

Hierarchical Agglomerative Clustering (HAC)

Different methods:

1. Single –nearest distance or single linkage method- minimum
2. Complete – farthest distance or complete linkage method - maximum
3. Average distance or average linkage method -average

Problem 1:

Find the clusters using single linkage method. Use Euclidean distance and draw the dendrogram.

	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

Calculate the Euclidean distance and create the distance matrix.

For example:

Distance between (x,y) and (a,b)= $\sqrt{(x-a)^2+(y-b)^2}$

Distance (p1,p2)= $\sqrt{(0.40-0.22)^2+(0.53-0.38)^2}$

$$=\sqrt{(0.18)^2+(0.15)^2}$$

$$=\sqrt{0.0324+0.0225}$$

$$=\sqrt{0.0549}$$

$$=0.23$$

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.23	0				
P3	0.22	0.15	0			
P4	0.37	0.20	0.15	0		
P5	0.34	0.14	0.28	0.29	0	
P6	0.23	0.25	0.11	0.22	0.39	0

Select the minimum entry in the distance matrix:

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.23	0				
P3	0.22	0.15	0			
P4	0.37	0.20	0.15	0		
P5	0.34	0.14	0.28	0.29	0	
P6	0.23	0.25	0.11	0.22	0.39	0

Update the distance matrix with the following criteria:

$$\text{Min}(\text{dist}(\text{P3}, \text{P6}), \text{P1}) = \min(\text{dist}(\text{P3}, \text{P1}), \text{dist}(\text{P6}, \text{P1}))$$

$$= \min(0.22, 0.23)$$

$$= 0.22$$

$$\text{Min}(\text{dist}(\text{P3}, \text{P6}), \text{P2}) = \min(\text{dist}(\text{P3}, \text{P2}), \text{dist}(\text{P6}, \text{P2}))$$

$$= \min(0.15, 0.25)$$

$$= 0.15$$

	P1	P2	P3,P6	P4	P5
P1	0				

P2	0.23	0			
P3,P6	0.22	0.15	0		
P4	0.37	0.20	0.15	0	
P5	0.34	0.14	0.28	0.29	0

$\text{Min}(\text{dist}(p2,p1), (p5,p1))$

$=\min(0.23,0.34)$

$=0.23$

$\text{Min}(\text{dist}(p2,(p3,p6)), (p5,(p3,p6)))$

$=\min(0.15,0.28)$

$=0.15$

$\text{Min}(\text{dist}(p2,p4), (p5,p4))$

$=\min(0.20,0.29)$

$=0.20$

Update the matrix based on the selected smallest distance between P2 and P5

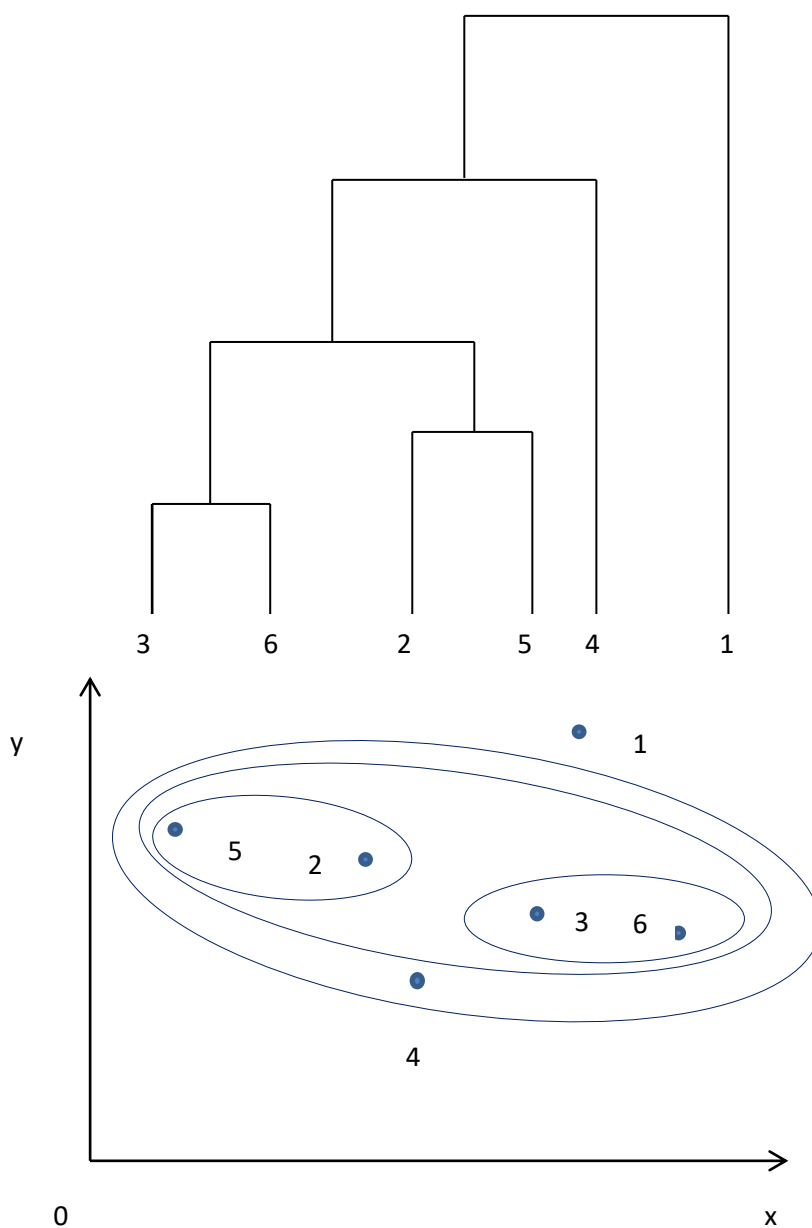
	P1	P2,P5	P3,P6	P4
P1	0			
P2,P5	0.23	0		
P3,P6	0.22	0.15	0	
P4	0.37	0.20	0.15	0

	P1	P2,p5,p3,p6	P4
P1	0		
P2,p5,p3,p6	0.22	0	
P4	0.37	0.15	0

The updated matrix for the clusters $\{p1\}$ and $\{p2,p5,p3,p6,p4\}$

	P1	P2,p5,p3,p6,p4
P1	0	
P2,p5,p3,p6,p4	0.22	0

Dendrogram:



Courtesy : <https://www.youtube.com/watch?v=RdT7bhm1M3E>