Divisive Clustering

Dr. Bhargavi
SCOPE
VIT Chennai

DIANA - Divisive Analysis Clustering

- DIANA (Divisive Analysis) Clustering, is a hierarchical clustering method.
- Follows divisive clustering approach.
- Works in a “top-down” manner.
- Starts with all data points in a single cluster and successively divides them into smaller clusters.
- At each step, DIANA splits the largest or most heterogeneous cluster (the one with the greatest internal dissimilarity) into two smaller clusters.
- The method looks for the data point that is the most different from the others and separates them from the cluster. This creates two new clusters.
- The splitting is continued until all clusters are made up of single data points or until a stopping criterion (like a predefined number of clusters) is met.

DIANA (Cont…)

- Divisive clustering **starts with one, all-inclusive cluster**.
- At each step, it **splits a cluster until each cluster contains a point**.
- Example:

Distance	a	b	c	d	e
a	0	2	6	10	9
b	2	0	5	9	8
c	6	5	0	4	5
d	10	9	4	0	3
e	9	8	5	3	0

Cont…

Step 1: Split entire data into 2 clusters

- Distance of a to others: mean(2,6,10,9) = 6.75
- Distance of b to others: mean(2,5,9,8) = 6.0
- Distance of c to others: mean(6,5,4,5) = 5.0
- Distance of d to others: mean(10,9,4,3)= 6.5
- Distance of e to others: mean(9,8,5,3)=6.25

- Since the data a is far from all others (in average) it makes a new cluster

Cont…

- Now check every point if it is closer to new cluster or the existing cluster and reassign the cluster if it is closer to new cluster.

	P (distance from old cluster)	Q (distance from new cluster)	P - Q	
b	(5+9+8)/3 = 7.33	2	> 0	b joins
c	(5+4+5)/3 = 4.67	6	< 0	No change
d	(9+4+3)/3 = 5.33	10	< 0	No change
e	(8+5+3)/3 = 5.33	9	< 0	No change

Cont…

- Above process is repeated again and again until the data points in the old cluster do not change

	P (distance from old cluster)	Q (distance from new cluster)	P - Q	
c	(4+5)/2 = 4.5	(6+5)/2 = 5.5	< 0	No change
d	(4+3)/2 = 3.5	(10 +9)/2 = 9.5	< 0	No change
e	(5+3)/2 = 4	(9+8)/2 = 8.5	< 0	No change

Cont…

- **Step 2.** Choose a current cluster and split it as in **Step 1.**
- If we want to split the cluster with the largest number of members, then the cluster {c,d,e} will be split.
- If we want to split the cluster with the largest diameter, then the cluster {c,d,e} will be split.
  - Diameter of the cluster {a,b} = 2
  - Diameter of the cluster {c,d,e} = 5
- Split the chosen cluster as in Step 1.
- **Step 3.** Repeat Step 2. until each cluster contains a point (or there are k clusters)