Hierarchical Clustering using Single Link technique.

Starting with the Individual clusters.

Given Points – A(2,2), B(2,6), C(3,7), D(5,2), E(5,5), F(5,8), G(6,6), H(7,3), I(8,4), J(10,6), K(12,8)

Initial Clusters – [A], [B], [C], [D], [E], [F], [G], [H], [I], [J], [K]

Now let's calculate the distance between every pair and then using the distance we will fill and create the distance matrix.

Formula to calculate the **Euclidean Distance:**

Distance(A,B) = $\sqrt{(x_b - x_a)^2 + (y_b - y_a)^2}$

Calculating the distances from point A –

Distance(A,B) = $\sqrt{(2-2)^2 + (6-2)^2}$ = 4

Distance(A,C) = $\sqrt{(3-2)^2 + (7-2)^2}$ = 5.10

Distance(A,D) = $\sqrt{(5-2)^2 + (2-2)^2}$ = 3

Distance(A,E) = $\sqrt{(5-2)^2 + (5-2)^2}$ = 4.25

Distance(A,F) = $\sqrt{(5-2)^2 + (8-2)^2}$ = 6.71

Distance(A,G) = $\sqrt{(6-2)^2 + (6-2)^2}$ = 5.65

Distance(A,H) = $\sqrt{(7-2)^2 + (3-2)^2}$ = 5.10

Distance(A,I) = $\sqrt{(8-2)^2 + (4-2)^2}$ = 6.32

Distance(A,J) = $\sqrt{(10-2)^2 + (6-2)^2}$ = 8.94

Distance(A,K) = $\sqrt{(12-2)^2 + (8-2)^2}$ = 11.66

Calculating the distances from point B –

Distance(B,C) = $\sqrt{(3-2)^2 + (7-6)^2}$ = 1.41

Distance(B,D) = $\sqrt{(5-2)^2 + (2-6)^2}$ = 5

Distance(B,E) = $\sqrt{(5-2)^2 + (5-6)^2}$ = 3.16

Distance(B,F) = $\sqrt{(5-2)^2 + (8-6)^2}$ = 3.60

Distance(B,G) = $\sqrt{(6-2)^2 + (6-6)^2}$ = 4

Distance(B,H) = $\sqrt{(7-2)^2 + (3-6)^2}$ = 5.83

Distance(B,I) = $\sqrt{(8-2)^2 + (4-6)^2}$ = 6.32

Distance(B,J) = $\sqrt{(10-2)^2 + (6-6)^2}$ = 8

Distance(B,K) = $\sqrt{(12-2)^2 + (8-6)^2}$ = 10.19

Calculating the distances from point C –

Distance(C,D) = $\sqrt{(5-3)^2 + (2-7)^2}$ = 5.38

Distance(C,E) = $\sqrt{(5-3)^2 + (5-7)^2}$ = 2.82

Distance(C,F) = $\sqrt{(5-3)^2 + (8-7)^2}$ = 2.23

Distance(C,G) = $\sqrt{(6-3)^2 + (6-7)^2}$ = 4

Distance(C,H) = $\sqrt{(7-3)^2 + (3-7)^2}$ = 3.16

Distance(C,I) = $\sqrt{(8-3)^2 + (4-7)^2}$ = 5.83

Distance(C,J) = $\sqrt{(10-3)^2 + (6-7)^2}$ = 7.07

Distance(C,K) = $\sqrt{(12-3)^2 + (8-7)^2}$ = 9.05

Calculating the distances from point D –

Distance(D,E) = $\sqrt{(5-5)^2 + (5-2)^2}$ = 3

Distance(D,F) = $\sqrt{(5-5)^2 + (8-2)^2}$ = 6

Distance(D,G) = $\sqrt{(6-5)^2 + (6-2)^2}$ = 4.12

Distance(D,H) = $\sqrt{(7-5)^2 + (3-2)^2}$ = 2.23

Distance(D,I) = $\sqrt{(8-5)^2 + (4-2)^2}$ = 3.6

Distance(D,J) = $\sqrt{(10-5)^2 + (6-2)^2}$ = 6.4

Distance(D,K) = $\sqrt{(12-5)^2 + (8-2)^2}$ = 9.21

Calculating the distances from point E –

Distance(E,F) = $\sqrt{(5-5)^2 + (8-5)^2}$ = 3

Distance(E,G) = $\sqrt{(6-5)^2 + (6-5)^2}$ = 1.41

Distance(E,H) = $\sqrt{(7-5)^2 + (3-5)^2}$ = 2.82

Distance(E,I) = $\sqrt{(8-5)^2 + (4-5)^2}$ = 3.16

Distance(E,J) = $\sqrt{(10-5)^2 + (6-5)^2}$ = 5.10

Distance(E,K) = $\sqrt{(12-5)^2 + (8-5)^2}$ = 7.61

Calculating the distances from point F –

Distance(F,G) = $\sqrt{(6-5)^2 + (6-8)^2}$ = 2.23

