Hierarchical / Agglomerative clustering

Example

	x	y
z_1	0.4	0.53
z_2	0.22	0.38
z_3	0.35	0.32
z_4	0.26	0.19
z_5	0.08	0.41
z_6	0.45	0.30

Find the distance matrix by using manhattan distance which is given by distance b/w the points

(p_1, p_2) & (q_1, q_2) is given by

$|p_1 - q_1| + |p_2 - q_2|$

Distance b/w the points z_1 & $z_2 = |0.4 - 0.22| + |0.53 - 0.38|$

$= 0.18 + 0.15$

Distance b/w z_1 & $z_2 = 0.33$

Distance b/w $(z_1, z_3) = |0.4 - 0.35| + |0.53 - 0.32|$

$= 0.05 + 0.21 = 0.26$

Distance b/w $(z_1, z_4) = |0.4 - 0.26| + |0.53 - 0.19|$

$= 0.14 + 0.34 = 0.48$

Distance b/w $(z_1, z_5) = |0.4 - 0.08| + |0.53 - 0.41|$

$= 0.32 + 0.12$

$= 0.44$

Distance b/w $(z_1, z_6) = |0.4 - 0.45| + |0.53 - 0.30|$

$= 0.05 + 0.23 = 0.28$

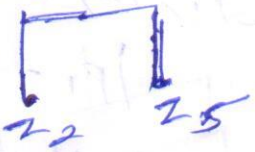
Similarly we calculate for all data points & the distance matrix is given by.

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
Z_1	0	0.33	0.26	0.35	0.44	0.28
Z_2	0.33	0	0.19	0.23	0.01	0.31
Z_3	0.26	0.19	0	0.22	0.36	0.12
Z_4	0.35	0.23	0.22	0	0.4	0.3
Z_5	0.44	0.01	0.36	0.4	0	0.48
Z_6	0.28	0.31	0.12	0.3	0.48	0

Values lying the upper half of the digonal is same as Values lying in the lower half of the digonal.

* The minimum distance is b/w ~~Z_5~~ Z_2 & Z_5 & they both form the cluster.

* The dendogram is drawn to represent the first cluster formation.



Step3 Update the distance matrix using the first cluster (Z_2, Z_5)

There are three methods to do this

1. single link
2. complete link
3. Average link.

* Distance b/w (Z_2, Z_5) & Z_1 is given by using a single link $\Rightarrow$ $\min((Z_2, Z_1), (Z_5, Z_1))$

$= \min(0.33, 0.44)$

$= \underline{0.33}$

* Distance b/w (Z_2, Z_5) & Z_1 is given by using a complete link $\Rightarrow$ $\max((Z_2, Z_1), (Z_5, Z_1))$

$= \max(0.33, 0.44)$

$= \underline{0.44}$

* Distance b/w (Z_2, Z_5) & Z_1 is given by using a Average link $= \frac{1}{2}[\text{dist}(Z_2, Z_1) + \text{dist}(Z_5, Z_1)]$

$= \frac{1}{2}[0.33 + 0.44]$

The updated matrix is given by.

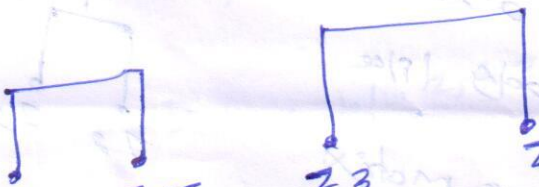
	z_1	z_2, z_5	z_3	z_4	z_6
z_1	0				
z_2, z_5	0.33	0			
z_3	0.26	0.19	0		
z_4	0.35	0.23	0.22	0	
z_6	0.28	0.31	0.12	0.13	0

distance b/w (z_2, z_5) & $(z_3) = 0.19$

dist b/w (z_2, z_5) & $(z_4) = 0.23$

dist b/w (z_2, z_5) & $z_6 = 0.31$

* The minimum distance is b/w z_3 & z_6 & they form the second cluster.



Step 4: Again update the distance matrix with the newly formed clusters.

	z_1	z_2, z_5	z_3, z_6	z_4
z_1	0			
z_2, z_5	0.33	0		
z_3, z_6	0.26	0.19	0	
z_4	0.35	0.23	0.22	0

(i) = distance $((z_2, z_5), (z_3, z_6))$

= mindistance b/w $(z_3 \& z_2, z_5)$, $(z_6 \& z_2, z_5)$

$= (\min(0.19, 0.31)) = 0.19$

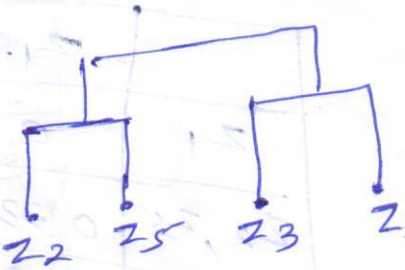
(ii) distance $((z_3, z_6), z_1) = \min(0.26, 0.28)$

dist ((z3, z6), z4)

= min (0.22, 0.3)

= 0.22

* The dendogram now looks like.



Step 5 The updated distance matrix is again formed.

	z1	z2, z5, z3, z6	z4
z1	0		
z2, z5, z3, z6	0.26	0	
z4	0.35	0.22	0

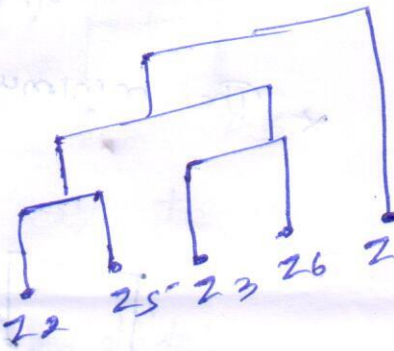
dist (z1, z2 z5 z3 z4) = 0.26

= min (0.33, 0.26)

dist (z4, z2 z5 z3 z4) = 0.22

= min (0.23, 0.22

* Now, dendogram looks like



Step 6 update distance matrix

	z1	z2 z5 z3 z6 z4
z1	0	0.33
z2 z5 z3 z6 z4	0.33	0

distance b/w (z1, (z2 z5 z3 z6) z4)

= min (0.33, 0.35)

= 0.33