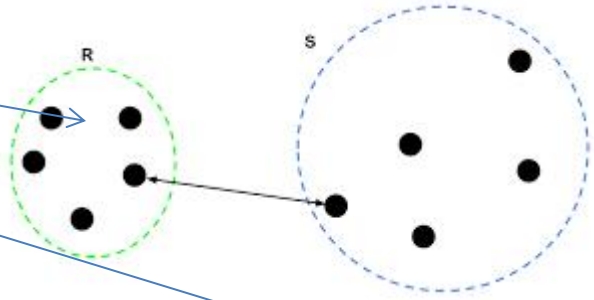
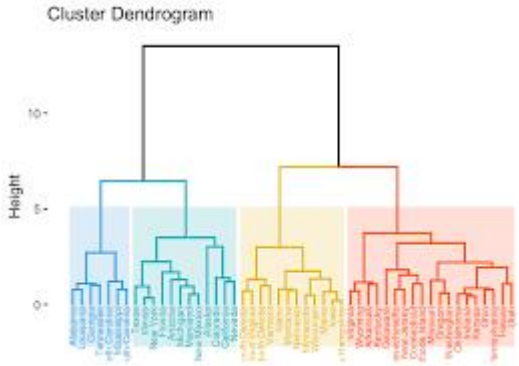
Hierarchical Agglomerative Clustering

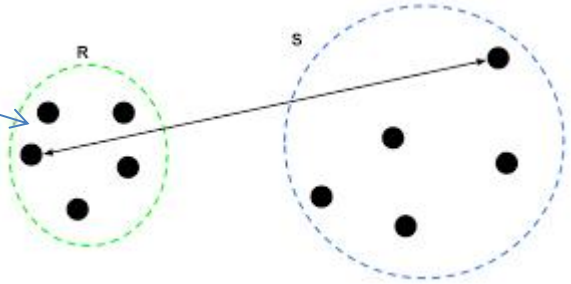
HAC

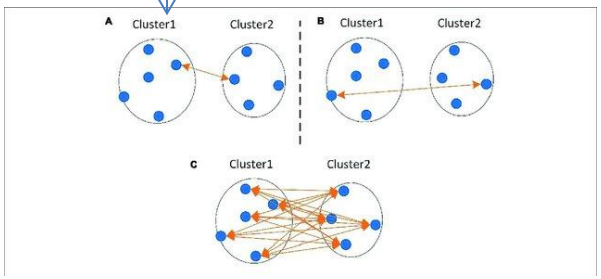
Agglomerative Clustering

- Bottom up approach
- Starts with one cluster and individual node
- Ends with Head, single cluster and all the data items will be merged.
- Dendrograms
- Why and When?
- Single linkage
- Compete linkage
- Average linkage



