Agglomerative Clustering Algorithm

More popular hierchical clustering technique

Basic algorithm is straightforward
- Compute the proximity matrix
- Let each data point be a cluster
- Repeat
  - Merge the two closest clusters
  - Update the proximity matrix
- Until only a single cluster matrix

Key operation is the computation of the proximity of two clusters.

Q cluster these six points using Agglomerative hierarchical clustering, Euclidean distance for the distance measure between objects and single link for the distance between clusters. Show clearly all iterations and give the Dendogram of clustering.

A(7,8), B(3,5), C(6,4), D(6,0), E(1,5), F(8,4

Step 1 Compute the proximity matrix.

	A	B	C	D	E	F
A	0					
B	5	0				
C	4.123	3.162	0			
D	8.062	5.830	4	0		
E	6.708	(2)	5.099	7.071	0	
F	4.123	5.099	2	4.472	7.071	0

Using Euclidean Distance for distance Measure

$d(A,B) = \sqrt{(7-3)^2 + (8-5)^2} = 5$

$d(A,C) = \sqrt{(7-6)^2 + (8-4)^2} = 4.123$

$d(A,D) = \sqrt{(7-6)^2 + (8-0)^2} = 8.062$

$d(A,E) = \sqrt{(7-1)^2 + (8-5)^2} = 6.708$

$d(A,F) = \sqrt{(7-8)^2 + (8-4)^2} = 4.123$

$d(B,C) = \sqrt{(3-6)^2+(5-4)^2} = 3.162$

$d(B,D) = \sqrt{(3-6)^2+(5-0)^2} = 5.830$

$d(B,E) = \sqrt{(3-1)^2+(5-5)^2} = 2$

$d(B,F) = \sqrt{(3-8)^2+(5-4)^2} = 5.099$

$d(C,D) = \sqrt{(6-6)^2+(4-0)^2} = 4$

$d(C,E) = \sqrt{(6-1)^2+(4-5)^2} = 5.099$

$d(C,F) = \sqrt{(6-8)^2+(4-4)^2} = 2$

$d(D,E) = \sqrt{(6-1)^2+(0-5)^2} = 9.071$

$d(D,F) = \sqrt{(6-8)^2+(0-4)^2} = 4.472$

$d(E,F) = \sqrt{(1-8)^2+(5-4)^2} = 7.071$

Step 2 Choose the smallest value other than '0'. Merge the smallest value

Selecting value '2' from (B & E) so we will merge the B & E.

Merge B & E Iteration #1

	A	(B,E)	C	D	F
A	0				
B,E	5	0			
C	4.123	3.162	0		
D	8.062	5.830	4	0	
F	4.123	5.099	(2)	4.472	0

Calculating Single Link method for the distance between clusters.

$$d((B,E),A) = \min(d(B,A), d(E,A))$$
$$= \min(5, 6.708)$$
$$= 5.$$

$$d((B,E),C) = 3.162$$

$$d((B,E),D) = \min(5.8301, 7.071) = 5.830$$
$$d((B,E),F) = \min(5.099, 7.071) = 5.099$$

Merge C and F Iteration 2

	A	(B,E)	(C,F)	D
A	0			
(B,E)	5	0		
(C,F)	4.123	3.162	0	
D	8.062	5.830	4	0

$$d((C,F),A) = \min(d(C,A), d(F,A))$$
$$= \min(4.123, 4.123)$$
$$= 4.123$$

$$d((C,F),(B,E)) = \min(d(C,B), d(C,E), d(F,B), d(F,E))$$
$$= \min(3.162, 5.099, 5.099, 7.071)$$
$$= 3.162.$$

$$d((C,F),D) = \min(d(C,D), d(F,D))$$
$$= \min(4, 4.472)$$
$$= 4.$$

Merge (B,E),(C,F) Iteration #3

	A	(B,E)(C,F)	D
A	0		
(B,E)(C,F)	4.123	0	
D	8.062	(4)	0

$d[(B,E,C,F),A] = \min(d(B,A), d(E,A), d(C,A), d(F,A))$
$= \min(5, 6.708, 4.123, 4.123)$
$= 4.123$

$d[(B,E,C,F),D] = \min(d(B,D), d(E,D), d(C,D), d(F,D))$
$= \min(5.830, 7.071, 4, 4.472)$
$= 4.$

Merge B,E,C,F,D Iteration #4

	A	B,E,C,F,D
A	0	
B,E,C,F,D	4.123	0

$d[(B,E,C,F,D),A] = \min(d(B,A), d(E,A), d(C,A), d(F,A), d(D,A))$

$= \min(5, 6.708, 4.123, 4.123, 8.062)$
$= 4.123$

Dendogram