**Homework 7**

1. **Describe the goal of all clustering methods.**

**Ans –** Clustering is a methosd of grouping records based on their classes or similarity between different objects.   
A cluster is collection of records that are similar and dissimilar to records in other clusters. In Clustering we donot have any target varaiable. The clustering rask does not classigy , estimate or predict value of target variable. Hence the clustering algorithms tend to segment or categorize or group data set into relatively homogenous subgroups or clusters.

Hence clustering can help in cases like target marketing, financial behaviour for accounting and auditing purposes, also might help in dimension reduction tool where data set has more number of attributes.

1. **Suppose that we have the following data (one variable). Use single linkage to identify the clusters.**

**Data: 0 0 1 3 3 6 7 9 10 10**

**Ans – Single linkage** agglomerative clustering  **to identify cluster**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 0 | 1 | 3 | 3 | 6 | 7 | 9 | 10 | 10 |

|  |  |
| --- | --- |
| 0 | 0 |

|  |  |
| --- | --- |
| 3 | 3 |

|  |  |
| --- | --- |
| 10 | 10 |

|  |  |
| --- | --- |
| 6 | 7 |

|  |
| --- |
| 9 |

Step 1:

|  |
| --- |
| 1 |

Step 2:

Combining clusters with minimal distances first-- { 0, 0} {1} and {9} {10, 10}

Resultant clusters –

{0,0,1} {3,3} {6,7} {9,10,10}

Step 3:   
Combining clusters with next minimal distances here-- |1 – 3| = 2 , |7-9| = 2   
{0, 0, 1, 3, 3} {6,7, 9,10,10}

Step 4 :   
Combining both the clusters now since the difference is |3-6| = 3

Resultant final single cluster – {0,0,1,3,3,6,7,9,10,10}

1. **Suppose that we have the following data (one variable). Use complete linkage to identify the clusters.**

**Data: 0 0 1 3 3 6 7 9 10 10**

**Ans – Complete linkage to identify the clusters**

**Step 1 –**

**{0, 0} {1} {3,3} {6} {7} {9} {10,10}**

**Step 2 – By complete linkage method distance between lowest of first and preceeding cluster should be lowest or minimal. Following pairs can be merged at this stage  
1. {0,0}{1} {3,3} 2. {6} {7} 3. {9} {10,10}  
Resultant cluster solution – {0,0,1} {3,3} {6, 7} {9, 10,10}**

**Step 3 – Next distance between – 0 & 3 – 3 is the lowest**

**Resultant cluster solution – {0,0,1, 3,3} {6, 7} {9, 10,10}**

**Step 4- Next distance between 0 and 7 -- 7 and distance between 6 and 10 is 4, hence merging these two clusters – {6, 7} and {9, 10,10}  
Resultant cluster –  
 {0,0,1,3,3} {6, 7, 9, 10, 10 }**

**Step 5 – Merging the last two clusters generated by complete linkage distance method –**

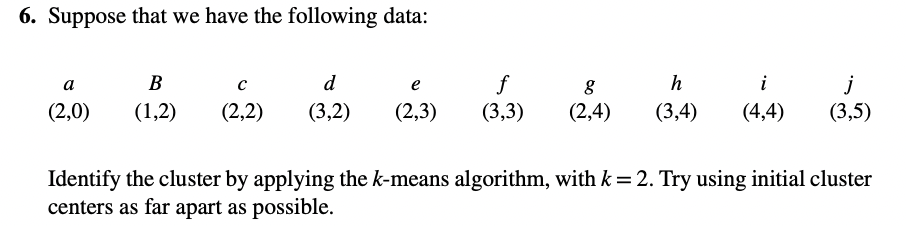
**{0, 0, 1, 3, 3, 6, 7, 9, 10, 10 }**

1. **What is an intuitive idea for the meaning of the *centroid* of a cluster?**

**Ans – Centroid of a cluster –**

Centroid of any object is the calculated center of the different values in a cluster, it represents a balance of a cluster and helps to define a distance between two clusters. Hence for each clusters in the network, the location of the center to the new centrouid varies as the values or the cluster member change. If there are no record shifted the cluster membership, and the cluster centroid remains unchanged.

Suppose that there are a few records that have been moved from cluster 1 too2 n account of its resemblance or relevance to the cluster properties, the resultant cluster centroid tends to move too with the change in cluster member values. Centroid is the Euclidean distance between the different members of a cluster or their values to the value considered as center of a cluster.



Ans –

K=2

Defining initial cluster centers - b(1,2) and i(4,4)

Graphical user interface, application, table, Excel

Description automatically generated

This has formed two clusters – C1 and C2 with following members

* C1 = { B,C, D, E}
* C2 = { A, F, G, H, I, J}

Where k =1 and initial cluster set are considered as. B & I with distant values.