COMP7/8118 M50

Data Mining

Hierarchical Clustering

Xiaofei Zhang

Slides compiled from Jiawei Han and Raymond C.W. Wong's work

THE UNIVERSITY OF MEMPHIS

Hierarchical Clustering Methods

- The partition of data is not done at a single step.
- There are two varieties of hierarchical clustering algorithms
  - Agglomerative – successively fusions of the data into groups
    - Start with the points as individual clusters
    - At each step, merge the closest pair of clusters until only one cluster (or k clusters) left
  - Divisive – separate the data successively into finer groups
    - Start with one, all-inclusive cluster
    - At each step, split a cluster until each cluster contains a point (or there are k clusters)

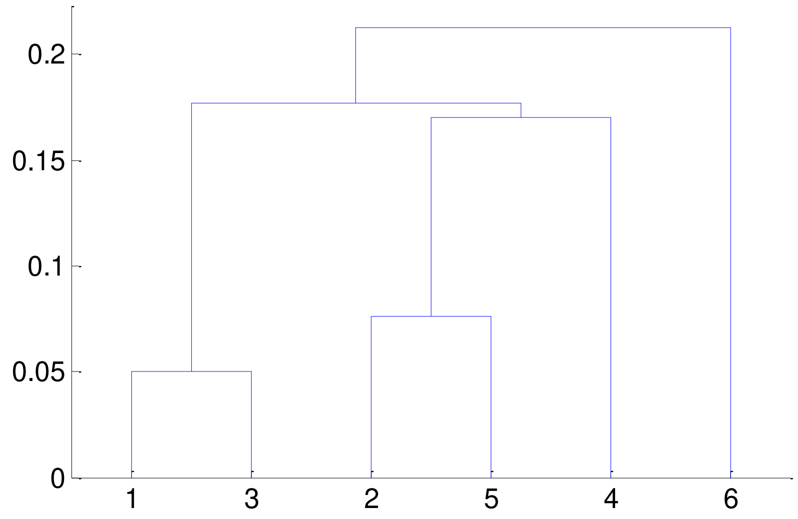
Strengths of Hierarchical Clustering

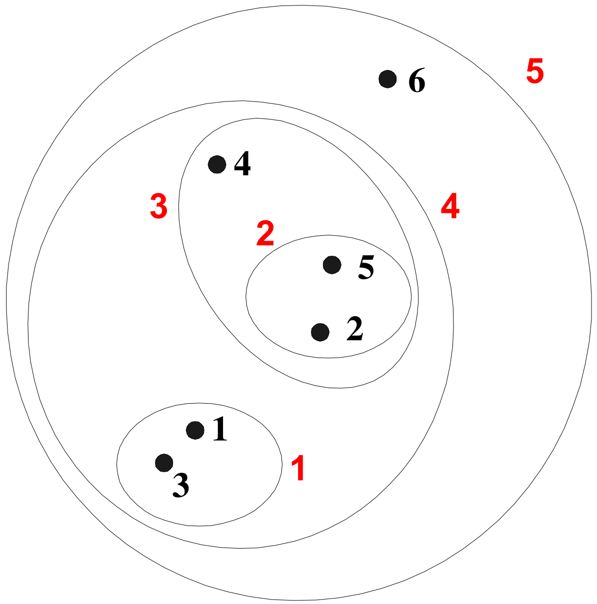
- Do not have to assume any particular number of clusters
  - Any desired number of clusters can be obtained by ‘cutting’ the dendrogram at the proper level

- They may correspond to meaningful taxonomies
  - Example in biological sciences (e.g., animal kingdom, phylogeny reconstruction, …)

Dendrogram

- Hierarchic grouping can be represented by two-dimensional diagram known as a **dendrogram**.



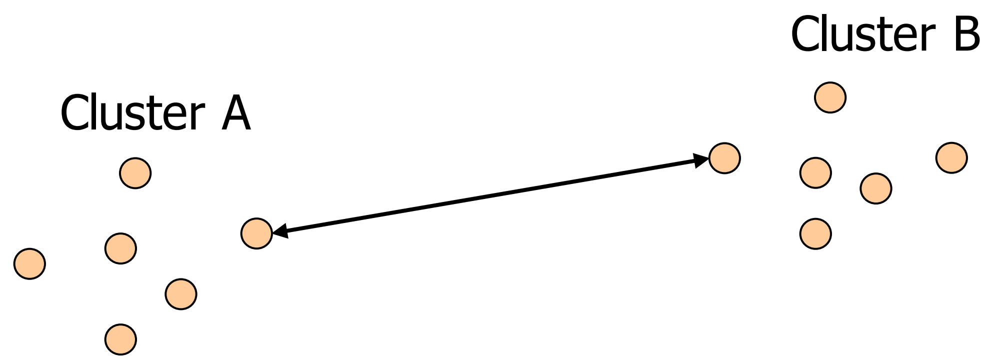


Distance of Clusters

- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

Single Linkage

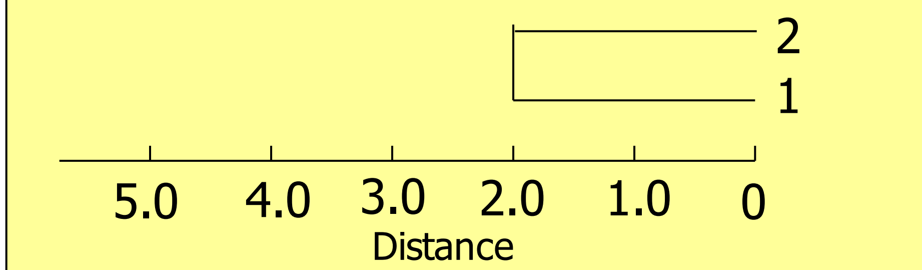
- Also, known as the **nearest neighbor** technique
- Distance between groups is defined as that of the closest pair of data, where only pairs consisting of one record from each group are considered