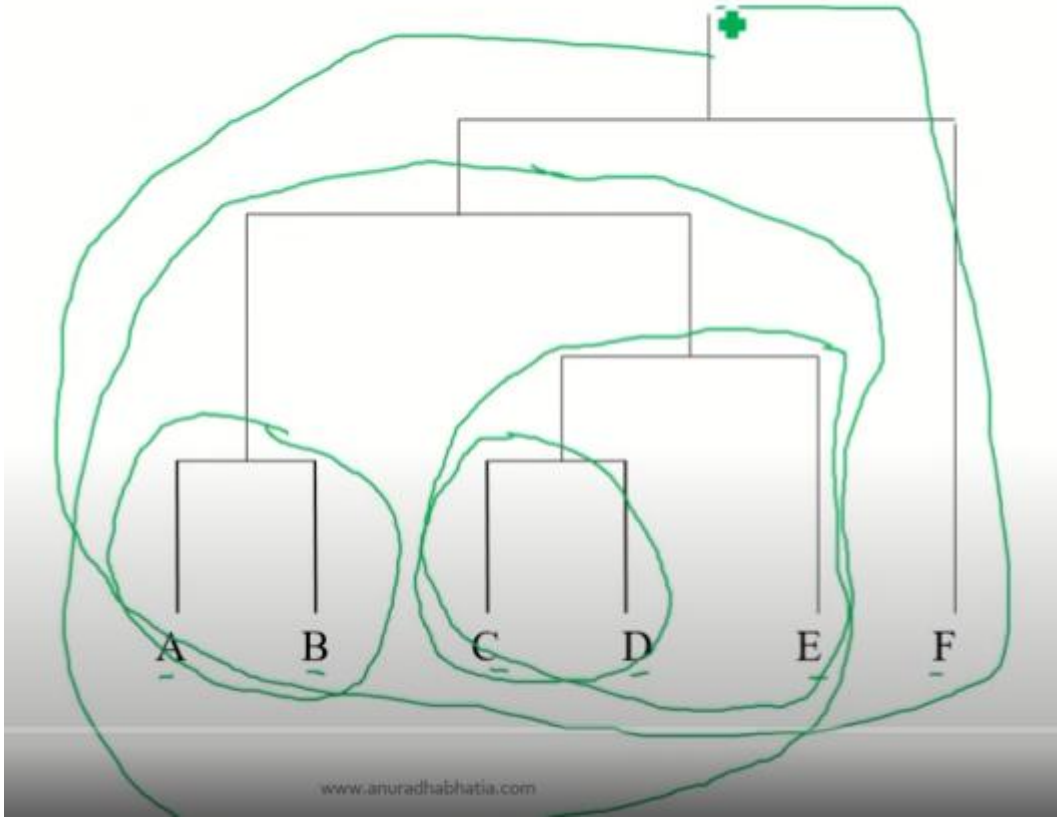






Dendrigrams – leaf Individual
and Root one cluster.

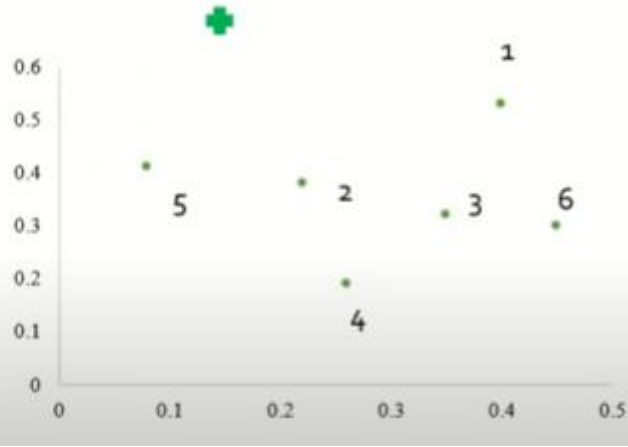
Dendrograms



Example

	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

	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



- Use Single link &
- Euclidean distance

- Calculate Euclidean distance, create the distance matrix.

$$\text{Distance } [(x,y), (a,b)] = \sqrt{(x-a)^2+(x-b)^2}$$

$$\text{Distance (P1,P2)} = \sqrt{(0.40-0.22)^2+(0.53-0.38)^2}$$

$$\text{(0.40,0.53), (0.22,0.38)} = \sqrt{(0.18)^2+(0.15)^2}$$

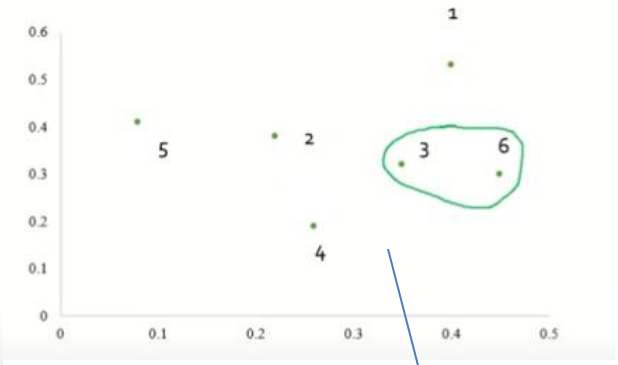
$$= \sqrt{0.0324+0.0225}$$

$$= \sqrt{0.0549}$$

Distance Matrix

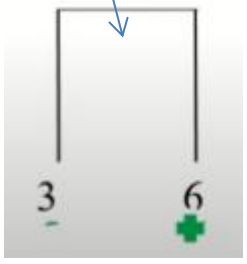
The distance matrix is

	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



P3,p6

Next step? – new cluster or extend



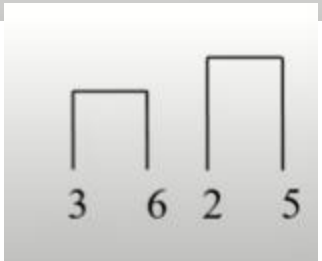
Update Distance matrix

- Min[dis(p3,p6),p1] = min((p3,p1),(p6,p1)) = 0.22
- Min[dis(p3,p6),P2)] = 0.15
- Min[dis(p3,p6),p4)] = 0.15
- Min[id(p3,p6),P5)] = 0.28

