Hierarchical methods - **agglomerative** and **divisive.**

Agglomerative methods:

- Start with partition P_n, where each object forms its own cluster.
- Merge the two closest clusters, obtaining P_{n-1}.
- Repeat merge until only one cluster is left.

Divisive methods

- Start with P_1.
- Split the collection into two clusters that are as homogenous (and as different from each other) as possible.
- Apply splitting procedure recursively to the clusters.

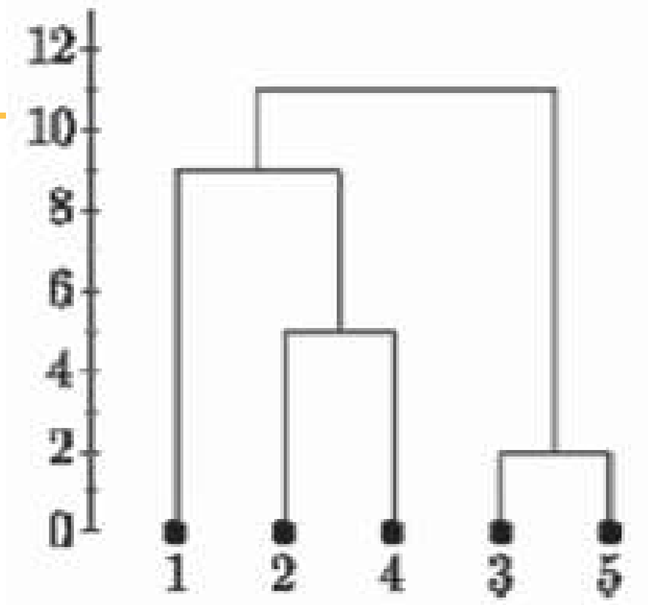
Only the lower triangle is shown, because the upper triangle can be filled in by reflection.

	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	$\boxed{2}$	8	0

•smallest distance is between three and five and they get linked up or merged first into a the cluster '35'.
•remove the 3 and 5 entries, and replace it by an entry "35" .
•distance between "35" and every other item is the maximum of the distance between this item and 3 and this item and 5.
•For example, d(1,3)= 3 and d(1,5)=11. So, D(1,"35")=11. This gives us the new distance matrix.

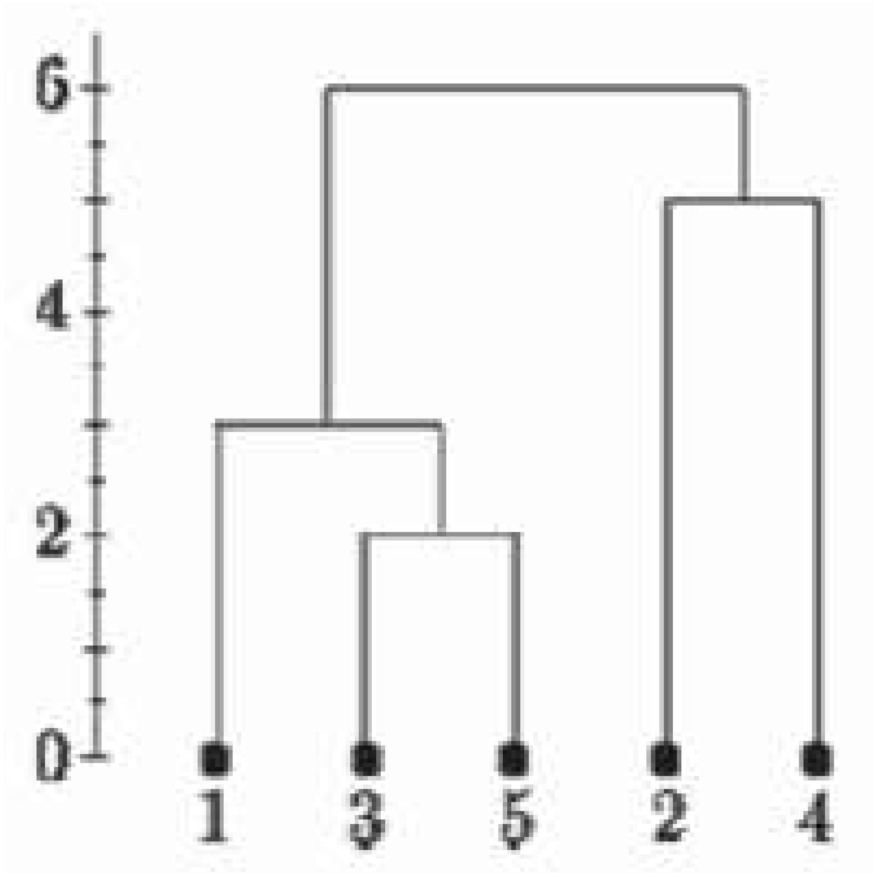
	35	1	2	4
35	0			
1	11	0		
2	10	9	0	
4	9	6	5	0

after 6 steps, everything is clustered. Below Plot (classically called as Dendrogram) On this plot, the y-axis shows the distance between the objects at the time they were clustered. This is called the cluster height. Different visualizations use different measures of cluster height.