Example with Single Linkage

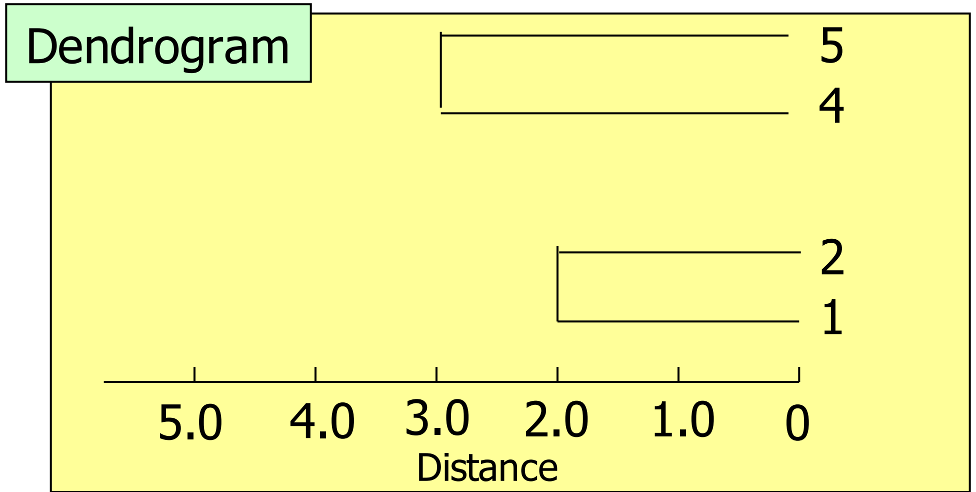
$$
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 0.0 & & & & \\
2 & 2.0 & 0.0 & & & \\
3 & 6.0 & 5.0 & 0.0 & & \\
4 & 10.0 & 9.0 & 4.0 & 0.0 & \\
5 & 9.0 & 8.0 & 5.0 & 3.0 & 0.0
\end{array}
\Rightarrow
\begin{array}{c|cccc}
 & (12) & 3 & 4 & 5 \\
\hline
(12) & 0.0 & & & \\
3 & 5.0 & 0.0 & & \\
4 & 9.0 & 4.0 & 0.0 & \\
5 & 8.0 & 5.0 & 3.0 & 0.0
\end{array}
$$

Dendrogram



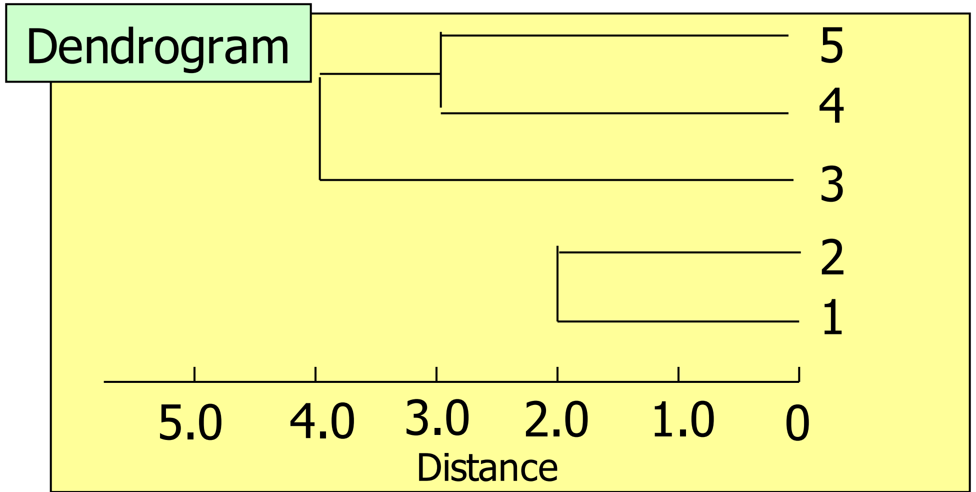
Example with Single Linkage

$$\begin{array}{c|c} & \begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\ \begin{array}{c} (12) \\ 3 \\ 4 \\ 5 \end{array} & \begin{pmatrix} 0.0 & & & \\ 5.0 & 0.0 & & \\ 9.0 & 4.0 & 0.0 & \\ 8.0 & 5.0 & 3.0 & 0.0 \end{pmatrix} \end{array} \Rightarrow \begin{array}{c|c} & \begin{array}{ccc} (12) & 3 & (4\ 5) \end{array} \\ \begin{array}{c} (12) \\ 3 \\ (4\ 5) \end{array} & \begin{pmatrix} 0.0 & & \\ 5.0 & 0.0 & \\ 8.0 & 4.0 & 0.0 \end{pmatrix} \end{array}$$