Distance(F,H) = $\sqrt{(7-5)^2 + (3-8)^2}$ = 5.38

Distance(F,I) = $\sqrt{(8-5)^2 + (4-8)^2}$ = 5

Distance(F,J) = $\sqrt{(10-5)^2 + (6-8)^2}$ = 5.38

Distance(F,K) = $\sqrt{(12-5)^2 + (8-8)^2}$ = 7

Calculating the distances from point G –

Distance(G,H) = $\sqrt{(7-6)^2 + (3-6)^2}$ = 3.16

Distance(G,I) = $\sqrt{(8-6)^2 + (4-6)^2}$ = 2.82

Distance(G,J) = $\sqrt{(10-6)^2 + (6-6)^2}$ = 4

Distance(G,K) = $\sqrt{(12-6)^2 + (8-6)^2}$ = 6.32

Calculating the distances from point H –

Distance(H,I) = $\sqrt{(8-7)^2 + (4-3)^2}$ = 1.41

Distance(H,J) = $\sqrt{(10-6)^2 + (6-6)^2}$ = 4.24

Distance(H,K) = $\sqrt{(12-6)^2 + (8-6)^2}$ = 7.07

Calculating the distances from point I –

Distance(I,J) = $\sqrt{(10-8)^2 + (6-4)^2}$ = 2.82

Distance(I,K) = $\sqrt{(12-7)^2 + (8-3)^2}$ = 5.65

Calculating the distances from point J –

Distance(J,K) = $\sqrt{(12-10)^2 + (8-6)^2}$ = 2.82

INITIAL MATRIX

Now as we have all the distances we can create the Distance Matrix –

	A(2,2)	B(2,6)	C(3,7)	D(5,2)	E(5,5)	F(5,8)	G(6,6)	H(7,3)	I(8,4)	J(10,6)	K(12,8)
A(2,2)	0										
B(2,6)	4	0									
C(3,7)	5.10	1.41	0								
D(5,2)	3	5	5.38	0							
E(5,5)	4.25	3.16	2.82	3	0						
F(5,8)	6.71	3.60	2.23	6	3	0					
G(6,6)	5.65	4	4	4.12	1.41	2.23	0				
H(7,3)	5.10	5.83	3.16	2.23	2.82	5.38	3.16	0			
I(8,4)	6.32	6.32	5.83	3.6	3.16	5	2.82	1.41	0		
J(10,6)	8.94	8	7.07	6.4	5.10	5.38	4	4.24	2.82	0	
K(12,8)	11.66	10.19	9.05	9.21	7.61	7	6.32	7.07	5.65	2.82	0

Now the next step is to find the **minimum distance** in this Distance Matrix. Here, the minimum distance is 1.41, which is between the points **C(3,7) and B(2,6)**. Now we will merge these two points into a new cluster and then update the distance matrix with the distances from the new cluster to all other data points using the single-linkage criterion (shortest distance).

ITERATION-1

Updating the distance between the cluster (C, B) to A

Distance = Min (dist (C, A), dist (B, A))

Distance = Min (5.10, 4)

Distance = 4

Updating the distance between the cluster (C, B) to D

Distance = Min (dist (C, D), dist (B, D))

Distance = Min (5.38, 5)

Distance = 5

Updating the distance between the cluster (C, B) to E

Distance = Min (dist (C, E), dist (B, E))

Distance = Min (2.82, 3.16)

Distance = 2.82

Updating the distance between the cluster (C, B) to F

Distance = Min (dist (C, F), dist (B, F))

Distance = Min (3.6, 2.23)

Distance = 2.23

Updating the distance between the cluster (C, B) to G

Distance = Min (dist (C, G), dist (B, G))

Distance = Min (4, 4)

Distance = 4

Updating the distance between the cluster (C, B) to H

Distance = Min (dist (C, H), dist (B, H))

Distance = Min (3.16, 5.83)

Distance = 3.16

Updating the distance between the cluster (C, B) to I

Distance = Min (dist (C, I), dist (B, I))

Distance = Min (5.83, 6.32)

