Hierarchical clustering

- A hierarchical clustering is a set of nested clusters that are organized as a tree.
- There are two basic approaches for generating a hierarchical clustering
  - Agglomerative
  - Divisive

Hierarchical clustering

- In agglomerative hierarchical clustering, we start with the points as individual clusters.
- At each step, we merge the closest pair of clusters.
- This requires defining a notion of cluster distance.

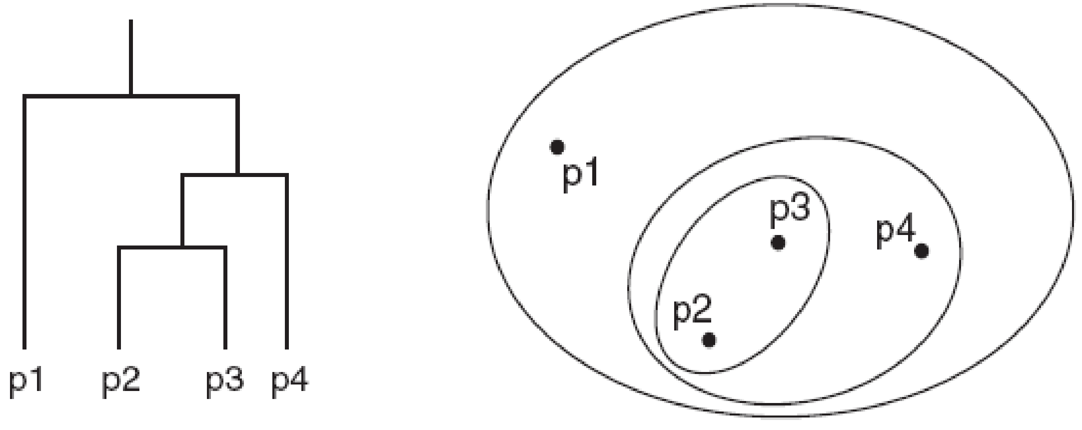
Hierarchical clustering

- In divisive hierarchical clustering, we start with one, all-inclusive cluster.
- At each step, we split a cluster.
- This process continues until only singleton clusters of individual points remain.
- In this case, we need to decide
  - Which cluster to split at each step and
  - How to do the splitting.

Hierarchical clustering

- A hierarchical clustering is often displayed graphically using a tree-like diagram called the dendrogram.
- The dendrogram displays both
  - the cluster-subcluster relationships and
  - the order in which the clusters are merged (agglomerative) or split (divisive).
- For sets of 2-D points, a hierarchical clustering can also be graphically represented using a nested cluster diagram.

Hierarchical clustering



(a) Dendrogram. (b) Nested cluster diagram.

Hierarchical clustering

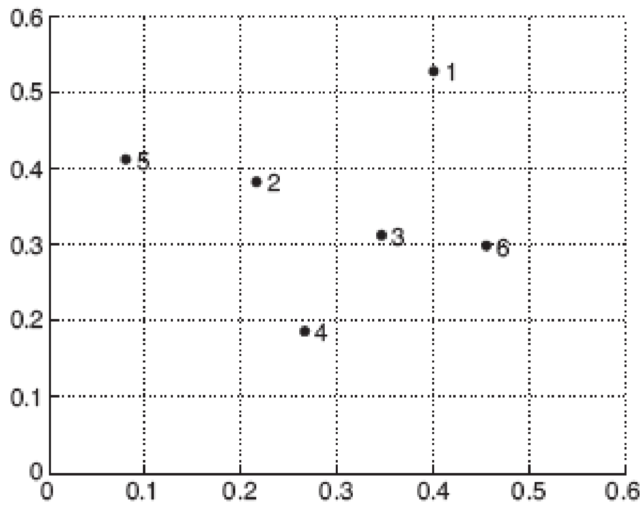
- The basic agglomerative hierarchical clustering algorithm is summarized as follows
  - Compute the distance matrix.
  - Repeat
    - Merge the closest two clusters
    - Update the distance matrix to reflect the distance between the new cluster and the original clusters.
  - Until only one cluster remains

Hierarchical clustering

- Different definitions of cluster distance leads to different versions of hierarchical clustering.
- These versions include
  - Single link or MIN
  - Complete link or MAX
  - Group average

Hierarchical clustering

- We consider the following set of data points.
- The Euclidean distance matrix for these data points is shown in the following slide.