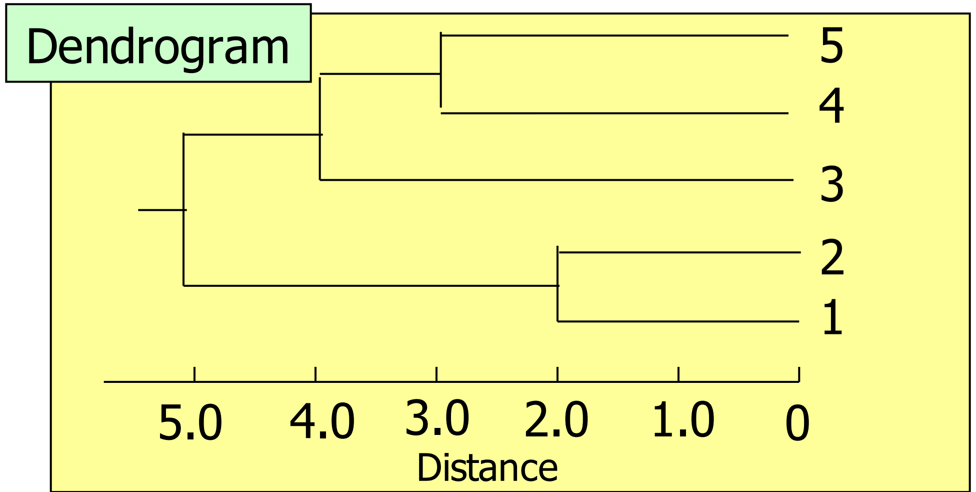
Example with Single Linkage

$$\begin{array}{cc} & \begin{array}{ccc} (12) & 3 & (4\ 5) \end{array} \\ \begin{array}{c} (12) \\ 3 \\ (4\ 5) \end{array} & \begin{pmatrix} 0.0 & & \\ 5.0 & 0.0 & \\ 8.0 & 4.0 & 0.0 \end{pmatrix} \end{array} \Rightarrow \begin{array}{cc} & \begin{array}{cc} (12) & (3\ 4\ 5) \end{array} \\ \begin{array}{c} (12) \\ (3\ 4\ 5) \end{array} & \begin{pmatrix} 0.0 & \\ 5.0 & 0.0 \end{pmatrix} \end{array}$$



Example with Single Linkage

$$\begin{array}{ccc} & (12) & (3\ 4\ 5) \\ (12) & \multicolumn{2}{c}{\begin{pmatrix} 0.0 & \\ 5.0 & 0.0 \end{pmatrix}} \\ (3\ 4\ 5) & & \end{array}$$

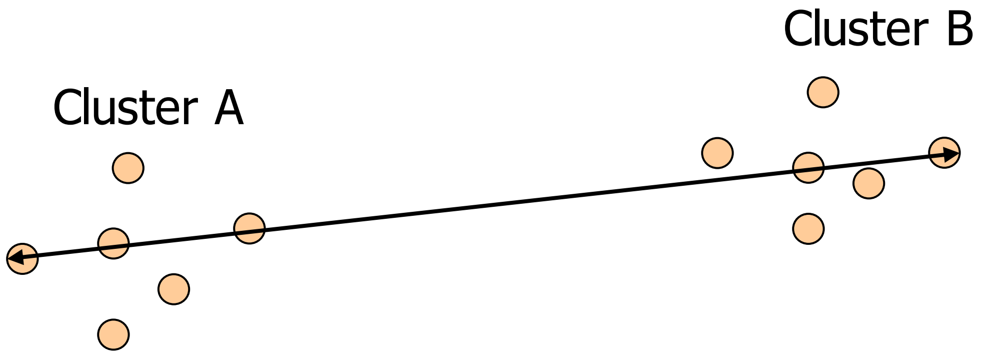


Distance of Clusters

- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

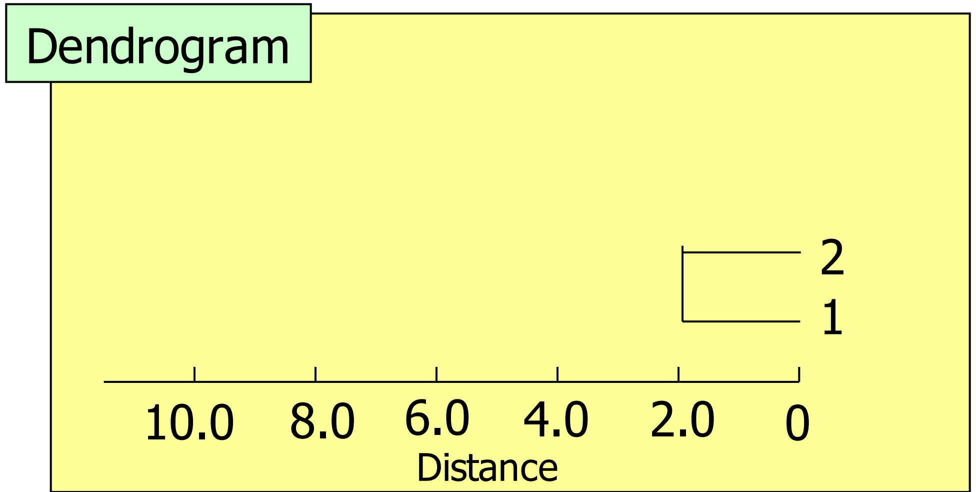
Complete Linkage

- The distance between two clusters is given by the distance between their most distant members