Distance = 5.83

Updating the distance between the cluster (C, B) to J

Distance = Min (dist (C, J), dist (B, J))

Distance = Min (7.07, 8)

Distance = 7.07

Updating the distance between the cluster (C, B) to K

Distance = Min (dist (C, K), dist (B, K))

Distance = Min (9.05, 10.19)

Distance = 9.05

Updated Matrix –

	A(2,2)	B(2,6)-C(3,7)	D(5,2)	E(5,5)	F(5,8)	G(6,6)	H(7,3)	I(8,4)	J(10,6)	K(12,8)
A(2,2)	0									
B(2,6)-C(3,7)	4	0								
D(5,2)	3	5	0							
E(5,5)	4.25	2.82	3	0						
F(5,8)	6.71	2.23	6	3	0					
G(6,6)	5.65	4	4.12	1.41	2.23	0				
H(7,3)	5.10	3.16	2.23	2.82	5.38	3.16	0			
I(8,4)	6.32	5.83	3.6	3.16	5	2.82	1.41	0		
J(10,6)	8.94	7.07	6.4	5.10	5.38	4	4.24	2.82	0	
K(12,8)	11.66	9.05	9.21	7.61	7	6.32	7.07	5.65	2.82	0

Now the next step is to find the minimum distance in this Distance Matrix. Here, the minimum distance is 1.41, which is between the points E(5,5) and G(6,6). Now we will merge these two points into a new cluster and then update the distance matrix with the distances from the new cluster to all other data points using the single-linkage criterion (shortest distance).

ITERATION-2

Updating the distance between the cluster (E,G) to A

Distance = Min (dist (E, A), dist (G, A))

Distance = Min (4.25, 5.65)

Distance = 4.25

Updating the distance between the cluster (E,G) to (B,C)

Distance = Min (dist (E, (B,C)), dist (G, (B,C)))

Distance = Min (2.82, 4)

Distance = 2.82

Updating the distance between the cluster (E,G) to D

Distance = Min (dist (E, D), dist (G, D))

Distance = Min (3, 4.12)

Distance = 3

Updating the distance between the cluster (E,G) to F

Distance = Min (dist (E, F), dist (G, F))

Distance = Min (3, 2.23)

Distance = 2.23

Updating the distance between the cluster (E,G) to H

Distance = Min (dist (E, H), dist (G, H))

Distance = Min (2.82, 3.16)

Distance = 2.82

Updating the distance between the cluster (E,G) to I

Distance = Min (dist (E, I), dist (G, I))

Distance = Min (3.16, 2.82)

Distance = 2.82

Updating the distance between the cluster (E,G) to J

Distance = Min (dist (E, J), dist (G, J))

Distance = Min (5.10, 4)

Distance = 4

Updating the distance between the cluster (E,G) to K

Distance = Min (dist (E, K), dist (G, K))

Distance = Min (7.61, 6.32)

Distance = 6.32

Updated Matrix –

	A(2,2)	B(2,6)-C(3,7)	D(5,2)	E(5,5)-G(6,6)	F(5,8)	H(7,3)	I(8,4)	J(10,6)	K(12,8)
A(2,2)	0								
B(2,6)-C(3,7)	4	0							
D(5,2)	3	5	0						
E(5,5)-G(6,6)	4.25	2.82	3	0					
F(5,8)	6.71	2.23	6	2.23	0				
H(7,3)	5.10	3.16	2.23	2.82	5.38	0			
I(8,4)	6.32	5.83	3.6	2.82	5	1.41	0		
J(10,6)	8.94	7.07	6.4	4	5.38	4.24	2.82	0	
K(12,8)	11.66	9.05	9.21	6.32	7	7.07	5.65	2.82	0

Now the next step is to find the minimum distance in this Distance Matrix. Here, the minimum distance is 1.41, which is between the points H(7,3) and I(8,4). Now we will merge these two points into a new cluster and then update the distance matrix with the distances from the new cluster to all other data points using the single-linkage criterion (shortest distance).

