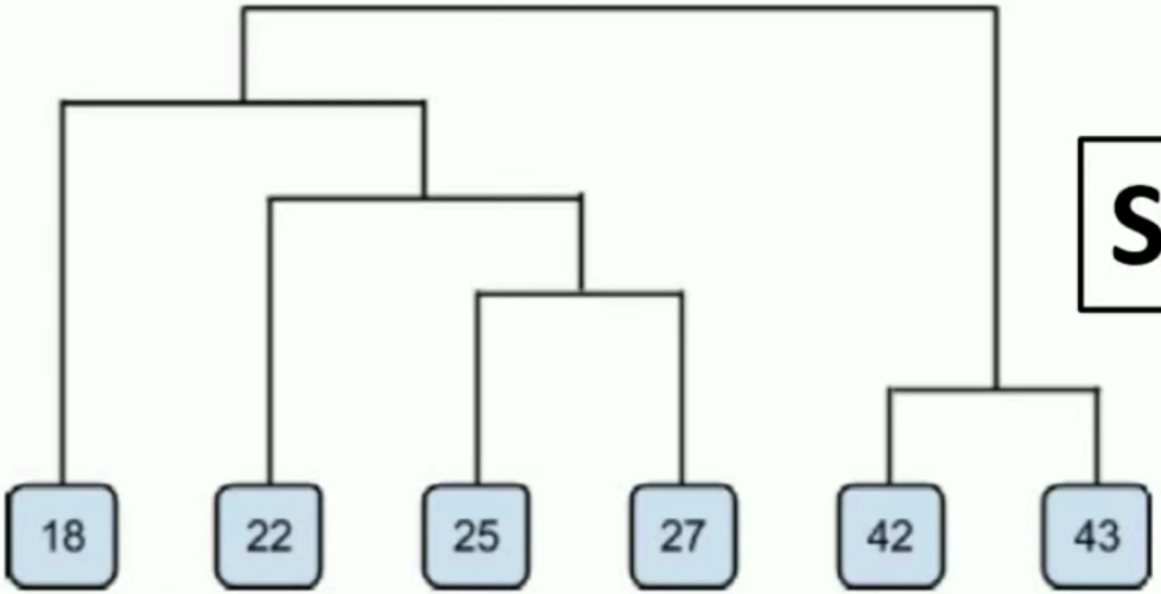
Agglomerative Hierarchical Clustering



Solved Example

Agglomerative Hierarchical Clustering Solved Example

- Consider the following set of 6 one dimensional data points:
- 18, 22, 25, 42, 27, 43
- Apply the **agglomerative hierarchical clustering** algorithm to build the hierarchical clustering **dendogram**.
- Merge the clusters using **Min distance** and update the proximity matrix accordingly.
- Clearly show the **proximity matrix** corresponding to each iteration of the algorithm.

Agglomerative Hierarchical Clustering Solved Example

- Step – 1

	18	22	25	27	42	43
18	0	4	7	9	24	25
22	4	0	3	5	20	21
25	7	3	0	2	17	18
27	9	5	2	0	15	16
42	24	20	17	15	0	1
43	25	21	18	16	1	0

Agglomerative Hierarchical Clustering Solved Example

- Step – 1

	18	22	25	27	42	43
18	0	4	7	9	24	25
22	4	0	3	5	20	21
25	7	3	0	2	17	18
27	9	5	2	0	15	16
42	24	20	17	15	0	1
43	25	21	18	16	1	0

(42, 43)

Agglomerative Hierarchical Clustering Solved Example

- Step – 2

	18	22	25	27	42, 43
18	0	4	7	9	24
22	4	0	3	5	20
25	7	3	0	2	17
27	9	5	2	0	15
42, 43	24	20	17	15	0

Agglomerative Hierarchical Clustering Solved Example

- Step – 2

	18	22	25	27	42, 43
18	0	4	7	9	24
22	4	0	3	5	20
25	7	3	0	2	17
27	9	5	2	0	15
42, 43	24	20	17	15	0

(42, 43), (25, 27)

Agglomerative Hierarchical Clustering Solved Example

- Step – 3

	18	22	25, 27	42, 43
18	0	4	7	24
22	4	0	3	20
25, 27	7	3	0	15
42, 43	24	20	15	0

Agglomerative Hierarchical Clustering Solved Example

- Step – 3

	18	22	25, 27	42, 43
18	0	4	7	24
22	4	0	3	20
25, 27	7	3	0	15
42, 43	24	20	15	0

(42, 43), ((25, 27), 22)

Agglomerative Hierarchical Clustering Solved Example

- Step – 4

	18	22, 25, 27	42, 43
18	0	4	24
22, 25, 27	4	0	15
42, 43	24	15	0

Agglomerative Hierarchical Clustering Solved Example

- Step – 4

	18	22, 25, 27	42, 43
18	0	4	24
22, 25, 27	4	0	15
42, 43	24	15	0

(42, 43), (((25, 27), 22), 18)