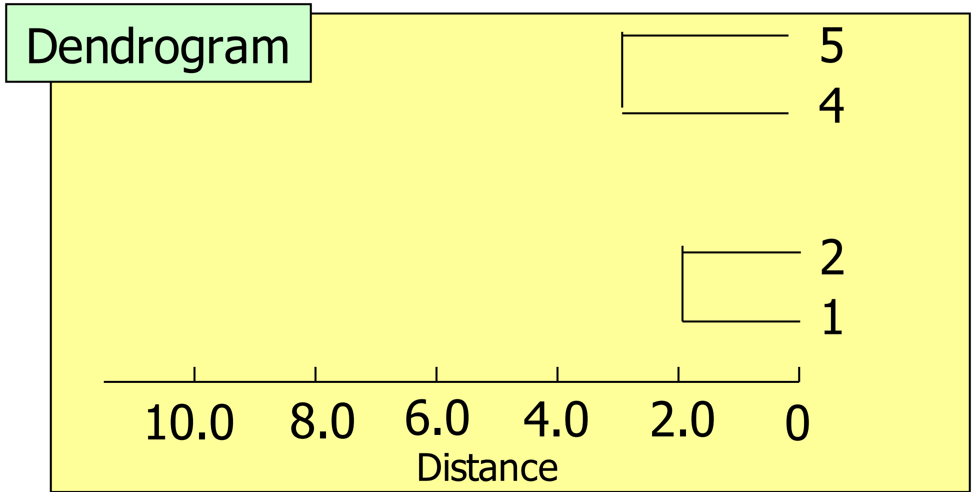
Example with Complete Linkage

$$
\begin{array}{c|c}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} &
\begin{pmatrix}
0.0 & & & & \\
2.0 & 0.0 & & & \\
6.0 & 5.0 & 0.0 & & \\
10.0 & 9.0 & 4.0 & 0.0 & \\
9.0 & 8.0 & 5.0 & 3.0 & 0.0
\end{pmatrix}
\end{array}
\Rightarrow
\begin{array}{c|c}
 & \begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\
\begin{array}{c} (12) \\ 3 \\ 4 \\ 5 \end{array} &
\begin{pmatrix}
0.0 & & & \\
6.0 & 0.0 & & \\
10.0 & 4.0 & 0.0 & \\
9.0 & 5.0 & 3.0 & 0.0
\end{pmatrix}
\end{array}
$$



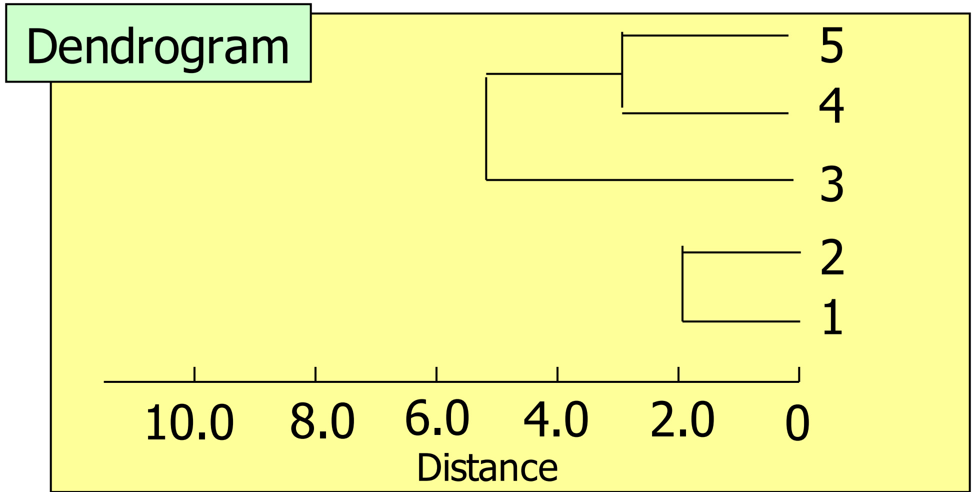
Example with Complete Linkage

$$
\begin{array}{c c}
 & \begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\
\begin{array}{c} (12) \\ 3 \\ 4 \\ 5 \end{array} &
\begin{pmatrix}
0.0 & & & \\
6.0 & 0.0 & & \\
10.0 & 4.0 & 0.0 & \\
9.0 & 5.0 & 3.0 & 0.0
\end{pmatrix}
\end{array}
\Rightarrow
\begin{array}{c c}
 & \begin{array}{ccc} (12) & 3 & (4\ 5) \end{array} \\
\begin{array}{c} (12) \\ 3 \\ (4\ 5) \end{array} &
\begin{pmatrix}
0.0 & & \\
6.0 & 0.0 & \\
10.0 & 5.0 & 0.0
\end{pmatrix}
\end{array}
$$