Below is the single linkage dendrogram for the same distance matrix. It starts with cluster "35" but the distance between "35" and each item is now the minimum of d(x,3) and d(x,5). So c(1,"35")=3.

Single Linkage

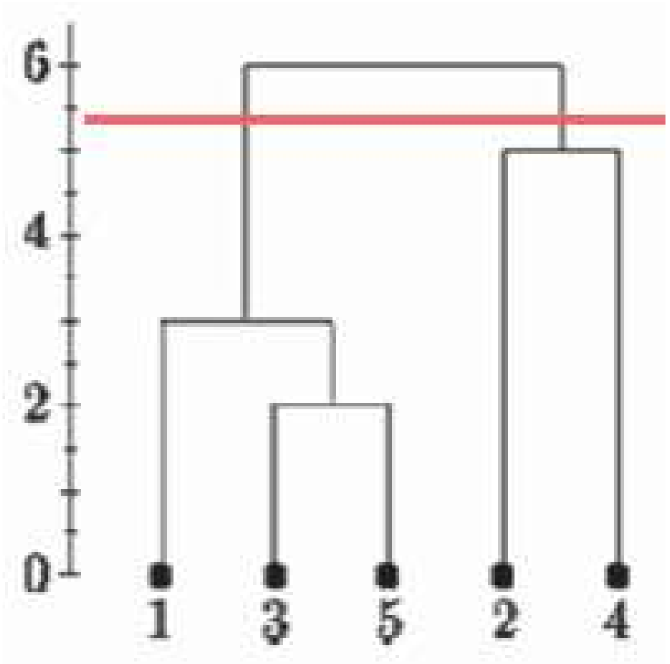


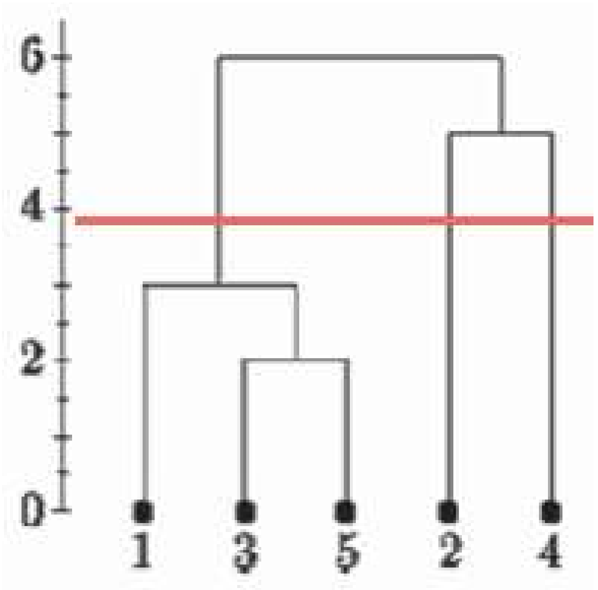
Determining clusters

One of the problems with hierarchical clustering is that there is no objective way to say how many clusters there are.

If we cut the single linkage tree at the point shown below, we would say that there are two clusters.

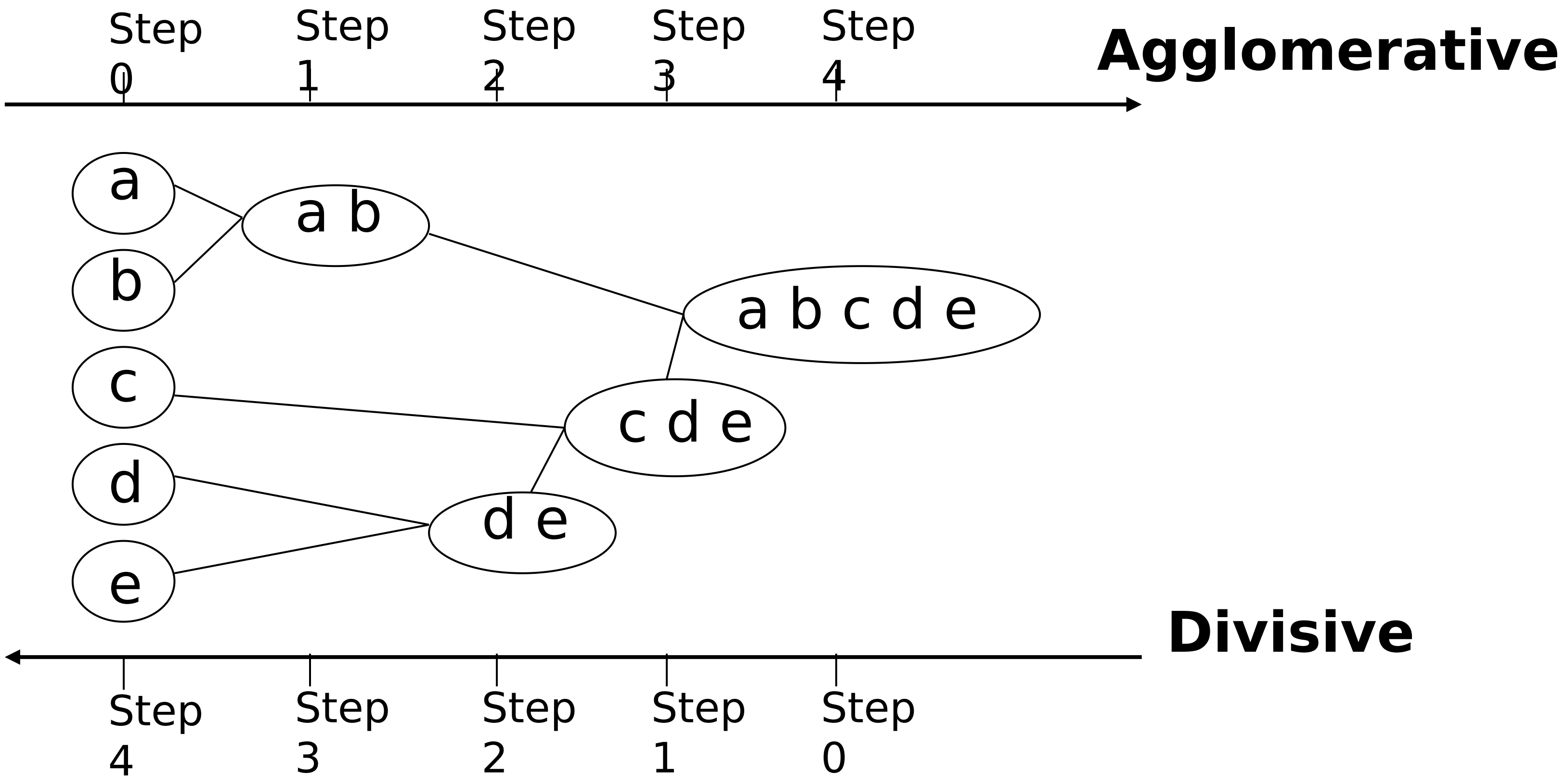
However, if we cut the tree lower we might say that there is one cluster and two singletons.





