Hierarchical agglomerative clustering:

The hierarchical agglomerative clustering uses the bottom-up approaches. In the HAC algorithm starts with every single data point as a single cluster. The similar clusters are successively merged until all clusters have merged into a one cluster and result is represents in tree structure as named dendogram.

Let's understanding with example:

Sample Data:

	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

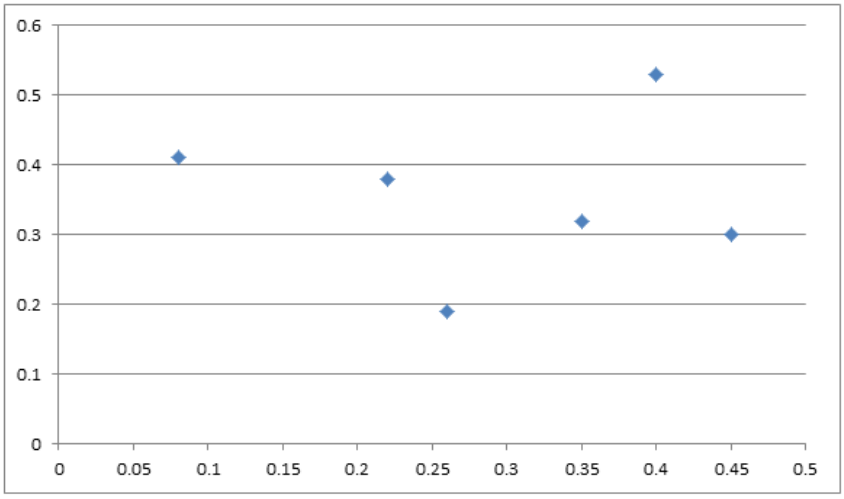
Formulas:

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

$$Min[dist(point1, point2)]$$

Calculation Steps:

Step 1: Draw the graph.



Step 2: Calculate Euclidean distance, create the distance matrix.

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

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

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

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

$$= \sqrt{0.0549}$$

$$= 0.23$$

Same formula can be used for (p1,p3),(p1,p4),(p1,p5),(p1,p6)

$$Distance(p2,p3) = \sqrt{(0.22-0.35)^2 + (0.38-0.32)^2}$$

$$= \sqrt{(-0.13)^2 + (0.06)^2}$$

$$= \sqrt{0.0169 + 0.0036}$$

$$= \sqrt{0.0205}$$

$$= 0.15$$

Same formula can be used for (p2,p4),(p2,p5),(p2,p6)

$$Distance(p3,p4) = \sqrt{(0.35-0.26)^2 + (0.32-0.19)^2}$$

$$= \sqrt{(0.09)^2 + (0.13)^2}$$

$$= \sqrt{0.0081 + 0.0169}$$

$$= \sqrt{0.025}$$

$$= 0.15$$

Same formula can be used for (p3,p5),(p3,p6)

$$Distance(p4,p5) = \sqrt{(0.26-0.08)^2 + (0.19-0.41)^2}$$

$$= \sqrt{(0.18)^2 + (-0.22)^2}$$

$$= \sqrt{0.0324 + 0.0484}$$

$$= \sqrt{0.0808}$$

$$= 0.29$$

Same formula can be used for (p4,p6)

$$Distance(p5,p6) = \sqrt{(0.08-0.45)^2 + (0.41-0.30)^2}$$

$$= \sqrt{(-0.37)^2 + (0.11)^2}$$

$$= \sqrt{0.1369 + 0.0121}$$

$$= \sqrt{0.149}$$

$$= 0.39$$

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

Step 3: Find the minimum value element from distance matrix.

The minimum value element is (p3,p6)and value is 0.11

i.e. our 1st cluster (p3,p6)

	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

Step 4: Recalculate or update the distance matrix for cluster(p3,p6)

$$Min[dist(point1, point2)]$$

$$Min[dist((p3, p6), p1)]$$

$$Min[dist((p3, p1), (p6, p1))]$$

$$Min[dist(0.22, 0.23)]$$

$$= 0.22$$

Same formula can be used for p2,p4,p5

Updated distance matrix:

	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

Step 5: repeat the step 3 & 4.

The minimum value element is (p2,p5)and value is 0.14

i.e. our 2nd cluster (p2,p5)

	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

Recalculate or update the distance matrix for cluster (p2,p5)

$$Min[dist(point1, point2)]$$

$$Min[dist((p2, p5), p1)]$$

$$Min[dist((p2, p1), (p5, p1))]$$

$$Min[dist(0.23,0.34)]$$

$$= 0.23$$

Same formula can be used for (p3,p6),p4

Updated distance matrix:

	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

Step 6: repeat the step 3 & 4.

The minimum value element is (p2,p5,p3,p6) and value is 0.15

Here 2 values are same then first element is choose as minimum value element

i.e. our 3rd cluster (p2,p5,p3,p6)

	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

Recalculate or update the distance matrix for cluster (p2,p5,p3,p6)

$$Min[dist(point1, point2)]$$

$$Min[dist((p2, p5), p1), ((p3, p6), p1)]$$

$$Min[dist(0.23,0.22)]$$

$$= 0.22$$

Same formula can be used for p4

Updated distance matrix:

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

Step 7: repeat the step 3 & 4.

The minimum value element is (p2,p5,p3,p6,p4) and value is 0.15

i.e. our 4th cluster (p2,p5,p3,p6,p4)

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

Recalculate or update the distance matrix for cluster (p2,p5,p3,p6,p4)

$$Min[dist(point1, point2)]$$

$$Min[dist((p2, p5, p3, p6), p1), (p4, p1)]$$

$$Min[dist(0.22,0.37)]$$

$$= 0.22$$

Updated distance matrix:

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

Step 8: repeat the step 3 & 4.

The minimum value element is (p2,p5,p3,p6,p4,p1) and value is 0.22

i.e. our 5th cluster (p2,p5,p3,p6,p4,p1)

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

Recalculate or update the distance matrix for cluster (p2,p5,p3,p6,p4,p1)

In this step only 1 value is remaining so it is by default cluster.

Step 9: Drawing the Dendogram.