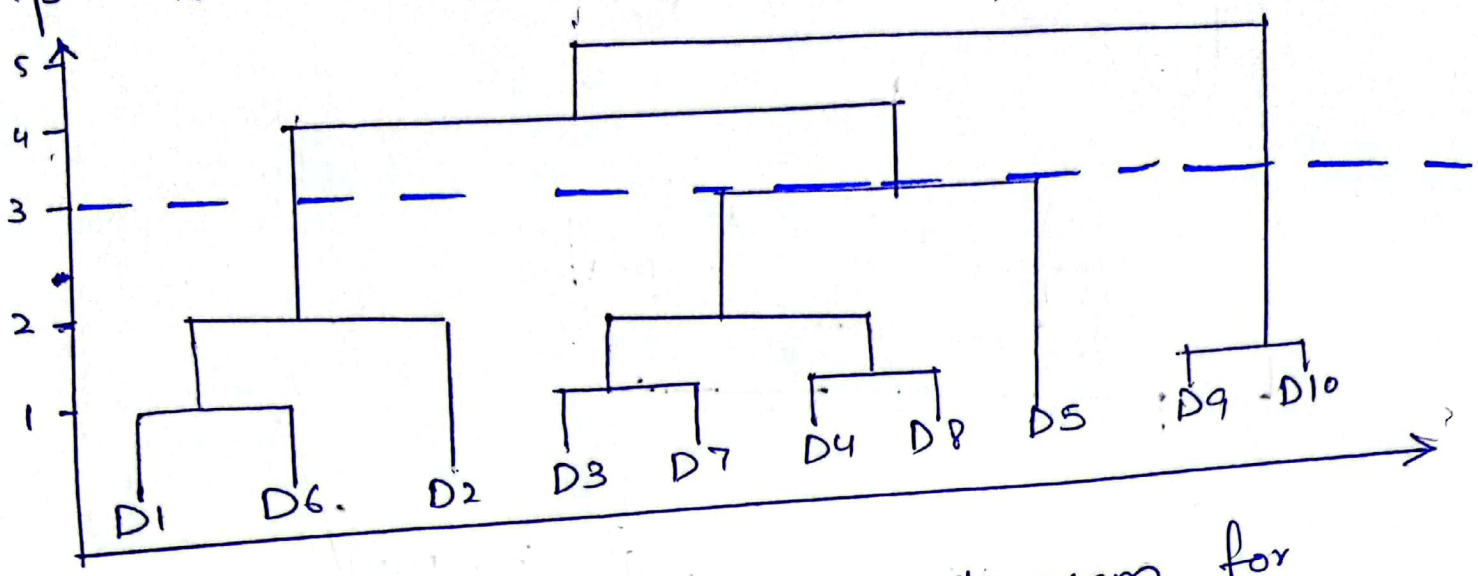
→ Distance $\left[\left[\left[\left((D1, D6)\, D2\right)\left(\left((D3, D7)(D4, D8)\right) D5\right)\right]\right],\ (D9; D10)\right]$

=n





As The Threshold = 3



So, there are '4' clusters in dendrogram for threshold = 3