- Illustrative Example: Agglomerative vs. Divisive

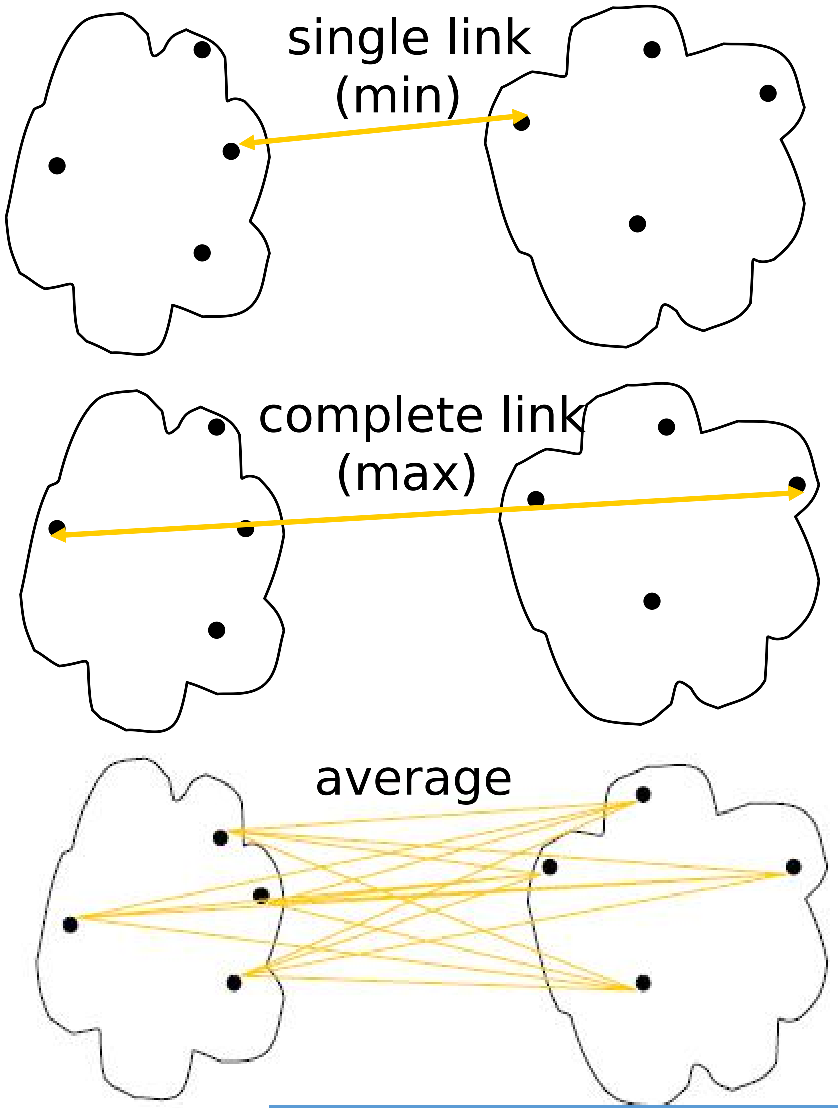
Agglomerative and divisive clustering on the data set {a, b, c, d ,e }