Agglomerative Hierarchical Clustering Solved Example

- Step – 5

	18, 22, 25, 27	42, 43
18, 22, 25, 27	0	15
42, 43	15	0

Agglomerative Hierarchical Clustering Solved Example

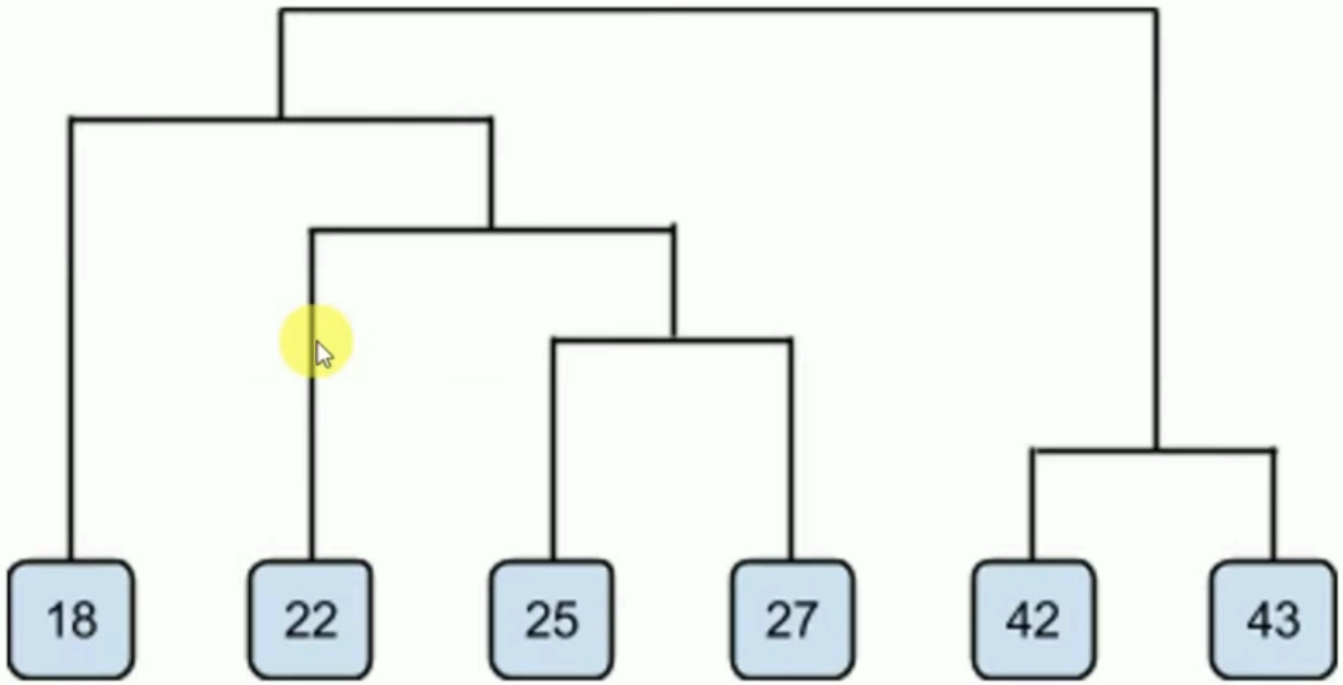
- Step – 6

	18, 22, 25, 27, 42, 43
18, 22, 25, 27, 42, 43	0

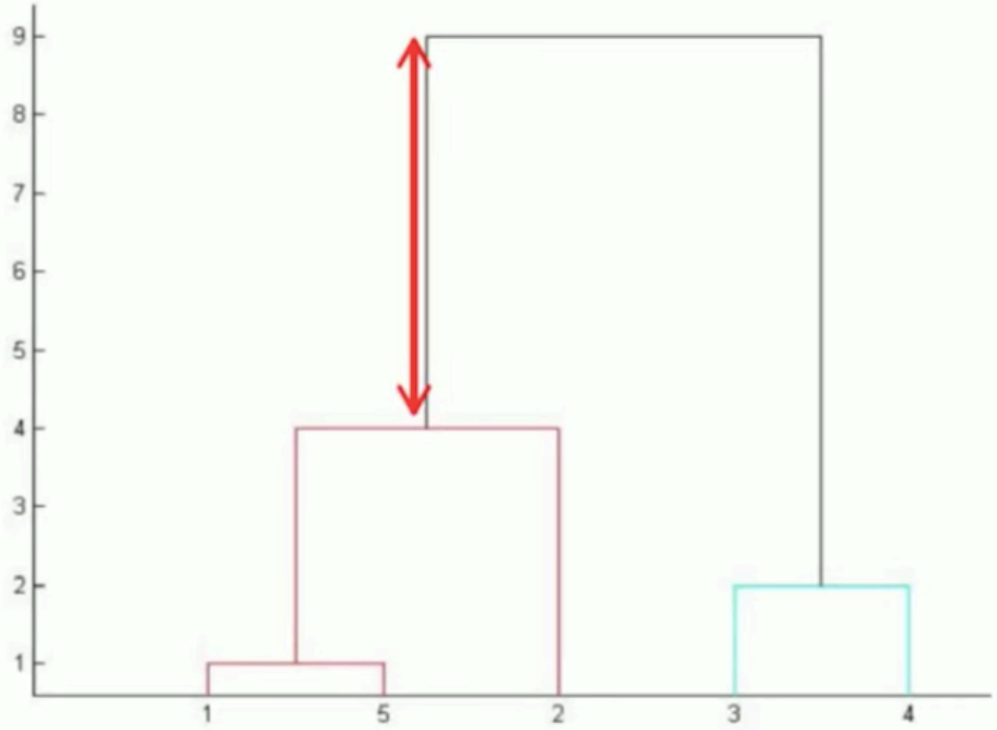
Agglomerative Hierarchical Clustering Solved Example