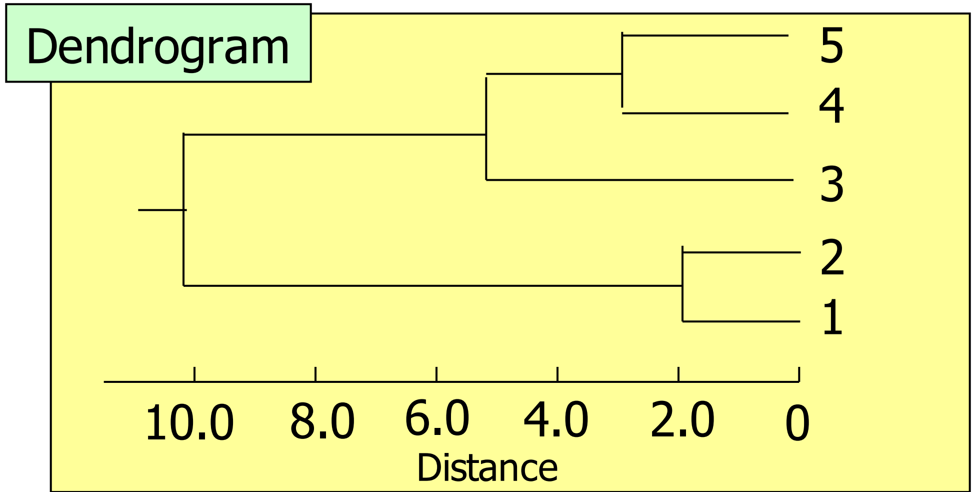
Example with Complete Linkage

$$\begin{array}{cc} & \begin{array}{ccc}(12) & 3 & (4\ 5)\end{array} \\ \begin{array}{c}(12)\\ 3\\ (4\ 5)\end{array} & \begin{pmatrix}0.0 & & \\ 6.0 & 0.0 & \\ 10.0 & 5.0 & 0.0\end{pmatrix}\end{array} \Rightarrow \begin{array}{cc} & \begin{array}{cc}(12) & (3\ 4\ 5)\end{array} \\ \begin{array}{c}(12)\\ (3\ 4\ 5)\end{array} & \begin{pmatrix}0.0 & \\ 10.0 & 0.0\end{pmatrix}\end{array}$$



Example with Complete Linkage

$$\begin{array}{cc} & \begin{array}{cc}(12) & (3\ 4\ 5)\end{array} \\ \begin{array}{c}(12) \\ (3\ 4\ 5)\end{array} & \begin{pmatrix} 0.0 & \\ 10.0 & 0.0 \end{pmatrix} \end{array}$$

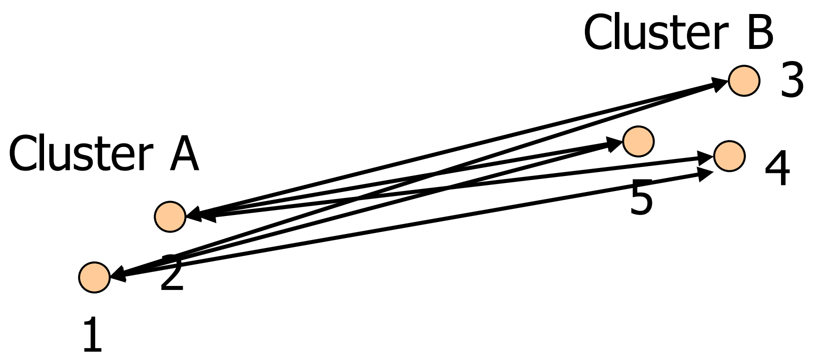


Distance of Clusters

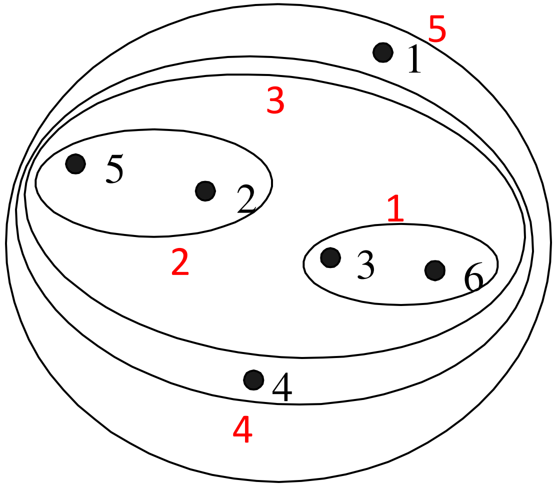
- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

Group Average Clustering

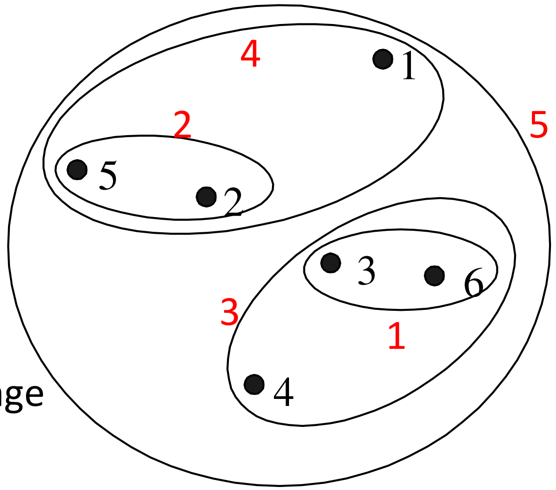
- The distance between two clusters is defined as the average of the distances between all pairs of records (one from each cluster).
- $d_{AB} = 1/6\ (d_{13} + d_{14} + d_{15} + d_{23} + d_{24} + d_{25})$