Cluster Distance Measures

- **Single link**: smallest distance between an element in one cluster and an element in the other, i.e., $d(C_i, C_j) = \min\{d(x_{ip}, x_{jq})\}$
- **Complete link**: largest distance between an element in one cluster and an element in the other, i.e., $d(C_i, C_j) = \max\{d(x_{ip}, x_{jq})\}$
- **Average**: avg distance between elements in one cluster and elements in the other, i.e.,

$d(C_i, C_j) = \text{avg}\{d(x_{ip}, x_{jq})\}$



$d(C, C)=0$

Cluster Distance Measures

Example: Given a data set of five objects characterised by a single continuous feature, assume that there are two clusters: C_1: {a, b} and C_2: {c, d, e}.

	a	**b**	**c**	**d**	**e**
Feature	1	2	4	5	6

1. Calculate the distance matrix .

2. Calculate three cluster distances between C_1 and C_2.

$$\begin{array}{c|c} & \begin{array}{ccccc} \mathbf{a} & \mathbf{b} & \mathbf{c} & \mathbf{d} & \mathbf{e} \end{array} \\ \begin{array}{c} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \\ \mathbf{d} \\ \mathbf{e} \end{array} & \begin{bmatrix} 0 & 1 & 3 & 4 & 5 \\ 1 & 0 & 2 & 3 & 4 \\ 3 & 2 & 0 & 1 & 2 \\ 4 & 3 & 1 & 0 & 1 \\ 5 & 4 & 2 & 1 & 0 \end{bmatrix} \end{array}$$

Single link

$$\mathrm{dist}(C_1, C_2) \quad \min\{d(\mathrm{a,c}), d(\mathrm{a,d}), d(\mathrm{a,e}), d(\mathrm{b,c}), d(\mathrm{b,d}), d(\mathrm{b,e})\}$$
$$\min\{3, 4, 5, 2, 3, 4\} \quad 2$$

Complete link

$$\mathrm{dist}(C_1, C_2) \quad \max\{d(\mathrm{a,c}), d(\mathrm{a,d}), d(\mathrm{a,e}), d(\mathrm{b,c}), d(\mathrm{b,d}), d(\mathrm{b,e})\}$$
$$\max\{3, 4, 5, 2, 3, 4\} \quad 5$$

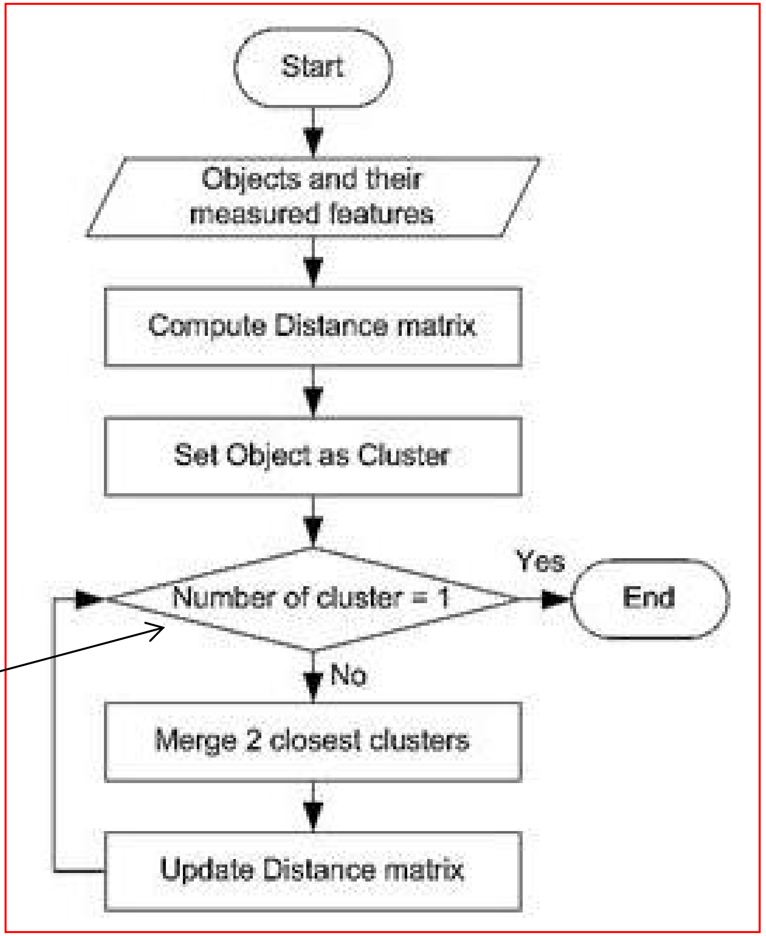
Average

$$\mathrm{dist}(C_1, C_2) \quad \frac{d(\mathrm{a,c}) \quad d(\mathrm{a,d}) \quad d(\mathrm{a,e}) \quad d(\mathrm{b,c}) \quad d(\mathrm{b,d}) \quad d(\mathrm{b,e})}{6}$$
$$\frac{3 \quad 4 \quad 5 \quad 2 \quad 3 \quad 4}{6} \quad \frac{21}{6} \quad 3.5$$