ITERATION-3

Updating the distance between the cluster (H, I) to A

Distance = Min (dist (H, A), dist (I, A))

Distance = Min (5.10, 6.32)

Distance = 5.10

Updating the distance between the cluster (H, I) to (B,C)

Distance = Min (dist (H, (B,C)), dist (I, (B,C)))

Distance = Min (3.16, 5.83)

Distance = 3.16

Updating the distance between the cluster (H, I) to D

Distance = Min (dist (H, D), dist (I, D))

Distance = Min (2.23, 3.6)

Distance = 2.23

Updating the distance between the cluster (H, I) to F

Distance = Min (dist (H, F), dist (I, F))

Distance = Min (5.38, 5)

Distance = 5

Updating the distance between the cluster (H, I) to J

Distance = Min (dist (H, J), dist (I, J))

Distance = Min (4.24, 2.82)

Distance = 2.82

Updating the distance between the cluster (H, I) to K

Distance = Min (dist (H, K), dist (I, K))

Distance = Min (7.07, 5.65)

Distance = 5.65

Updated Matrix –

	A(2,2)	B(2,6)-C(3,7)	D(5,2)	E(5,5)-G(6,6)	F(5,8)	H(7,3)-I(8,4)	J(10,6)	K(12,8)
A(2,2)	0							
B(2,6)-C(3,7)	4	0						
D(5,2)	3	5	0					
E(5,5)-G(6,6)	4.25	2.82	3	0				
F(5,8)	6.71	2.23	6	2.23	0			
H(7,3)-I(8,4)	5.10	3.16	2.23	2.82	5	0		
J(10,6)	8.94	7.07	6.4	4	5.38	2.82	0	
K(12,8)	11.66	9.05	9.21	6.32	7	5.65	2.82	0

Now the next step is to find the minimum distance in this Distance Matrix. Here, the minimum distance is 2.23, which is between the points B(2,6)-C(3,7) and F(5,8). Now we will merge these two points into a new cluster and then update the distance matrix with the distances from the new cluster to all other data points using the single-linkage criterion (shortest distance).

ITERATION-4

Updating the distance between the cluster (BC, F) to A

Distance = Min (dist (BC, A), dist (F, A))

Distance = Min (4, 6.71)

Distance = 4

Updating the distance between the cluster (BC, F) to D

Distance = Min (dist (BC, D), dist (F, D))

Distance = Min (5, 6)

Distance = 5

Updating the distance between the cluster (BC, F) to J

Distance = Min (dist (BC, J), dist (F, J))

Distance = Min (7.07, 5.38)

Distance = 5.38

Updating the distance between the cluster (BC, F) to K

Distance = Min (dist (BC, K), dist (F, K))

Distance = Min (9.05, 7)

Distance = 7

Updated Matrix –

	A(2,2)	B(2,6)-C(3,7)-F(5,8)	D(5,2)	E(5,5)-G(6,6)	H(7,3)-I(8,4)	J(10,6)	K(12,8)
A(2,2)	0						
B(2,6)-C(3,7)-F(5,8)	4	0					
D(5,2)	3	5	0				
E(5,5)-G(6,6)	4.25	2.82	3	0			
H(7,3)-I(8,4)	5.10	3.16	2.23	2.82	0		
J(10,6)	8.94	5.38	6.4	4	2.82	0	
K(12,8)	11.66	7	9.21	6.32	5.65	2.82	0

Now the next step is to find the minimum distance in this Distance Matrix. Here, the minimum distance is 2.23, which is between the points H(7,3)-I(8,4) and D(5,2). Now we will merge these two points into a new cluster and then update the distance matrix with the distances from the new cluster to all other data points using the single-linkage criterion (shortest distance).

ITERATION-5

Updating the distance between the cluster (HI, D) to A

Distance = Min (dist (HI, A), dist (D, A))

Distance = Min (5.10, 3)

Distance = 3

Updating the distance between the cluster (HI, D) to J

Distance = Min (dist (HI, J), dist (D, J))

Distance = Min (2.82, 6.4)

Distance = 2.82

Updating the distance between the cluster (HI, D) to K

Distance = Min (dist (HI, K), dist (D, K))

Distance = Min (5.65, 9.21)

Distance = 5.65

Updated Matrix –

	A(2,2)	B(2,6)-C(3,7)-F(5,8)	E(5,5)-G(6,6)	H(7,3)-I(8,4)-D(5,2)	J(10,6)	K(12,8)
A(2,2)	0					
B(2,6)-C(3,7)-F(5,8)	4	0				
E(5,5)-G(6,6)	4.25	2.82	0			
H(7,3)-I(8,4)-D(5,2)	3	3.16	2.82	0		
J(10,6)	8.94	5.38	4	2.82	0	
K(12,8)	11.66	7	6.32	5.65	2.82	0

Now the next step is to find the minimum distance in this Distance Matrix. Here, the minimum distance is 2.82, which is between the points (BCF) and (EG) Now we will merge these two points into a new cluster and then update the distance matrix with the distances from the new cluster to all other data points using the single-linkage criterion (shortest distance).

