AGGLOMERATIVE CLUSTERING

To form clusters we need to check the distances between points by using distance calculator. So always standardize your data first before calculating distance between points.

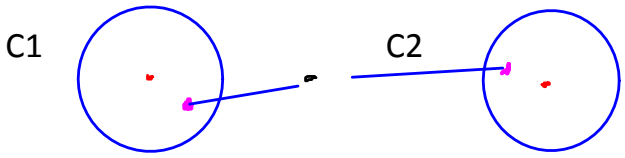
Euclidian Distance:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

To calculate Euclidian distance method we need to follow some linkage methods.

There are :
- Single Linkage
- Complete Linkage
- Average Linkage
- Centroid Linkage
- Median Linkage
- Ward Linkage

• **Single Linkage (Nearest Neighbourhood Distance/Minimum distance):**



Lets work on example on Single Linkage method

The distance matrix is given by as follows.

	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	2	8	0

Let's assume that we already calculated distance from some data set and above are those values. Assuming that each individual observation is a cluster point. Now we are reducing to minimum clusters.

Step 1: Take out the minimum distance from all the above points. d(5,3) = 2.

Now, I will form (5,3) in to one group and rest of the observations are separate clusters as 1,2,4

	(3,5)	1	2	4
(3,5)	0			
1	3	0		
2	7	9	0	
4	8	6	5	0

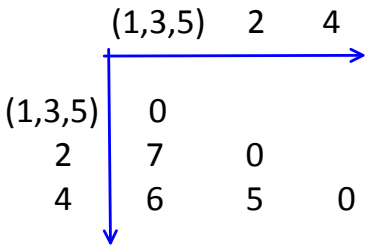
Find the distance between d (3,5) and 1
Let's check distance between (3 and 1) and (5 and 1) = min(3,11)=3

Find the distance between d (3,5) and 2
Let's check distance between (3 and 2) and (5 and 2) = min(7,10)=7

Find the distance between d (3,5) and 4
Let's check distance between (3 and 4) and (5 and 4) = min(9,8)=8

Step 2: Take out the minimum distance from all the above points. d(1,(3,5)) = 3.

Now, I will form (1,3,5) in to one group and rest of the observations are separate clusters as 2,4



Find the distance between d (1,3,5) and 2

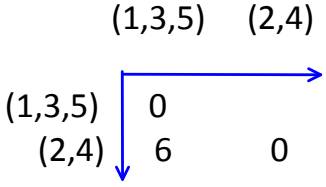
Let's check distance between (1 and 2) and (3 and 2) and (5 and 2) = min(9,7,10)=7

Find the distance between d (1,3,5) and 4

Let's check distance between (1 and 4) and (3 and 4) and (5 and 4) = min(6,9,8)=6

Step 3: Take out the minimum distance from all the above points. d(2,4)) = 5.

Now, I will form (1,3,5) in to one group and (2,4) as another group

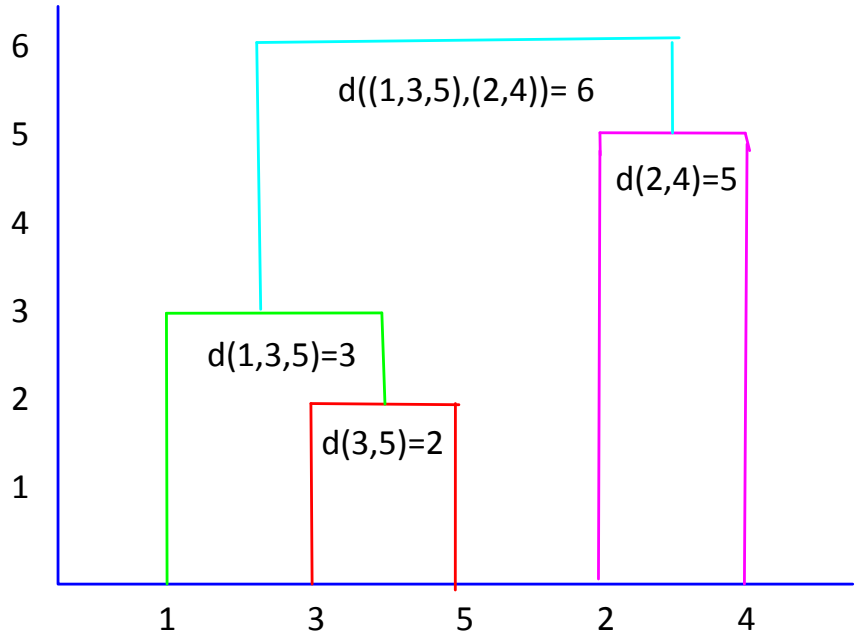


Find the distance between d (1,3,5) and (2,4)

Let's check distance between d(1,2), d(1,4), d(3,2), d(3,4), d(5,2), d(5,4)

Min(9,6,7,9,10,8) = 6

Now, the process will be stopped and we need to construct the "**Dendogram**"


6
5
4
3
2
1
d((1,3,5),(2,4))= 6
d(2,4)=5
d(1,3,5)=3
d(3,5)=2
1
3
5
2
4