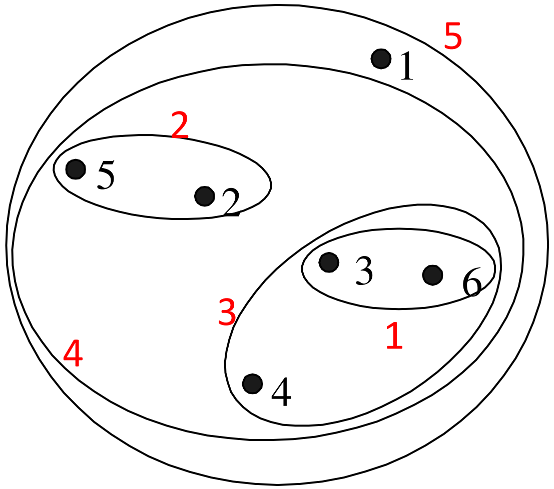
Distance of Clusters: Comparison



Single Linkage



Complete Linkage



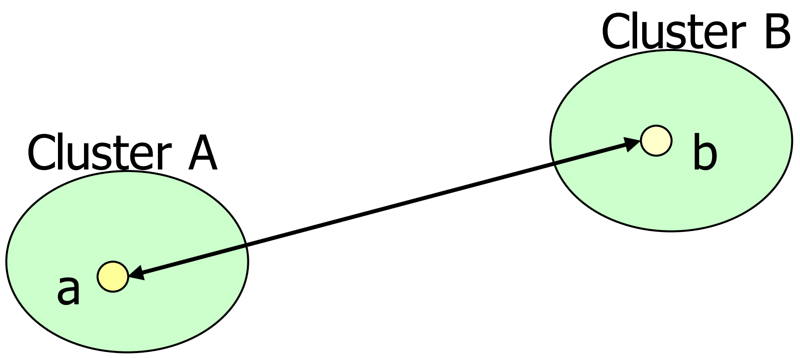
Group Average

Distance of Clusters

- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

Centroid Linkage

- The distance between two clusters is defined as the distance between the mean vectors of the two clusters.
- $d_{AB} = d_{ab}$
- where a is the mean vector of the cluster A and b is the mean vector of the cluster B.

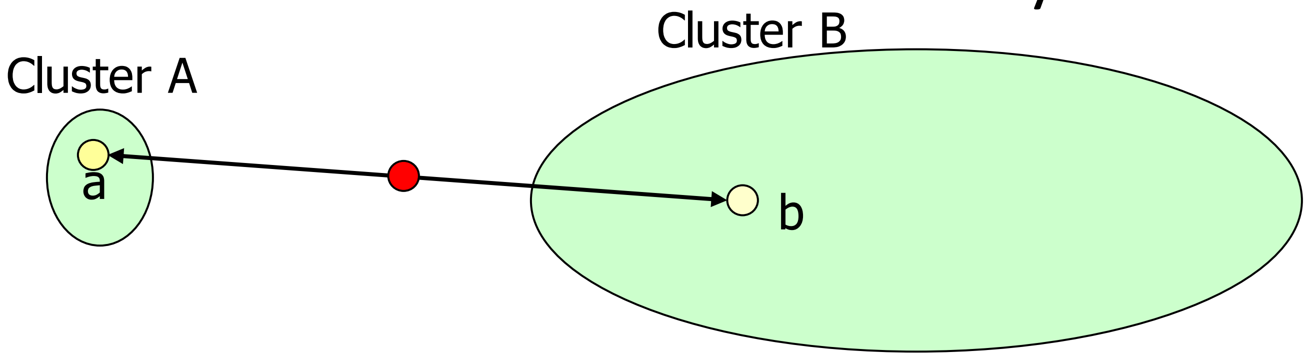


Distance of Clusters

- Single Linkage
- Complete Linkage
- Group Average Linkage
- Centroid Linkage
- Median Linkage

Median Clustering

- Disadvantage of the Centroid Clustering: When a large cluster is merged with a small one, the centroid of the combined cluster would be closer to the large one, ie. the characteristic properties of the small one are lost
- After we have combined two groups, the mid-point of the original two cluster centers is used as the center of the newly combined group



Clustering Methods

- Hierarchical Clustering Methods
  - Agglomerative methods
  - Divisive methods – polythetic approach

Divisive Methods

- In a divisive algorithm, we start with the assumption that all the data is part of one cluster.
- We then use a distance criterion to divide the cluster in two, and then subdivide the clusters until a stopping criterion is achieved.
  - Polythetic – divide the data based on the values by all attributes

Polythetic Approach (based on group average linkage)

$$\begin{array}{c|c}
 & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} & \begin{pmatrix} 0 & & & & & & \\ 10 & 0 & & & & & \\ 7 & 7 & 0 & & & & \\ 30 & 23 & 21 & 0 & & & \\ 29 & 25 & 22 & 7 & 0 & & \\ 38 & 34 & 31 & 10 & 11 & 0 & \\ 42 & 36 & 36 & 13 & 17 & 9 & 0 \end{pmatrix}
\end{array}$$

A = {1 }

B = {2, 3, 4, 5, 6, 7}

D(1, *) = 26.0

D(2, *) = 22.5

D(3, *) = 20.7

D(4, *) = 17.3

D(5, *) = 18.5

D(6, *) = 22.2

D(7, *) = 25.5

Polythetic Approach (based on group average linkage)

$$
\begin{array}{c|ccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
1 & 0 & & & & & & \\
2 & 10 & 0 & & & & & \\
3 & 7 & 7 & 0 & & & & \\
4 & 30 & 23 & 21 & 0 & & & \\
5 & 29 & 25 & 22 & 7 & 0 & & \\
6 & 38 & 34 & 31 & 10 & 11 & 0 & \\
7 & 42 & 36 & 36 & 13 & 17 & 9 & 0
\end{array}
$$