Agglomerative Algorithm

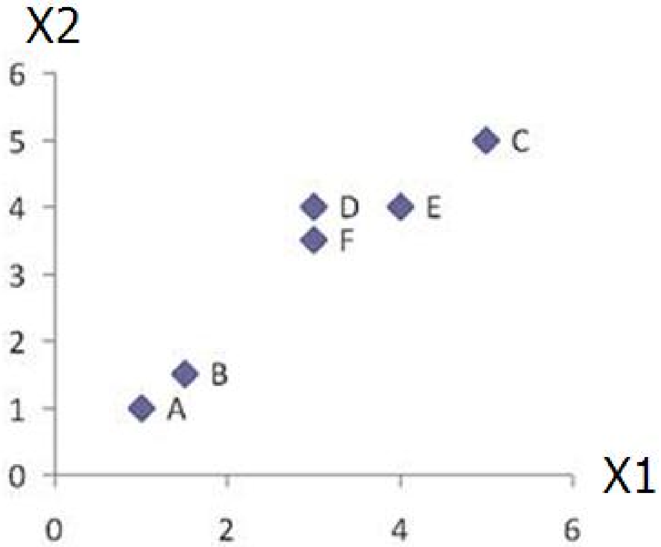
- The *Agglomerative* algorithm is carried out in three steps:

1) Convert all object features into a distance matrix
2) Set each object as a cluster (thus if we have N objects, we will have N clusters at the beginning)
3) Repeat until number of cluster is one (or known # of clusters)
   - Merge two closest clusters
   - Update "distance matrix"



Example

- Problem: clustering analysis with agglomerative algorithm



	X1	X2
A	1	1
B	1.5	1.5
C	5	5
D	3	4
E	4	4
F	3	3.5

data matrix

$$d_{AB} = \left((1-1.5)^2 + (1-1.5)^2\right)^{\frac{1}{2}} = \sqrt{\tfrac{1}{2}} = 0.7071$$

$$d_{DF} = \left((3-3)^2 + (4-3.5)^2\right)^{\frac{1}{2}} = 0.5$$

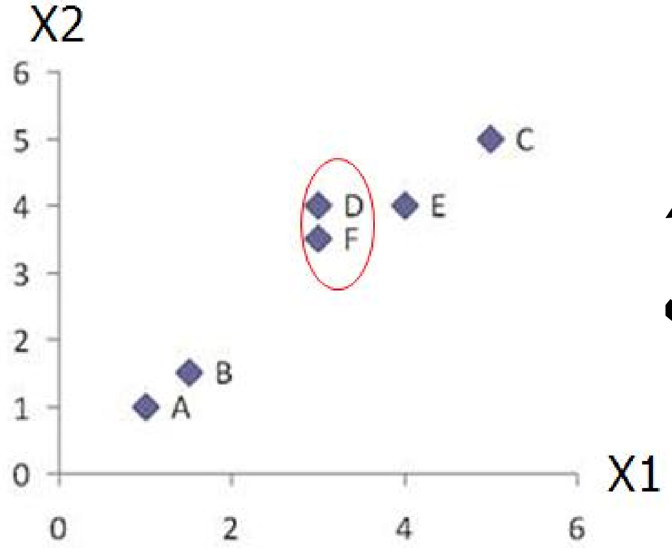
Euclidean distance

Dist	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

distance matrix

Example

- Merge two closest clusters (iteration 1)



Dist	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	?	4.24
B	0.71	0.00	4.95	?	3.54
C	5.66	4.95	0.00	?	1.41
D, F	?	?	?	0.00	?
E	4.24	3.54	1.41	?	0.00

Example

- Update distance matrix (iteration 1)

Dist	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

$$d_{(D,F)\mapsto A} = \min(d_{DA}, d_{FA}) = \min(3.61, 3.20) = 3.20$$

$$d_{(D,F)\mapsto B} = \min(d_{DB}, d_{FB}) = \min(2.92, 2.50) = 2.50$$

$$d_{(D,F)\mapsto C} = \min(d_{DC}, d_{FC}) = \min(2.24, 2.50) = 2.24$$

$$d_{E\to(D,F)} = \min(d_{ED}, d_{EF}) = \min(1.00, 1.12) = 1.00$$

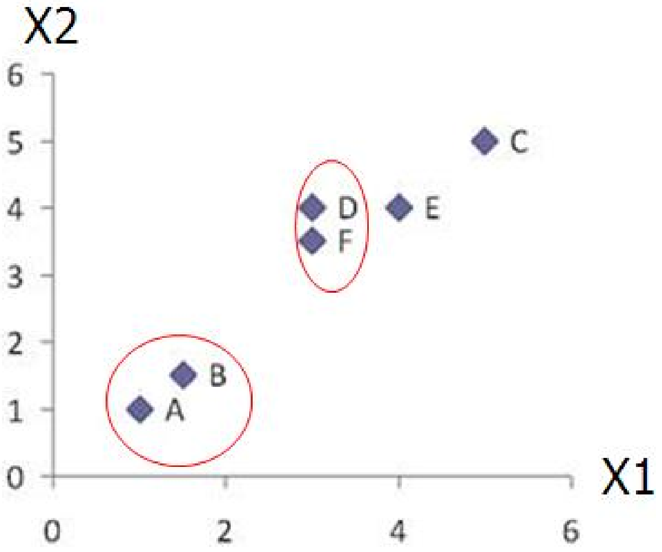
Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	?	4.24
B	0.71	0.00	4.95	?	3.54
C	5.66	4.95	0.00	?	1.41
D, F	?	?	?	0.00	?
E	4.24	3.54	1.41	?	0.00