- **Dendrogram**

((42, 43), (((25, 27), 22), 18))



Complete Linkage



Agglomerative

Hierarchical

Clustering

Solved Example

Complete Linkage - Agglomerative Clustering

- Given a one-dimensional data set {1, 5, 8, 10, 2}, use the agglomerative clustering algorithms with the complete link with Euclidean distance to establish a hierarchical grouping relationship.
- By using the cutting threshold of 5, how many clusters are there?

Complete Linkage - Agglomerative Clustering

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

$$Euclidean\ distance = \sqrt{(x_2 - x_1)^2}$$

- In order to use the agglomerative algorithm,
- we need to calculate the distance matrix.
- One-dimensional data set {1, 5, 8, 10, 2}

	1	5	8	10	2
1	0	4	7	9	1
5	4	0	3	5	3
8	7	3	0	2	6
10	9	5	2	0	8
2	1	3	6	8	0

Complete Linkage - Agglomerative Clustering

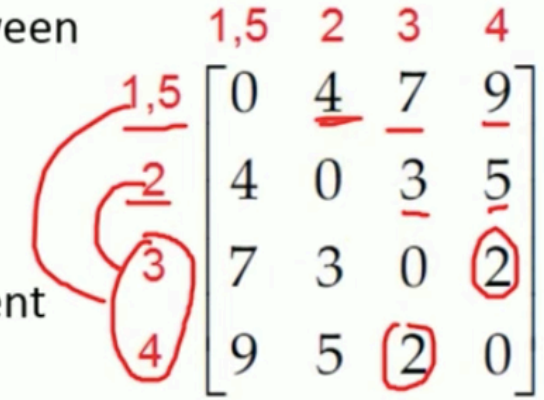
- d(2, {1,5}) = max{ d(2,1), d(2,5) } = max {4, 3} = 4
- d(3, {1,5}) = max{ d(3,1), d(3,5) } = max {7, 6} = 7
- d(4, {1,5}) = max{ d(4,1), d(4,5) } = max {9, 8} = 9
- Let the 1st column (row) denote the distances between this cluster and other points, we have the following distance matrix:

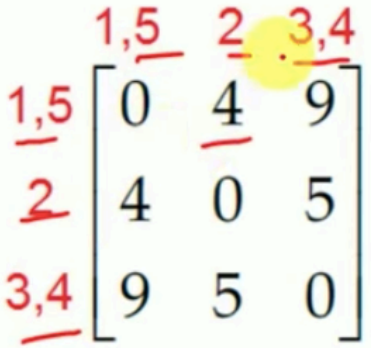
	1	2	3	4	5
1	0	4	7	9	1
2	4	0	3	5	3
3	7	3	0	2	6
4	9	5	2	0	8
5	1	3	6	8	0

	1,5	2	3	4
1,5	0	4	7	9
2	4	0	3	5
3	7	3	0	2
4	9	5	2	0

Complete Linkage - Agglomerative Clustering

- From the above distance matrix, we can see the distance between points 3 and 4 is smallest.
- Hence, they merge together to form a cluster {3, 4}.
- Using the complete link, we have the distance between different points/clusters as follows:
- d({1,5}, {3, 4}) = max{ d({1,5}, 3), d({1,5}, 4) } = max{ 7, 9} = 9
- d(2, {3,4}) = max{ d(2,3), d(2,4) } = max {3, 5} = 5
- Thus, we can update the distance matrix, where row 2 corresponds to point 2, rows 1 and 3 correspond to clusters {1, 5} and {3, 4}, as follows:





Complete Linkage - Agglomerative Clustering

- Following the same procedure, we merge point 2 with the cluster {1, 5} to form {1, 2, 5} and update the distance matrix as follows:

	[1,5],2	[3,4]
[1,5],2	0	9
[3,4]	9	0

	1,5	2	3,4
1,5	0	4	9
2	4	0	5
3,4	9	5	0

- After increasing the distance threshold to 9, all clusters would merge.
- Based on all above distance matrices, we draw the dendrogram tree as follows:

Complete Linkage - Agglomerative Clustering