Point	x Coordinate	y Coordinate
p1	0.40	0.53
p2	0.22	0.38
p3	0.35	0.32
p4	0.26	0.19
p5	0.08	0.41
p6	0.45	0.30

Hierarchical clustering

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

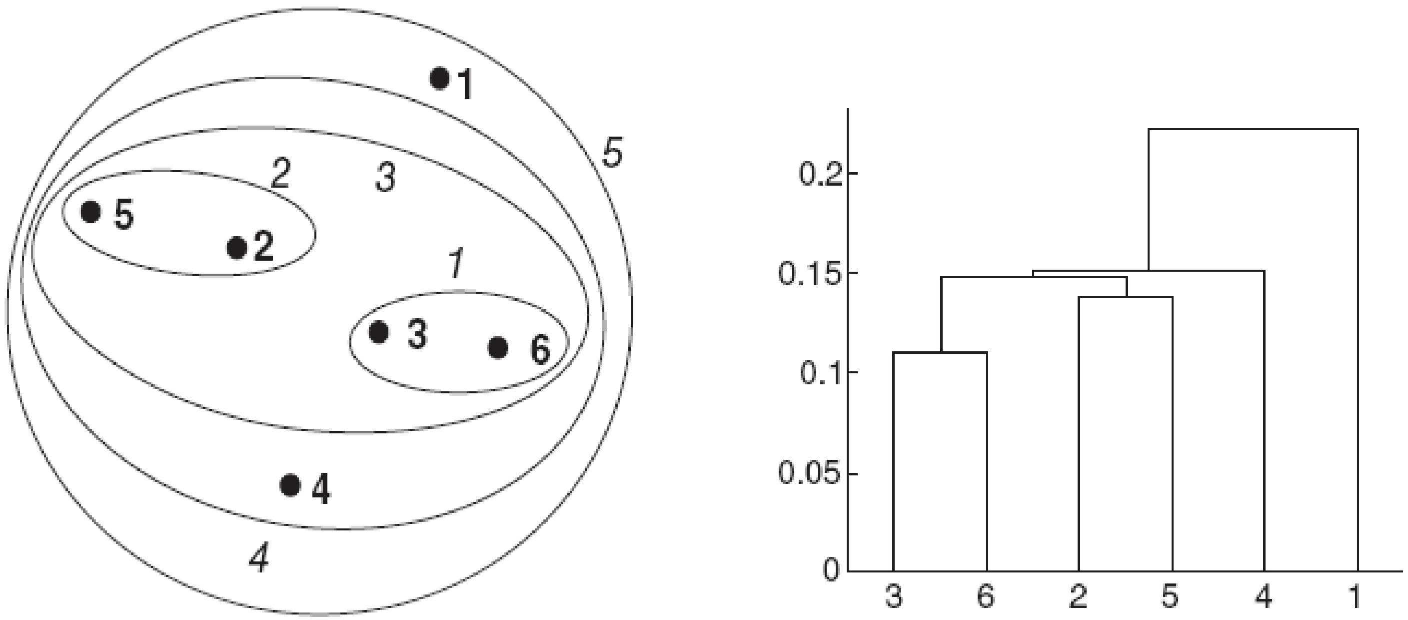
Single link

- We now consider the single link or MIN version of hierarchical clustering.
- In this case, the distance between two clusters is defined as the minimum of the distance between any two points, with one of the points in the first cluster, and the other point in the second cluster.
- This technique is good at handling clusters with non-elliptical shapes.

Single link

- The following figure shows the result of applying the single link technique to our example data.
- The left figure shows the nested clusters as a sequence of nested ellipses.
- The numbers associated with the ellipses indicate the order of the clustering.
- The right figure shows the same information in the form of a dendrogram.
- The height at which two clusters are merged in the dendrogram reflects the distance of the two clusters.

Single link



(a) Single link clustering.

(b) Single link dendrogram.