The updated distance matrix for cluster P3, P6

	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

	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



Update Distance matrix

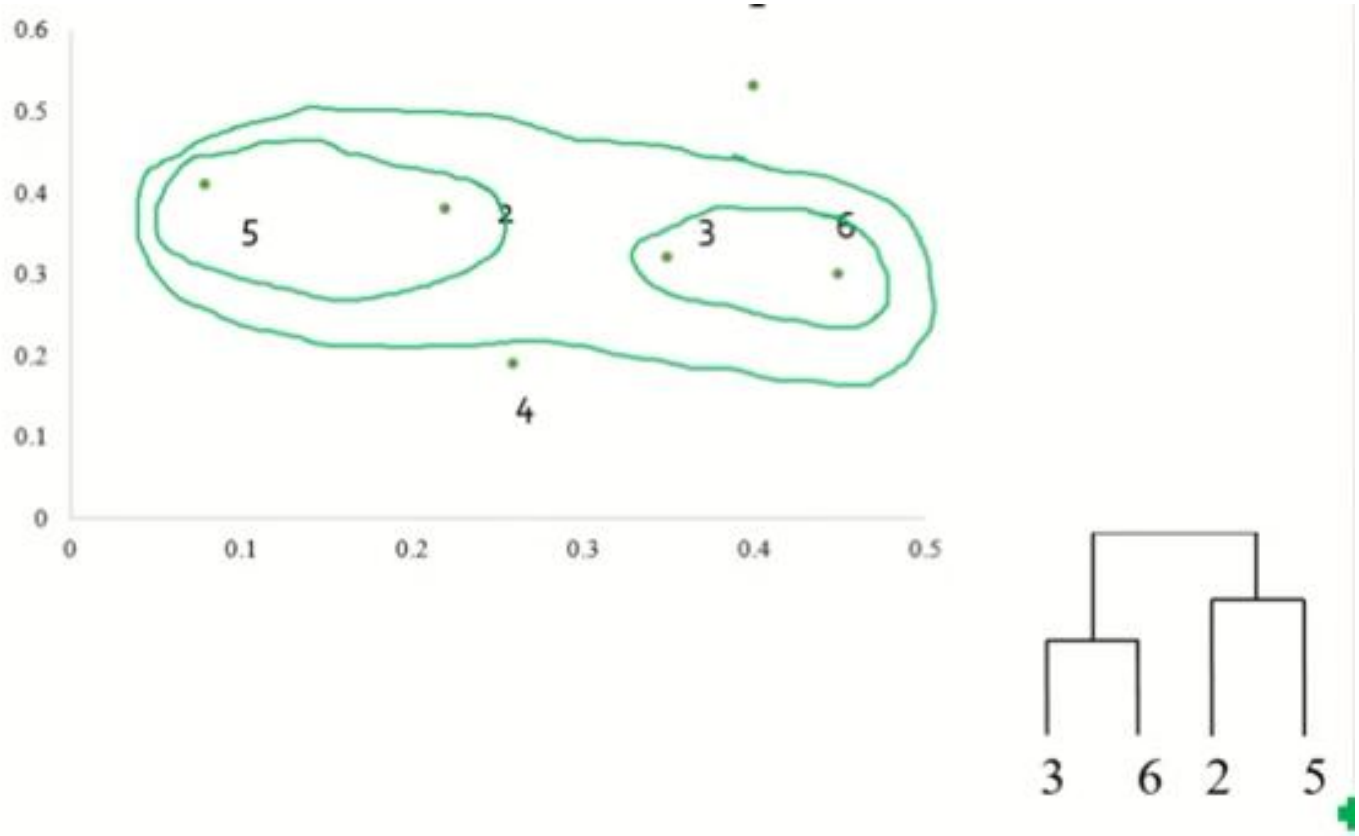
- Min[dist(p2,p5),p1)] = 0.23
- Min[dist(p2,p5),(P3,p6)] = 0.15
- Min[dist(p2,p5),p4)] = 0.20

Next link?

The updated distance matrix for cluster P2,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

