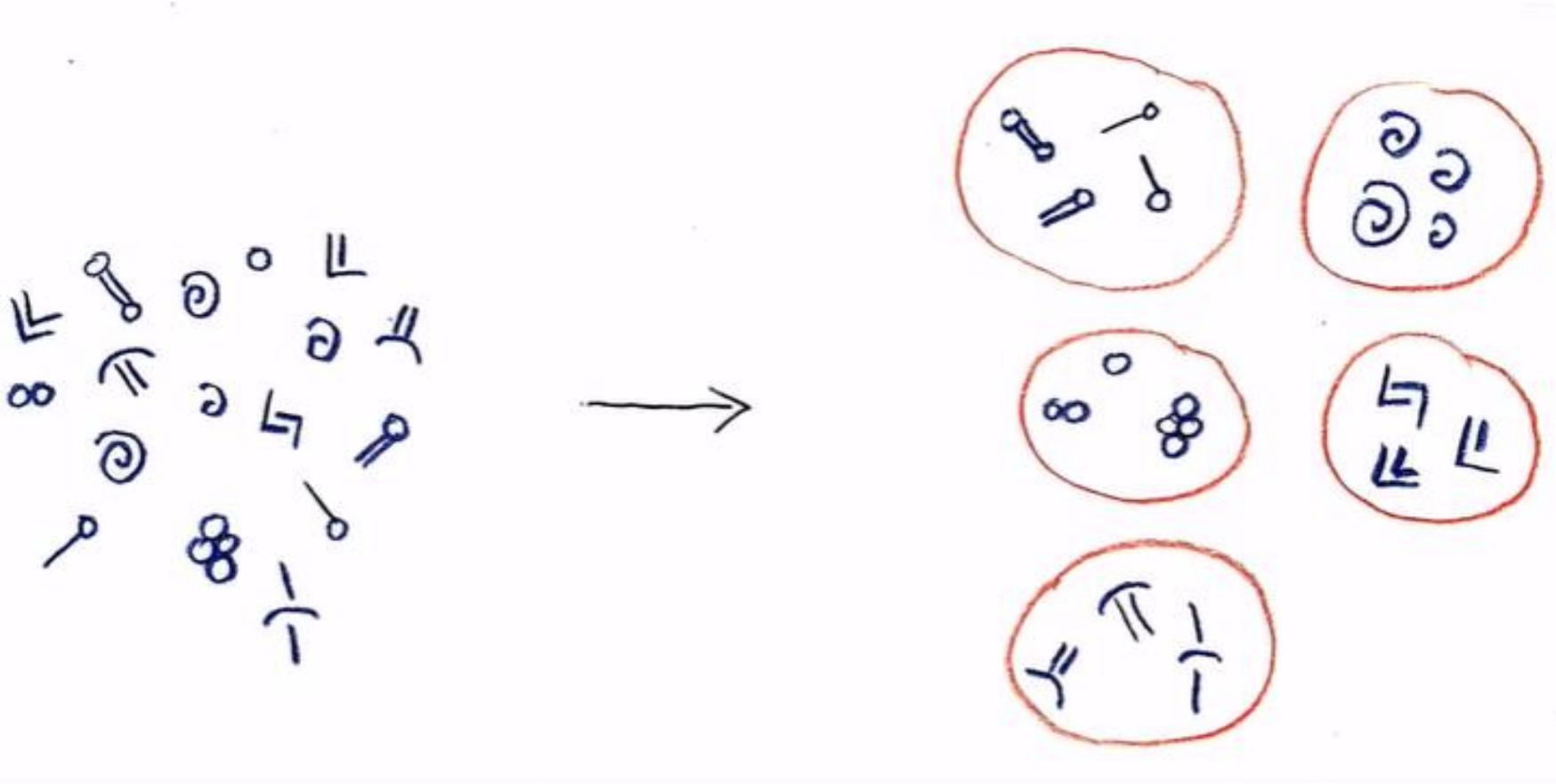
