**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.

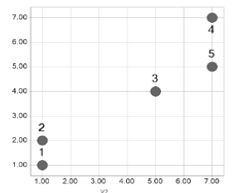
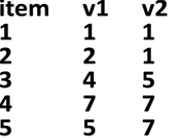
At every step, pair of clusters with smallest distance are merged.

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

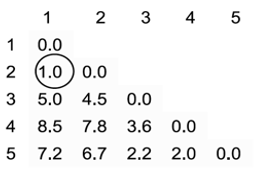
* Two rows+columns are merged into single row+column.
* Distance to the newly merged clusters are recalculated.
* Repeat the last step until a single cluster is formed.

**Slide-2**

Two variables, n= 5 items



**Euclidean matrix**



What happens next?

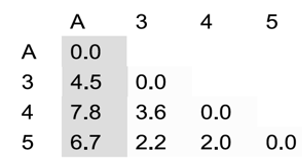
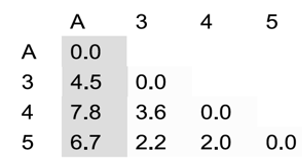
. Merge 1&2 into cluster A.

. Use single linkage to calculate distances from cluster A.

**Slide-3**

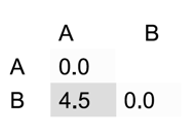
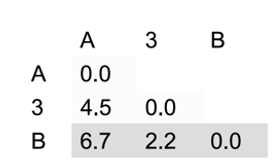
**What happens next?**

**🡪**



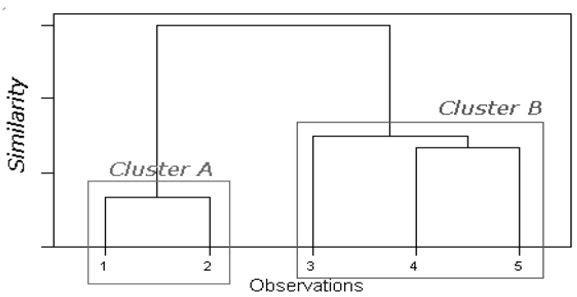
**Slide-4**

**Merge 4&5 (Cluster B) Merge 3& B**



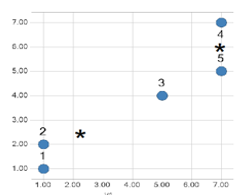
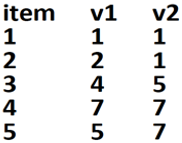
**Slide-5**

**Finally: Summarize process in a Dendrogram**



**SLIDE-6**

**Example k=2**

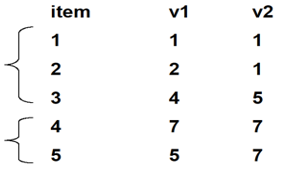


**Start with cluster A:1,2,3 and with cluster B: 4,5**

**Compute cluster centroids.**

**SLIDE-7**

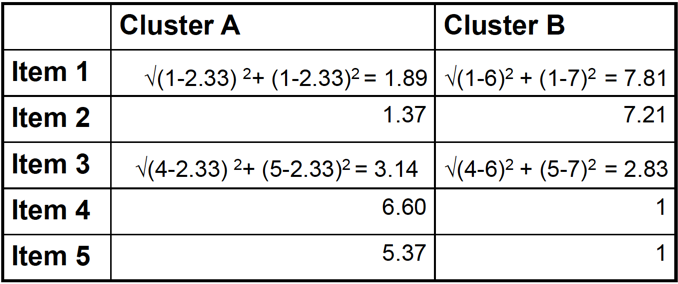
**What are the centroids of the cluster A & B?**



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.**



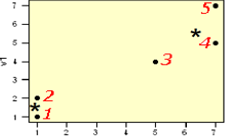
**Slide-9**

**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**