Single link

- For example, we see that the distance between points 3 and 6 is 0.11.
- That is the height at which they are joined into one cluster in the dendrogram.
- As another example, the distance between clusters {3,6} and {2,5} is

$$
\begin{aligned}
d(\{3,6\},\{2,5\}) &= \min(d(3,2), d(6,2), d(3,5), d(6,5)) \\
&= \min(0.15, 0.25, 0.28, 0.39) \\
&= 0.15
\end{aligned}
$$

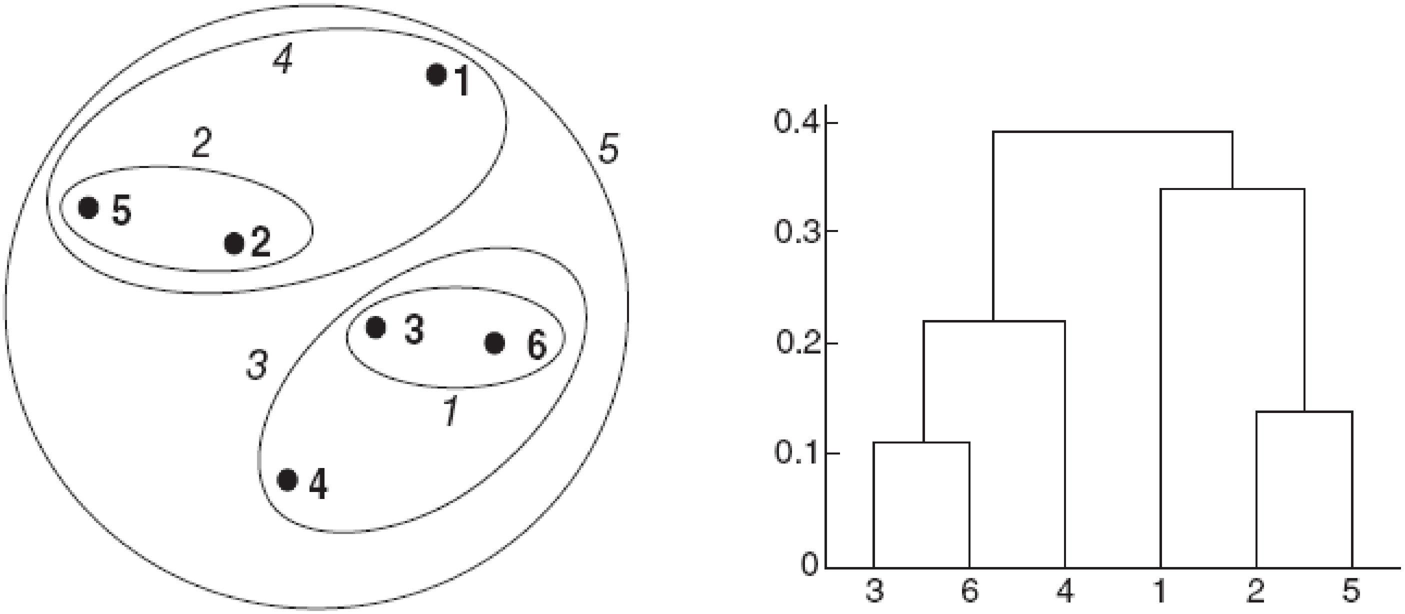
Complete link

- We now consider the complete link or MAX version of hierarchical clustering.
- In this case, the distance between two clusters is defined as the maximum of the distance between any two points, with one of the points in the first cluster, and the other point in the second cluster.
- Complete link tends to produce clusters with globular shapes.

Complete link

- The following figure shows the results of applying the complete link approach to our sample data points.
- As with single link, points 3 and 6 are merged first.
- Points 2 and 5 are then merged.
- After that, {3,6} is merged with {4}.

Complete link



(a) Complete link clustering.

(b) Complete link dendrogram.

Complete link

- This can be explained by the following calculations

$$
\begin{aligned}
d(\{3,6\},\{4\}) &= \max(d(3,4), d(6,4)) \\
&= \max(0.15, 0.22) \\
&= 0.22
\end{aligned}
$$

$$
\begin{aligned}
d(\{3,6\},\{1\}) &= \max(d(3,1), d(6,1)) \\
&= \max(0.22, 0.23) \\
&= 0.23
\end{aligned}
$$

$$
\begin{aligned}
d(\{3,6\},\{2,5\}) &= \max(d(3,2), d(6,2), d(3,5), d(6,5)) \\
&= \max(0.15, 0.25, 0.28, 0.39) \\
&= 0.39
\end{aligned}
$$

Complete link

$$d(\{4\},\{1\}) = 0.37$$

$$\begin{aligned} d(\{4\},\{2,5\}) &= \max(d(4,2), d(4,5)) \\ &= \max(0.20, 0.29) \\ &= 0.29 \end{aligned}$$

$$\begin{aligned} d(\{1\},\{2,5\}) &= \max(d(1,2), d(1,5)) \\ &= \max(0.24, 0.34) \\ &= 0.34 \end{aligned}$$

Group average

- We now consider the group average version of hierarchical clustering.
- In this case, the distance between two clusters is defined as the average distance across all pairs of points, with one point in each pair belonging to the first cluster, and the other point of the pair belonging to the second cluster.
- This is an intermediate approach between the single and complete link approaches.

