Slide-1

H clustering Hands-on:

The Hierarchical Clustering Algorithm:

Start with n clusters (record =cluster)

Step 1: Two closet records are merged into one cluster.

At every step, pair of clusters with smallest distance are merged.

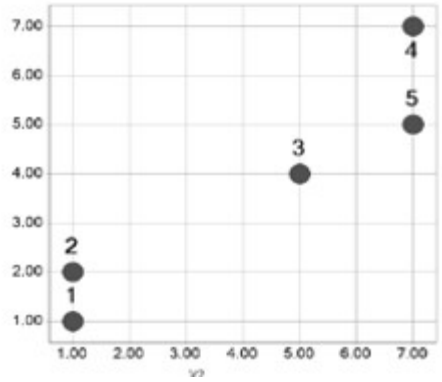
At this point, the distance matrix is re-computed.

- Two rows+columns are merged into single row+column.
- Distance to the newly merged clusters are recalculated.
- Repeat the last step until a single cluster is formed.

Slide-2

Two variables, n= 5 items

item	v1	v2
1	1	1
2	2	1
3	4	5
4	7	7
5	5	7



	1	2	3	4	5
1	0.0				
2	1.0	0.0			
3	5.0	4.5	0.0		
4	8.5	7.8	3.6	0.0	
5	7.2	6.7	2.2	2.0	0.0

Euclidean matrix

What happens next?

. Merge 1&2 into cluster A.

. Use single linkage to calculate distances from cluster A.

Slide-3

What happens next?

	A	3	4	5
A	0.0			
3	4.5	0.0		
4	7.8	3.6	0.0	
5	6.7	2.2	2.0	0.0

→

	A	3	4	5
A	0.0			
3	4.5	0.0		
4	7.8	3.6	0.0	
5	6.7	2.2	2.0	0.0

Slide-4

Merge 4&5 (Cluster B)

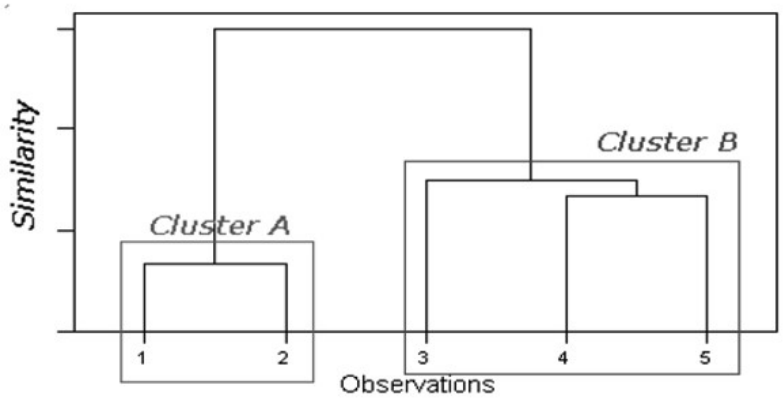
	A	3	B
A	0.0		
3	4.5	0.0	
B	6.7	2.2	0.0

Merge 3& B

	A	B
A	0.0	
B	4.5	0.0

Slide-5

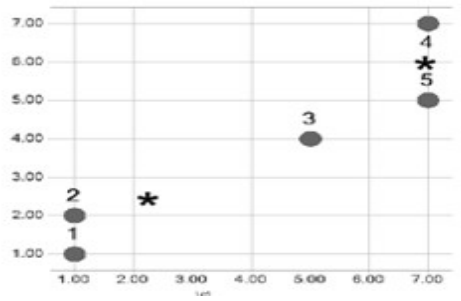
Finally: Summarize process in a Dendrogram



SLIDE-6

Example k=2

item	v1	v2
1	1	1
2	2	1
3	4	5
4	7	7
5	5	7



Start with cluster A:1,2,3 and with cluster B: 4,5

Compute cluster centroids.

SLIDE-7

What are the centroids of the cluster A & B?

item	v1	v2
1	1	1
2	2	1
3	4	5
4	7	7
5	5	7

1) A=(1,1.5,4.5) and B=(7,6)
2) A=(2.33) and B=(6.5)
3) A=(2.33,2.33) and B=(6,7)

Slide-8

Compute Euclidean distance of each record of from each centroid, and re-assign to closest cluster.

	Cluster A	Cluster B
Item 1	$\sqrt{(1-2.33)^2+(1-2.33)^2} = 1.89$	$\sqrt{(1-6)^2+(1-7)^2} = 7.81$
Item 2	1.37	7.21
Item 3	$\sqrt{(4-2.33)^2+(5-2.33)^2} = 3.14$	$\sqrt{(4-6)^2+(5-7)^2} = 2.83$
Item 4	6.60	1
Item 5	5.37	1

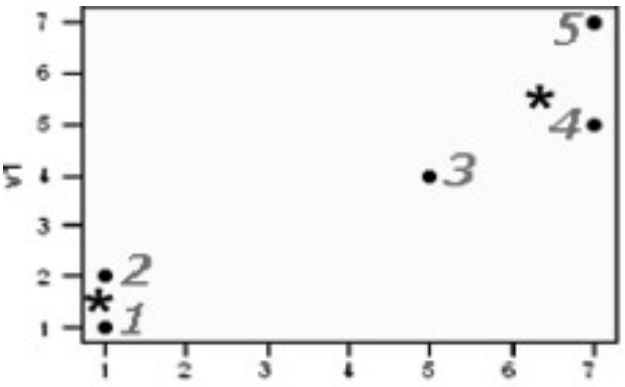
Slide-9

First iteration results:

Cluster A:1,2 and Cluster B: 3,4,5

Re-Compute centroids:

Centroid(A) = (1.5,1) Centroid(B) = (5.33,6.33)



Slide-10

Re-compute distances of records to centroids

	Cluster A	Cluster B
Item 1	$\sqrt{(1-1.5)^2 + (1-1)^2} = 0.5$	$\sqrt{(1-5.33)^2+(1-6.33)^2} = 6.87$
Item 2	0.5	6.29
Item 3	$\sqrt{(4-1.5)^2+ (5-1)^2} = 4.72$	$\sqrt{(4-5.33)^2+ (5-6.33)^2} =1.89$
Item 4	8.14	1.80
Item 5	6.95	0.75