**ISOM 2600 Introduction to Business Analytics**

**Weekly Exercise 4**

Q1-Q2: The table below summarizes the attributes of members from two clusters:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Member ID | X1 | X2 | X3 | Cluster ID |
| 1 | 5 | 6 | 4 | 1 |
| 2 | 1 | 6 | 7 | 2 |
| 3 | 6 | 5 | 3 | 1 |
| 4 | 2 | 7 | 6 | 2 |
| 5 | 4 | 4 | 3 | 1 |

Q1:

What is the distance between Member#1 and Member#3?

1. 3.7417
2. 1.7321
3. 6.4807
4. 4.5826

Solution:

Q2:

What is the single linkage of the two clusters if the distance is defined by Euclidean distance.

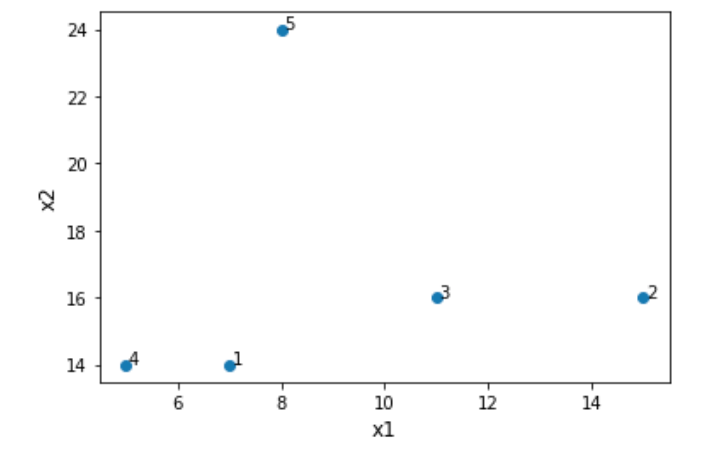
1. 3.7417
2. 1.7321
3. 6.4807
4. 5

Solution:  
The distances between all the possible pairs of the two clusters are given below:

|  |  |  |
| --- | --- | --- |
| Cluster 1 Member | Cluster 2 Member | distance |
| 1 | 2 | 5 |
| 3 | 2 | 6.4807 |
| 5 | 2 | 5.3852 |
| 1 | 4 | 3.7417 |
| 3 | 4 | 5.3852 |
| 5 | 4 | 4.6904 |

As the minimum distance of above is 3.7417, so the single linkage between the two clusters is 3.7417.

Q3-Q5: Five data points are visualized in the following scatter plot:



Q3: According to the scatter plot, which two data points will merge first when we use hierarchical clustering? (no need to consider standardization)

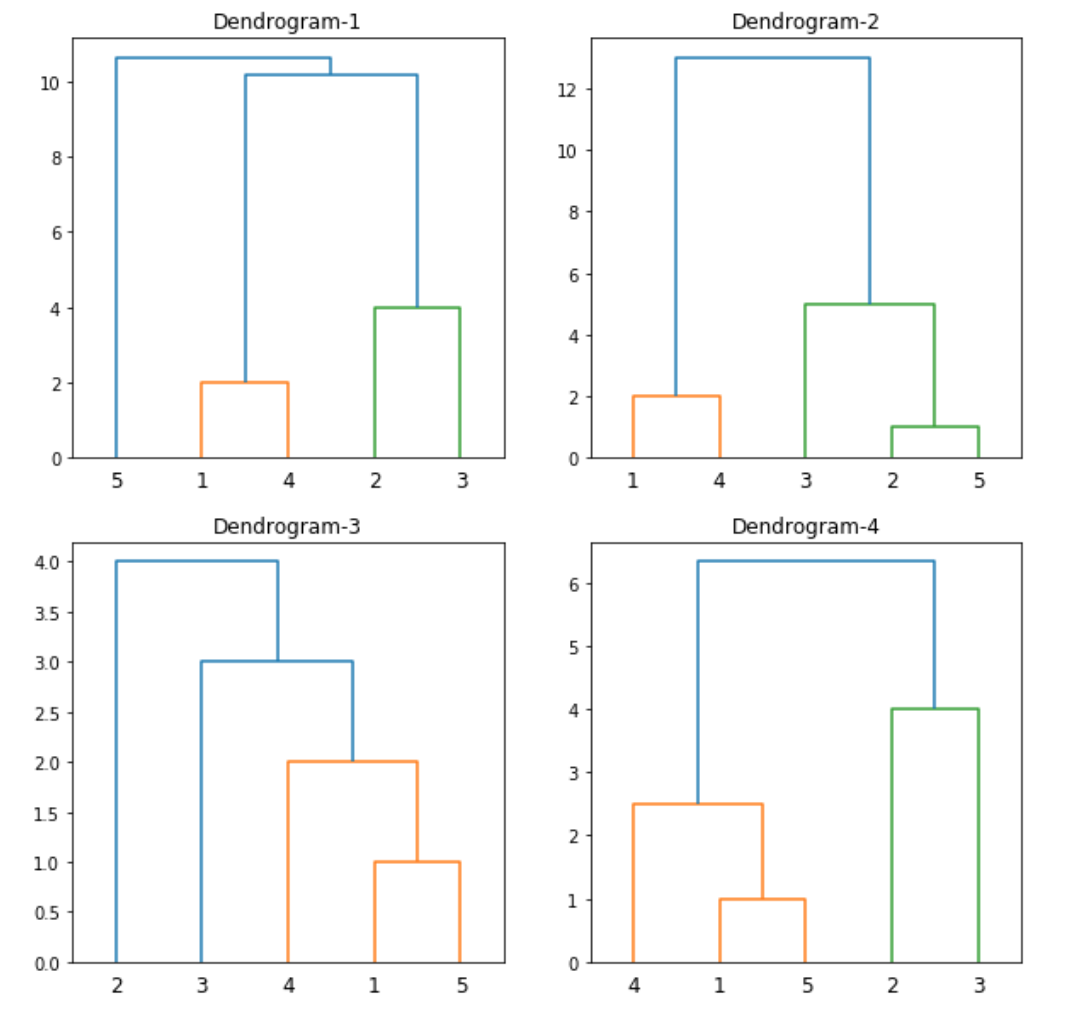
1. 2 and 3
2. 3 and 5
3. 1 and 4
4. 2 and 5