Min Distance (Single Linkage)

Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	3.20	4.24
B	0.71	0.00	4.95	2.50	3.54
C	5.66	4.95	0.00	2.24	1.41
D, F	**3.20**	**2.50**	**2.24**	0.00	1.00
E	4.24	3.54	1.41	**1.00**	0.00

Example

- Merge two closest clusters (iteration 2)



Min Distance (Single Linkage)

Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	3.20	4.24
B	0.71	0.00	4.95	2.50	3.54
C	5.66	4.95	0.00	2.24	1.41
D, F	**3.20**	**2.50**	**2.24**	0.00	1.00
E	4.24	3.54	1.41	**1.00**	0.00

Dist	A,B	C	(D, F)	E
A,B	0	?	?	?
C	?	0	2.24	1.41
(D, F)	?	2.24	0	1.00
E	?	1.41	1.00	0

Example

- Update distance matrix (iteration 2)

Min Distance (Single Linkage)

Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	3.20	4.24
B	0.71	0.00	4.95	2.50	3.54
C	5.66	4.95	0.00	2.24	1.41
D, F	**3.20**	**2.50**	**2.24**	0.00	1.00
E	4.24	3.54	1.41	**1.00**	0.00

$$d_{C\to(A,B)} = \min\left(d_{CA}, d_{CB}\right) = \min\left(5.66, 4.95\right) = 4.95$$

$$d_{(D,F)\to(A,B)} = \min\left(d_{DA}, d_{DB}, d_{FA}, d_{FB}\right)$$
$$= \min\left(3.61, 2.92, 3.20, 2.50\right) = 2.50$$

$$d_{E\to(A,B)} = \min\left(d_{EA}, d_{EB}\right) = \min\left(4.24, 3.54\right) = 3.54$$

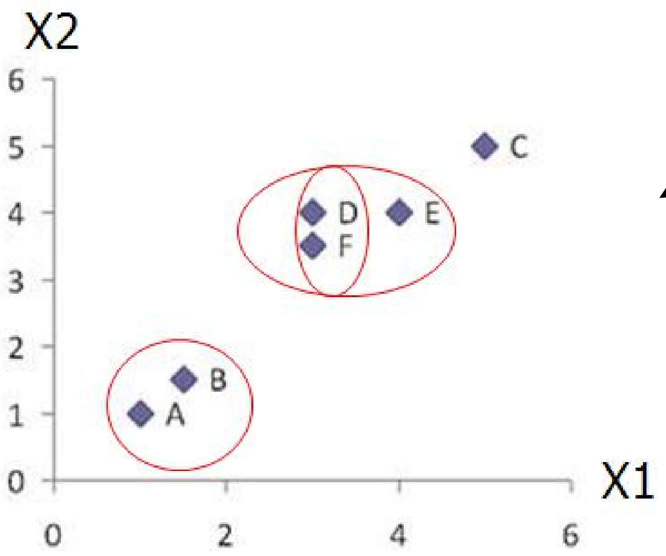
Dist	A,B	C	(D, F)	E
A,B	0	?	?	?
C	**?**	0	2.24	1.41
(D, F)	**?**	2.24	0	1.00
E	**?**	1.41	1.00	0

Min Distance (Single Linkage)

Dist	A,B	C	(D, F)	E
A,B	0	4.95	2.50	3.54
C	**4.95**	0	2.24	1.41
(D, F)	**2.50**	2.24	0	1.00
E	**3.54**	1.41	1.00	0

Example

- Merge two closest clusters/update distance matrix (iteration 3)



Min Distance (Single Linkage)

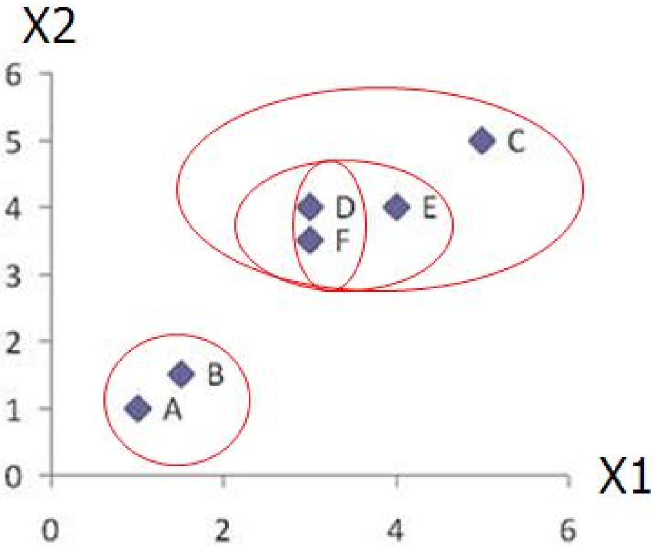
Dist	A,B	C	(D, F)	E
A,B	0	4.95	2.50	3.54
C	**4.95**	0	2.24	1.41
(D, F)	**2.50**	2.24	0	1.00
E	**3.54**	1.41	1.00	0