New Cluster



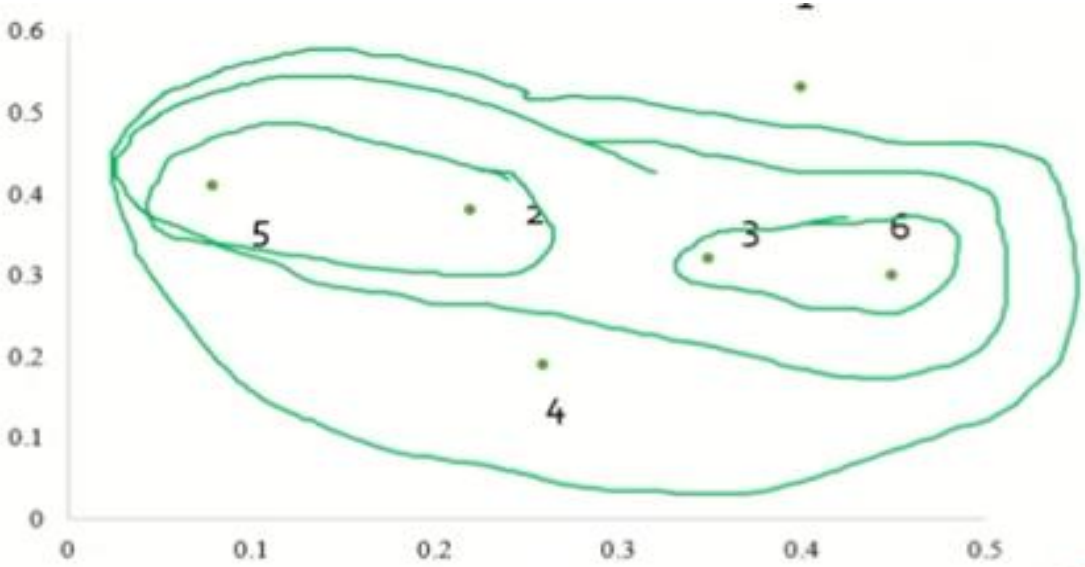
Update Distance vector Matrix

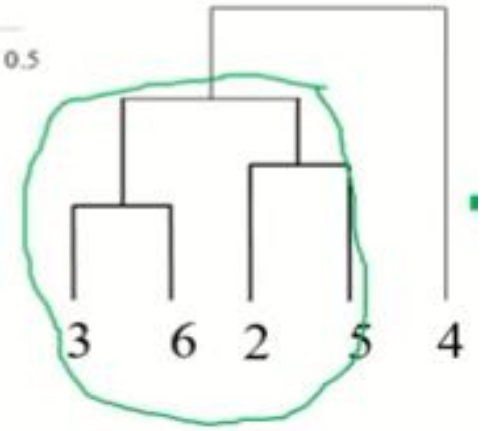
- Min[dist(p2,p5),(p3,p6)),p1] = 0.22
- Min[dist(p2,p5),(p3,p6)),p4] = 0.15

Next Link?

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

New cluster



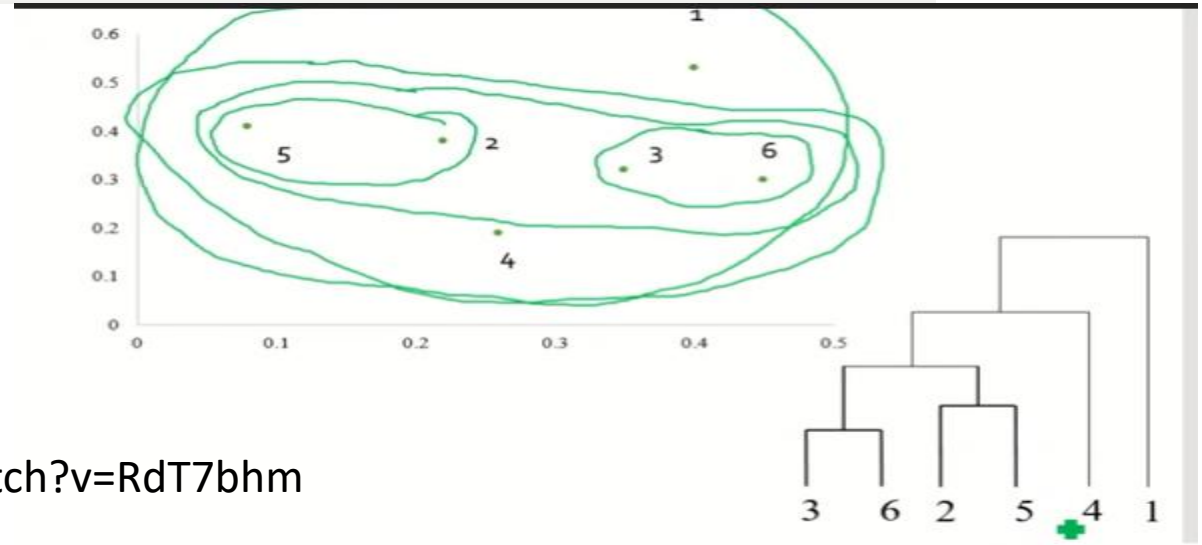


- To update the distance matrix MIN[dist(P2,P5,P3,P6),P4]
- MIN[dist((P2,P5,P3,P6),P1), (P4,P1)]

  = min[(0.22,0.37)]

  = 0.22

	P1	P2,P5,P3,P6,P4
P1	0	
P2,P5,P3,P6,P4	0.22	0



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