Solution: As the distance between 1 and 4 is the shortest among all the pairs, so 1 and 4 will merge together first.

Q4: Suppose there is one more attribute (x3) is given on the dataset, and each data point has the same value (a very large value) in x3. Which two data points will merge first when we use hierarchical clustering? (no need to consider standardization)

1. 2 and 3
2. 3 and 5
3. 1 and 4
4. 2 and 5

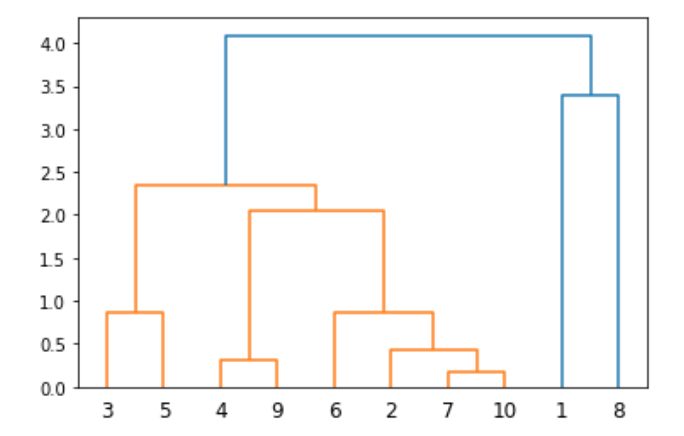
Solution: Since they share the same value in x3, so x3 is not a useful variable and has no effect in the distance at all, so everything remains the same.

Q5: Suppose we only use x1 to perform hierarchical clustering with centroid linkage, which is the corresponding dendrogram? (no need to consider standardization)  


1. Dendrogram-1
2. Dendrogram-2
3. Dendrogram-3
4. Dendrogram-4

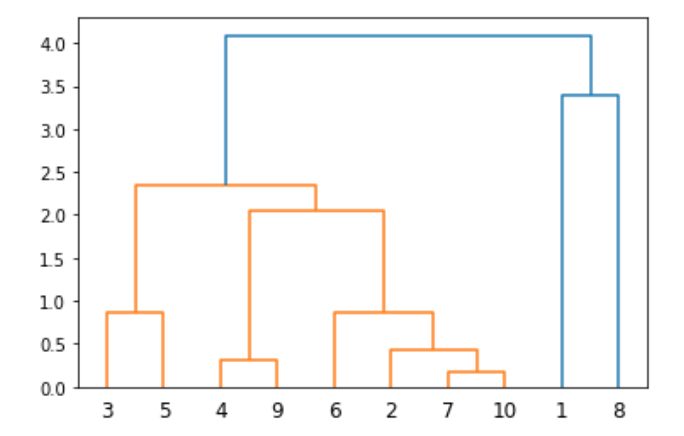
Solution: #1 and #5 merge into a single cluster (say cluster#1) first since they are the closest, and their centroid is 7.5; Then, #4 will merge to cluster#1 since it is closest to cluster’s centre then the individual #3 and #2. The updated cluster#1’s centroid is 6.67; The next merge will happen between #2 and #3 since #2 is a bit closer to #3 rather than the centroid of cluster#1. As a result, only Dendrogram-4 matches all the moves of above.

Q6: If we decide to form 7 clusters from the dendrogram below, which one is the neighbour to #6?



1. #2
2. #2,#7 and #10
3. #4, #9, #2,#7 and #10
4. No neighbour

Solution:



#6 forms a single cluster on its own.