Min Distance (Single Linkage)

Dist	(A,B)	C	(D, F), E
(A,B)	0.00	4.95	2.50
C	4.95	0.00	1.41
(D, F), E	**2.50**	**1.41**	0.00

Example

- Merge two closest clusters/update distance matrix (iteration 4)



Min Distance (Single Linkage)

Dist	(A,B)	C	(D, F), E
(A,B)	0.00	4.95	2.50
C	4.95	0.00	1.41
(D, F), E	2.50	1.41	0.00

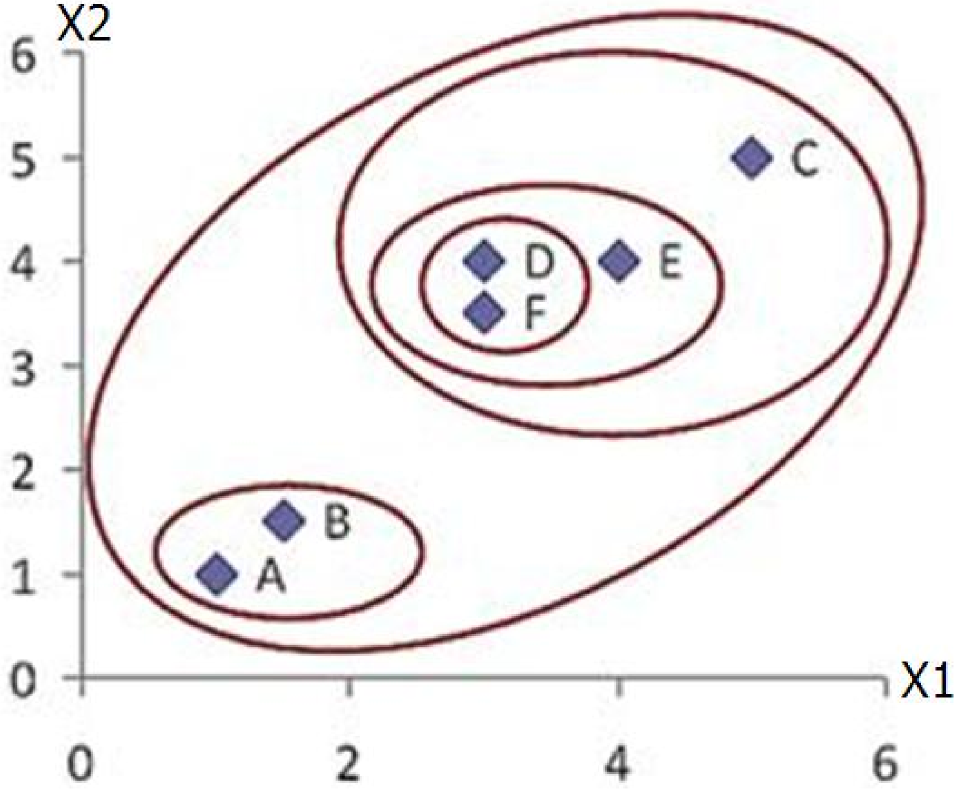
Min Distance (Single Linkage)

Dist	(A,B)	((D, F), E),C
(A,B)	0.00	2.50
((D, F), E),C	2.50	0.00

Example

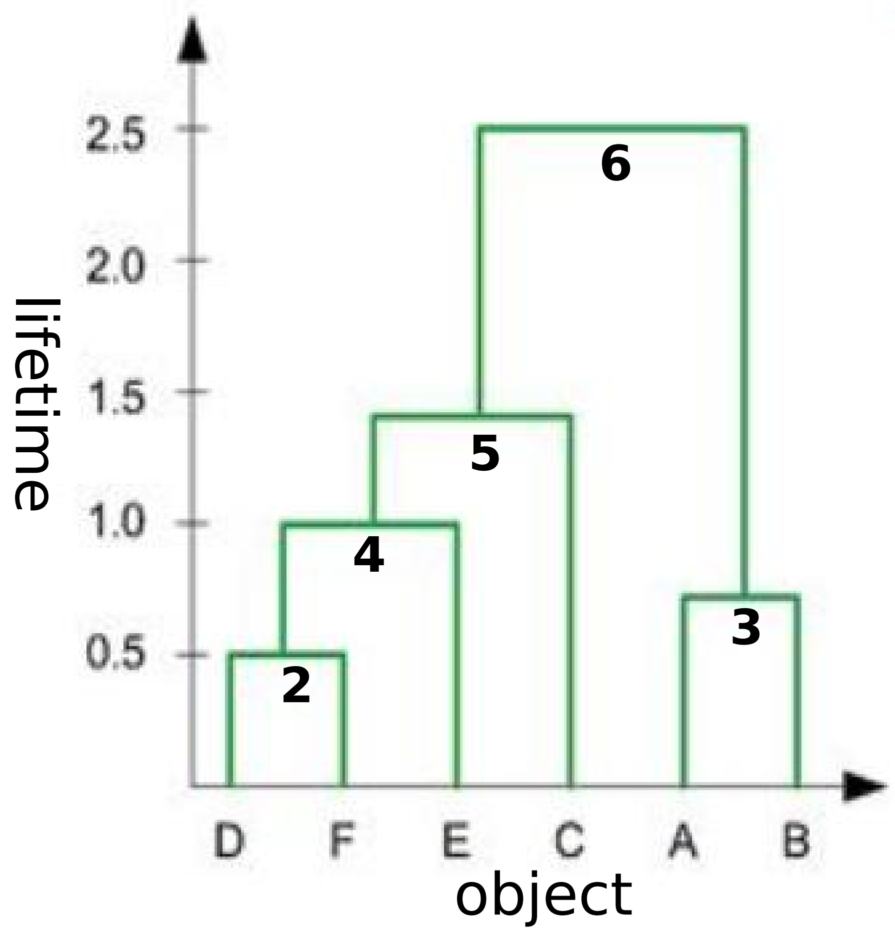
- Final result (meeting termination condition)