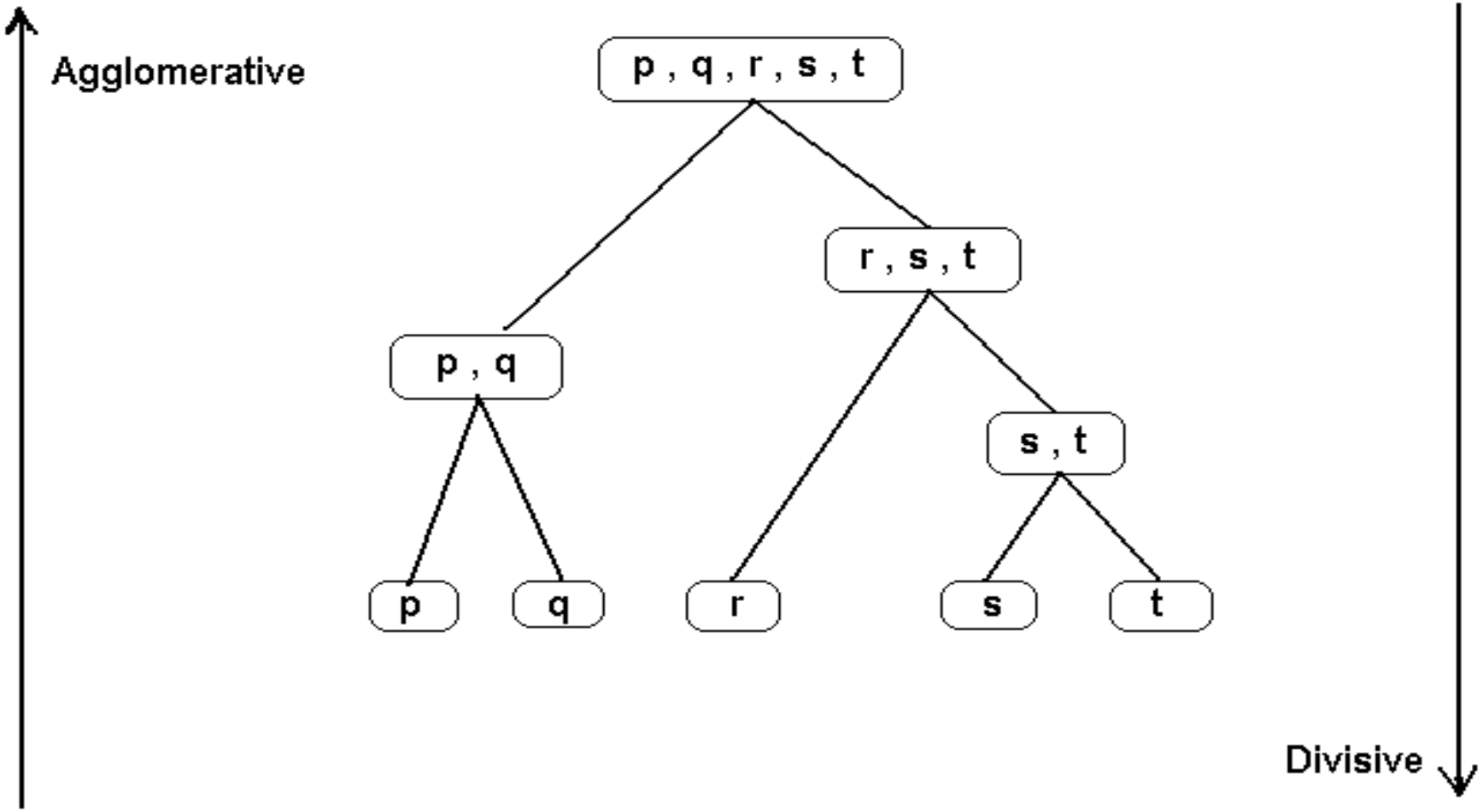
CLUSTERING