D(2, A) = 10

D(3, A) = 7

D(4, A) = 30

D(5, A) = 29

D(6, A) = 38

D(7, A) = 42

A = {1 }

B = {2, 3, 4, 5, 6, 7}

Polythetic Approach (based on group average linkage)

$$
\begin{array}{c|ccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
1 & 0 & & & & & & \\
2 & 10 & 0 & & & & & \\
3 & 7 & 7 & 0 & & & & \\
4 & 30 & 23 & 21 & 0 & & & \\
5 & 29 & 25 & 22 & 7 & 0 & & \\
6 & 38 & 34 & 31 & 10 & 11 & 0 & \\
7 & 42 & 36 & 36 & 13 & 17 & 9 & 0
\end{array}
$$

A = {1 }

B = {2, 3, 4, 5, 6, 7}

D(2, A) = 10

D(2, B) = 25.0

D(3, A) = 7

D(3, B) = 23.4

D(4, A) = 30

D(4, B) = 14.8

D(5, A) = 29

D(5, B) = 16.4

D(6, A) = 38

D(6, B) = 19.0

D(7, A) = 42

D(7, B) = 22.2

Polythetic Approach (based on group average linkage)

$$
\begin{array}{c|ccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
1 & 0 & & & & & & \\
2 & 10 & 0 & & & & & \\
3 & 7 & 7 & 0 & & & & \\
4 & 30 & 23 & 21 & 0 & & & \\
5 & 29 & 25 & 22 & 7 & 0 & & \\
6 & 38 & 34 & 31 & 10 & 11 & 0 & \\
7 & 42 & 36 & 36 & 13 & 17 & 9 & 0
\end{array}
$$

A = {1, 3 }

B = {2, ~~3~~, 4, 5, 6, 7}

D(2, A) = 10	D(2, B) = 25.0	Δ_2 = 15.0
D(3, A) = 7	D(3, B) = 23.4	Δ_3 = 16.4
D(4, A) = 30	D(4, B) = 14.8	Δ_4 = -15.2
D(5, A) = 29	D(5, B) = 16.4	Δ_5 = -12.6
D(6, A) = 38	D(6, B) = 19.0	Δ_6 = -19.0
D(7, A) = 42	D(7, B) = 22.2	Δ_7 = -19.8

Polythetic Approach (based on group average linkage)

$$
\begin{array}{c}
\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\begin{pmatrix}
0 & & & & & & \\
10 & 0 & & & & & \\
7 & 7 & 0 & & & & \\
30 & 23 & 21 & 0 & & & \\
29 & 25 & 22 & 7 & 0 & & \\
38 & 34 & 31 & 10 & 11 & 0 & \\
42 & 36 & 36 & 13 & 17 & 9 & 0
\end{pmatrix}
\end{array}
$$

A = {1, 3 }

B = {2, 4, 5, 6, 7}

D(2, A) = 10	D(2, B) = 25.0	Δ_2 = 15.0
D(3, A) = 7	D(3, B) = 23.4	Δ_3 = 16.4
D(4, A) = 30	D(4, B) = 14.8	Δ_4 = -15.2
D(5, A) = 29	D(5, B) = 16.4	Δ_5 = -12.6
D(6, A) = 38	D(6, B) = 19.0	Δ_6 = -19.0
D(7, A) = 42	D(7, B) = 22.2	Δ_7 = -19.8

Polythetic Approach (based on group average linkage)

$$
\begin{array}{c|ccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
1 & 0 & & & & & & \\
2 & 10 & 0 & & & & & \\
3 & 7 & 7 & 0 & & & & \\
4 & 30 & 23 & 21 & 0 & & & \\
5 & 29 & 25 & 22 & 7 & 0 & & \\
6 & 38 & 34 & 31 & 10 & 11 & 0 & \\
7 & 42 & 36 & 36 & 13 & 17 & 9 & 0
\end{array}
$$

D(2, A) = 8.5

D(4, A) = 25.5

D(5, A) = 25.5

D(6, A) = 34.5

D(7, A) = 39.0

A = {1, 3 }

B = {2, 4, 5, 6, 7}

Polythetic Approach (based on group average linkage)

$$
\begin{array}{c|ccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
1 & 0 & & & & & & \\
2 & 10 & 0 & & & & & \\
3 & 7 & 7 & 0 & & & & \\
4 & 30 & 23 & 21 & 0 & & & \\
5 & 29 & 25 & 22 & 7 & 0 & & \\
6 & 38 & 34 & 31 & 10 & 11 & 0 & \\
7 & 42 & 36 & 36 & 13 & 17 & 9 & 0
\end{array}
$$