	X1	X2
A	1	1
B	1.5	1.5
C	5	5
D	3	4
E	4	4
F	3	3.5



Key Concepts in Hierarchal Clustering

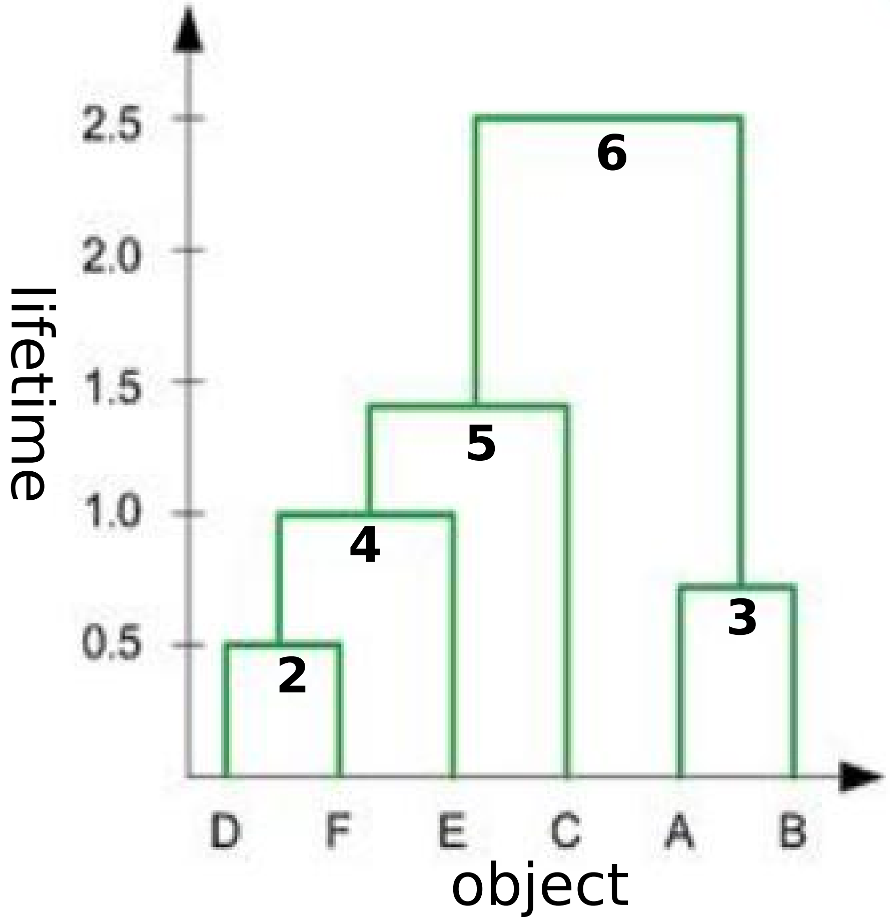
- Dendrogram tree representation



1. In the beginning we have 6 clusters: A, B, C, D, E and F
2. We merge clusters D and F into cluster (D, F) at distance 0.50
3. We merge cluster A and cluster B into (A, B) at distance 0.71
4. We merge clusters E and (D, F) into ((D, F), E) at distance 1.00
5. We merge clusters ((D, F), E) and C into (((D, F), E), C) at distance 1.41
6. We merge clusters (((D, F), E), C) and (A, B) into ((((D, F), E), C), (A, B)) at distance 2.50
7. The last cluster contain all the objects, thus conclude the computation

Key Concepts in Hierarchal Clustering

- Lifetime vs *K*-cluster Lifetime



- **Lifetime**

  The distance between that a cluster is created and that it disappears (merges with other clusters during clustering).

  e.g. lifetime of A, B, C, D, E and F are 0.71, 0.71, 1.41,

  0.50, 1.00 and 0.50, respectively, the life time of (A, B) is

  2.50 – 0.71 = 1.79, ……

- ***K*-cluster Lifetime**

  The distance from that *K* clusters emerge to that *K* clusters

  vanish (due to the reduction to *K-1* clusters).

  e.g.

  5-cluster lifetime is 0.71 - 0.50 = 0.21

  4-cluster lifetime is 1.00 - 0.71 = 0.29

  3-cluster lifetime is 1.41 – 1.00 = 0.41

  2-cluster lifetime is 2.50 – 1.41 = 1.09