ITERATION-6

Updating the distance between the cluster (BCF, EG) to A

Distance = Min (dist (BCF, A), dist (EG, A))

Distance = Min (4, 4.25)

Distance = 4

Updating the distance between the cluster (BCF, EG) to J

Distance = Min (dist (BCF, J), dist (EG, J))

Distance = Min (5.38, 4)

Distance = 4

Updating the distance between the cluster (BCF, EG) to K

Distance = Min (dist (BCF, K), dist (EG, K))

Distance = Min (7, 6.32)

Distance = 6.32

Updated Matrix –

	A(2,2)	B(2,6)-C(3,7)-F(5,8)-E(5,5)-G(6,6)	H(7,3)-I(8,4)-D(5,2)	J(10,6)	K(12,8)
A(2,2)	0				
B(2,6)-C(3,7)-F(5,8)-E(5,5)-G(6,6)	4	0			
H(7,3)-I(8,4)-D(5,2)	3	3.16	0		
J(10,6)	8.94	5.38	2.82	0	
K(12,8)	11.66	7	5.65	2.82	0

Now the next step is to find the minimum distance in this Distance Matrix. Here, the minimum distance is 2.82, which is between the points (HID) and (J) Now we will merge these two points into a new cluster and then update the distance matrix with the distances from the new cluster to all other data points using the single-linkage criterion (shortest distance).

ITERATION-7

Updating the distance between the cluster (HID, J) to A

Distance = Min (dist (HID, A), dist (J, A))

Distance = Min (3, 8.94)

Distance = 3

Updating the distance between the cluster (HID, J) to (BCFEG)

Distance = Min (dist (HID,BCFEG), dist (J,BCFEG))

Distance = Min (3.16, 5.38)

Distance = 3.16

Updating the distance between the cluster (HID, J) to K

Distance = Min (dist (HID, K), dist (J, K))

Distance = Min (5.65, 2.82)

Distance = 2.82

Updated Matrix –

	A(2,2)	B(2,6)-C(3,7)-F(5,8)-E(5,5)-G(6,6)	H(7,3)-I(8,4)-D(5,2)-J(10,6)	K(12,8)
A(2,2)	0			
B(2,6)-C(3,7)-F(5,8)-E(5,5)-G(6,6)	4	0		
H(7,3)-I(8,4)-D(5,2)-J(10,6)	3	3.16	0	
K(12,8)	11.66	7	2.82	0

Now the next step is to find the minimum distance in this Distance Matrix. Here, the minimum distance is 2.82, which is between the points (HIDJ) and (K) Now we will merge these two points into a new cluster and then update the distance matrix with the distances from the new cluster to all other data points using the single-linkage criterion (shortest distance).

ITERATION-8

Updating the distance between the cluster (HIDJ, K) to A

Distance = Min (dist (HIDJ, A), dist (K, A))

Distance = Min (3, 11.66)

Distance = 3

Updating the distance between the cluster (HIDJ, K) to (BCFEG)

Distance = Min (dist (HIDJ,BCFEG), dist (K,BCFEG))

Distance = Min (3.16, 7)

Distance = 3.16

Updated Matrix –

	A(2,2)	B(2,6)-C(3,7)-F(5,8)- E(5,5)-G(6,6)	H(7,3)-I(8,4)-D(5,2)-J(10,6)-K(12,8)
A(2,2)	0		
B(2,6)-C(3,7)-F(5,8)- E(5,5)-G(6,6)	4	0	
H(7,3)-I(8,4)-D(5,2)-J(10,6)-K(12,8)	3	3.16	0

Now the next step is to find the minimum distance in this Distance Matrix. Here, the minimum distance is 3, which is between the points (HIDJK) and (A) Now we will merge these two points into a new cluster and then update the distance matrix with the distances from the new cluster to all other data points using the single-linkage criterion (shortest distance).

ITERATION-9

Updating the distance between the cluster (HIDJK, A) to (BCFEG)

Distance = Min (dist (HIDJK,BCFEG), dist (A,BCFEG))

Distance = Min (3.16, 4)

Distance = 3.16

Updated Matrix –

	B(2,6)-C(3,7)-F(5,8)- E(5,5)-G(6,6)	H(7,3)-I(8,4)-D(5,2)-J(10,6)-K(12,8)-A(2,2)
B(2,6)-C(3,7)-F(5,8)- E(5,5)-G(6,6)	0	
H(7,3)-I(8,4)-D(5,2)-J(10,6)-K(12,8)-A(2,2)	3.16	0

With this, we are done with obtaining a single cluster.