Group average

- We consider two clusters C_i and C_j, which are of sizes m_i and m_j respectively.
- The distance between the two clusters can be expressed by the following equation

$$d(C_i, C_j) = \frac{\sum_{\mathbf{x} \in C_i, \mathbf{y} \in C_j} d(\mathbf{x}, \mathbf{y})}{m_i m_j}$$

Group average

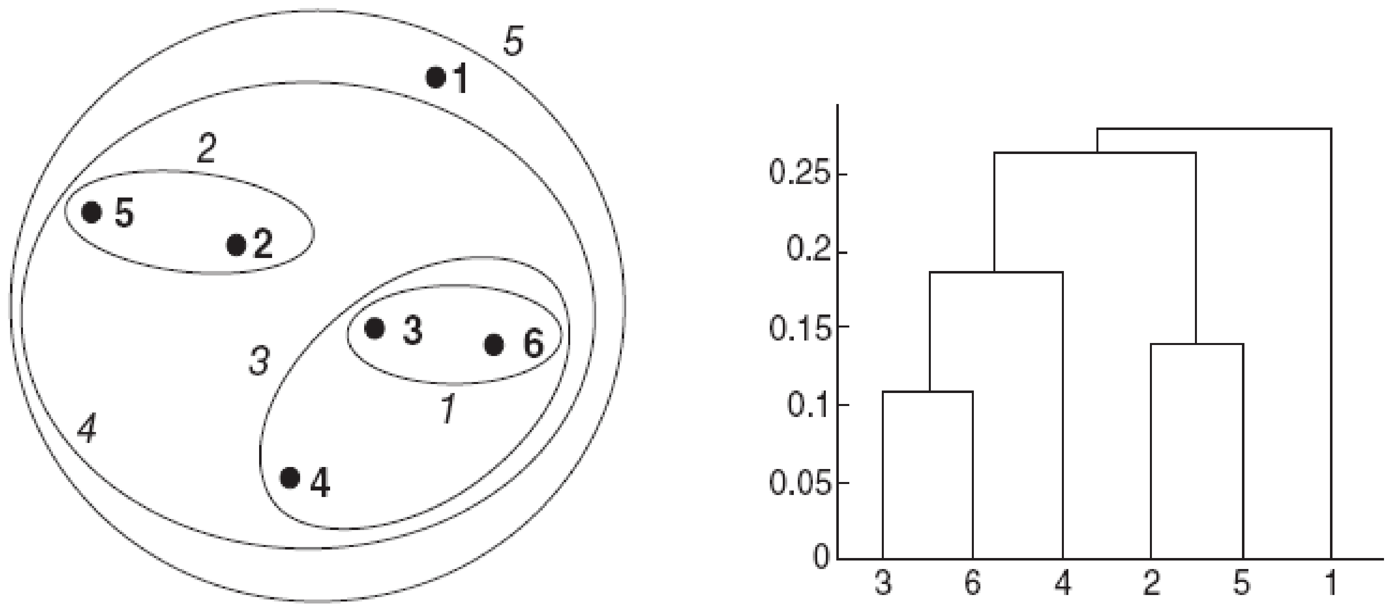
- The following figure shows the results of applying the group average to our sample data.
- The distances between some of the clusters are calculated as follows:

$$d(\{3,6,4\},\{1\}) = \frac{0.22 + 0.37 + 0.23}{3 \times 1} = 0.27$$

$$d(\{2,5\},\{1\}) = \frac{0.24 + 0.34}{2 \times 1} = 0.29$$

$$d(\{3,6,4\},\{2,5\}) = \frac{0.15 + 0.28 + 0.25 + 0.39 + 0.20 + 0.29}{3 \times 2} = 0.26$$

Group average



(a) Group average clustering. (b) Group average dendrogram.

Group average

- We observe that d({3,6,4},{2,5}) is smaller than d({3,6,4},{1}) and d({2,5},{1}).
- As a result, {3,6,4} and {2,5} are merged at the fourth stage.

Key issues

- Hierarchical clustering is effective when the underlying application requires the creation of a multi-level structure.
- However, they are expensive in terms of their computational and storage requirements.
- In addition, once a decision is made to merge two clusters, it cannot be undone at a later time.