Hierarchical Clustering



K-means Clustering - Example

- Given a cluster $K_i=\{t_{i1},t_{i2},\ldots,t_{im}\}$, let the ***cluster mean*** be $m_i = (1/m)(t_{i1} + \ldots + t_{im})$

Given: {2,4,10,12,3,20,30,11,25}, k=2

- Randomly pick some initial means: $m_1=3$, $m_2=4$
- $K_1=\{2,3\}$, $K_2=\{4,10,12,20,30,11,25\}$, $m_1=2.5$, $m_2=16$
- $K_1=\{2,3,4\}$, $K_2=\{10,12,20,30,11,25\}$, $m_1=3$, $m_2=18$
- $K_1=\{2,3,4,10\}$, $K_2=\{12,20,30,11,25\}$, $m_1=4.75$, $m_2=19.6$
- $K_1=\{2,3,4,10,11,12\}$, $K_2=\{20,30,25\}$, $m_1=7$, $m_2=25$

Stop as the clusters with these means are the same.

K-means Clustering

- Partitional clustering approach
- Each cluster is associated with a centroid (center point)
- Each point is assigned to the cluster with the closest centroid
- Number of clusters, K, must be specified
- The basic algorithm is very simple

1: Select K points as the initial centroids.
2: **repeat**
3: Form K clusters by assigning all points to the closest centroid.
4: Recompute the centroid of each cluster.
5: **until** The centroids don't change

Data : { 2,3,4,10,11,12,20,25,30}

m1=4

m2=12

k1={2,3,4}

k2={10,11,12,20,25,30}

m1=3

m2=18

k1={2,3,4,10}

k2={11,12,20,25,30}

m1=4.75~5

m2=19.6

k1={2,3,4,10,11,12}

k2={20,25,30}

m1=7

m2=25

k1={2,3,4,10,11,12}

k2={20,25,30}

K-Means Algorithms

- Step 1: Take Mean Value of Each Cluster

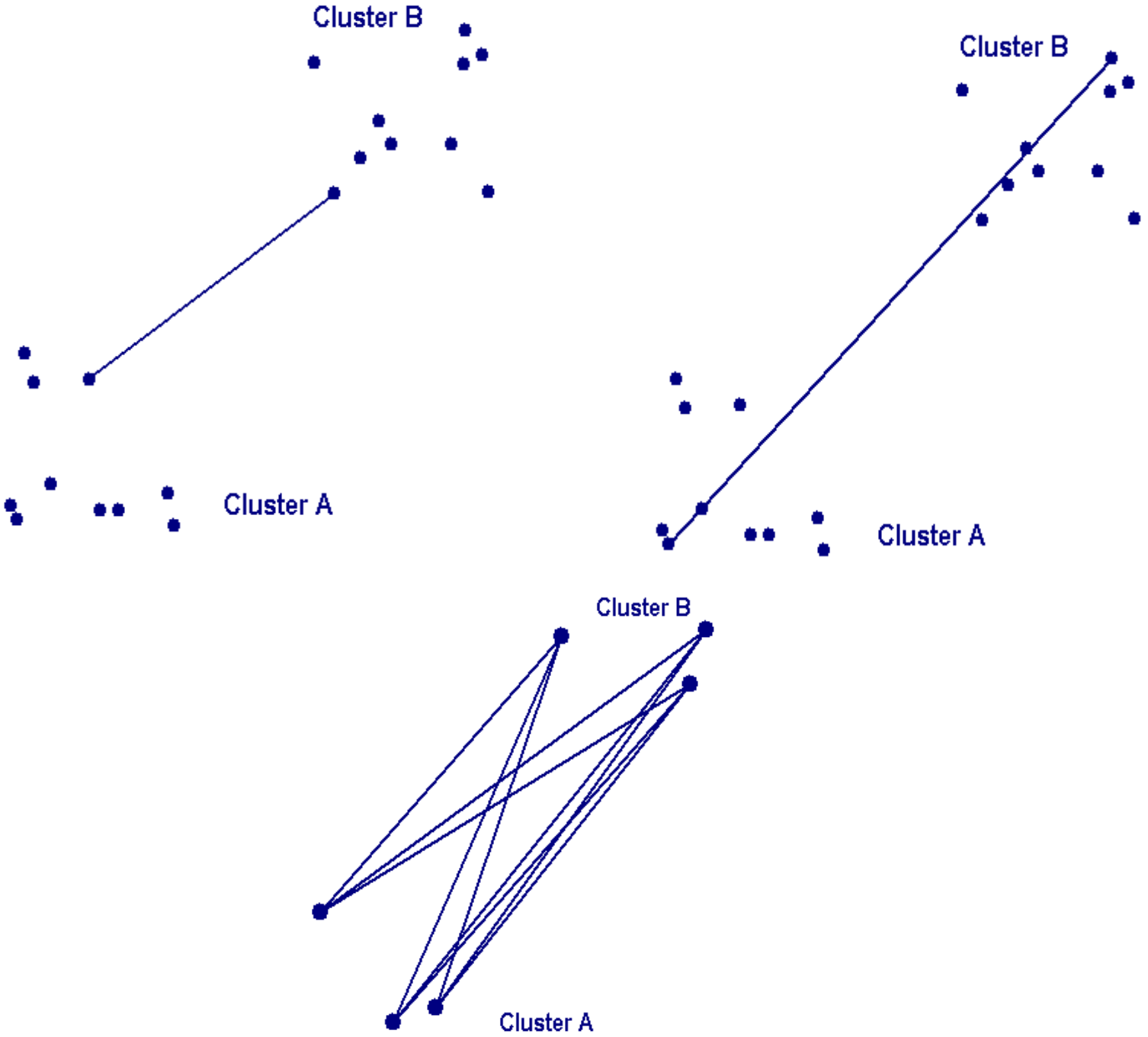
  (Random value for First Time)

- Step 2: Find nearest number of mean and put it in the cluster

- Step 3: Repeat Step 1 and 2 until we get same mean .

Point Location		Distance to Point	
X	Y	Red cluster	Green Cluster
9	1		
8	4		
7	2		
2	3		
1	5		
3	5		
7	5		

$$d_{1,2} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$


Cluster B
Cluster A
Cluster B
Cluster A
Cluster B
Cluster A