D(2, A) = 8.5

D(2, B) = 29.5

D(4, A) = 25.5

D(4, B) = 13.2

D(5, A) = 25.5

D(5, B) = 15.0

D(6, A) = 34.5

D(6, B) = 16.0

D(7, A) = 39.0

D(7, B) = 18.75

A = {1, 3 }

B = {2, 4, 5, 6, 7}

Polythetic Approach (based on group average linkage)

$$
\begin{array}{c|c}
 & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} &
\begin{pmatrix}
0 & & & & & & \\
10 & 0 & & & & & \\
7 & 7 & 0 & & & & \\
30 & 23 & 21 & 0 & & & \\
29 & 25 & 22 & 7 & 0 & & \\
38 & 34 & 31 & 10 & 11 & 0 & \\
42 & 36 & 36 & 13 & 17 & 9 & 0
\end{pmatrix}
\end{array}
$$

D(2, A) = 8.5	D(2, B) = 29.5	Δ_2 = 21.0
D(4, A) = 25.5	D(4, B) = 13.2	Δ_4 = -12.3
D(5, A) = 25.5	D(5, B) = 15.0	Δ_5 = -10.5
D(6, A) = 34.5	D(6, B) = 16.0	Δ_6 = -18.5
D(7, A) = 39.0	D(7, B) = 18.75	Δ_7 = -20.25

A = {1, 3 , 2 }

B = {~~2~~, 4, 5, 6, 7}

Polythetic Approach (based on group average linkage)

$$
\begin{array}{c|ccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
1 & 0 & & & & & & \\
2 & 10 & 0 & & & & & \\
3 & 7 & 7 & 0 & & & & \\
4 & 30 & 23 & 21 & 0 & & & \\
5 & 29 & 25 & 22 & 7 & 0 & & \\
6 & 38 & 34 & 31 & 10 & 11 & 0 & \\
7 & 42 & 36 & 36 & 13 & 17 & 9 & 0
\end{array}
$$

D(2, A) = 8.5	D(2, B) = 29.5	Δ_2 = 21.0
D(4, A) = 25.5	D(4, B) = 13.2	Δ_4 = -12.3
D(5, A) = 25.5	D(5, B) = 15.0	Δ_5 = -10.5
D(6, A) = 34.5	D(6, B) = 16.0	Δ_6 = -18.5
D(7, A) = 39.0	D(7, B) = 18.75	Δ_7 = -20.25

A = {1, 3 , 2 }

B = {   4, 5, 6, 7}

Polythetic Approach (based on group average linkage)

$$
\begin{array}{c}
 & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} &
\begin{pmatrix}
0 & & & & & & \\
10 & 0 & & & & & \\
7 & 7 & 0 & & & & \\
30 & 23 & 21 & 0 & & & \\
29 & 25 & 22 & 7 & 0 & & \\
38 & 34 & 31 & 10 & 11 & 0 & \\
42 & 36 & 36 & 13 & 17 & 9 & 0
\end{pmatrix}
\end{array}
$$

D(4, A) = 24.7

D(5, A) = 25.3

D(6, A) = 34.3

D(7, A) = 38.0

A = {1, 3, 2 }

B = {4, 5, 6, 7}

Polythetic Approach (based on group average linkage)

$$\begin{array}{c|ccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline 1 & 0 & & & & & & \\ 2 & 10 & 0 & & & & & \\ 3 & 7 & 7 & 0 & & & & \\ 4 & 30 & 23 & 21 & 0 & & & \\ 5 & 29 & 25 & 22 & 7 & 0 & & \\ 6 & 38 & 34 & 31 & 10 & 11 & 0 & \\ 7 & 42 & 36 & 36 & 13 & 17 & 9 & 0 \end{array}$$

D(4, A) = 24.7

D(4, B) = 10.0

D(5, A) = 25.3

D(5, B) = 11.7

D(6, A) = 34.3

D(6, B) = 10.0

D(7, A) = 38.0

D(7, B) = 13.0

A = {1, 3 , 2 }

B = {4, 5, 6, 7}

Polythetic Approach (based on group average linkage)

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}\begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} 0 & & & & & & \\ 10 & 0 & & & & & \\ 7 & 7 & 0 & & & & \\ 30 & 23 & 21 & 0 & & & \\ 29 & 25 & 22 & 7 & 0 & & \\ 38 & 34 & 31 & 10 & 11 & 0 & \\ 42 & 36 & 36 & 13 & 17 & 9 & 0 \end{pmatrix} \end{array}$$

D(4, A) = 24.7 | D(4, B) = 10.0 | Δ_4 = -14.7

D(5, A) = 25.3 | D(5, B) = 11.7 | Δ_5 = -13.6

D(6, A) = 34.3 | D(6, B) = 10.0 | Δ_6 = -24.3

D(7, A) = 38.0 | D(7, B) = 13.0 | Δ_7 = -25.0

A = {1, 3 , 2 }

B = {4, 5, 6, 7}

All differences are negative. The process would continue on each subgroup separately.

Hierarchical Clustering: Time and Space requirements

- $O(N^2)$ space since it uses the proximity matrix.
  - N is the number of points.

- $O(N^3)$ time in many cases
  - There are N steps and at each step the size, N^2, proximity matrix must be updated and searched
  - Complexity can be reduced to $O(N^2 \log(N))$ time for some approaches

Hierarchical Clustering: Problems and Limitations

- Computational complexity in time and space

- Once a decision is made to combine two clusters, it cannot be undone

- No objective function is directly minimized

- Different schemes have problems with one or more of the following:
  - Sensitivity to noise and outliers
  - Difficulty handling different sized clusters and convex shapes
  - Breaking large clusters