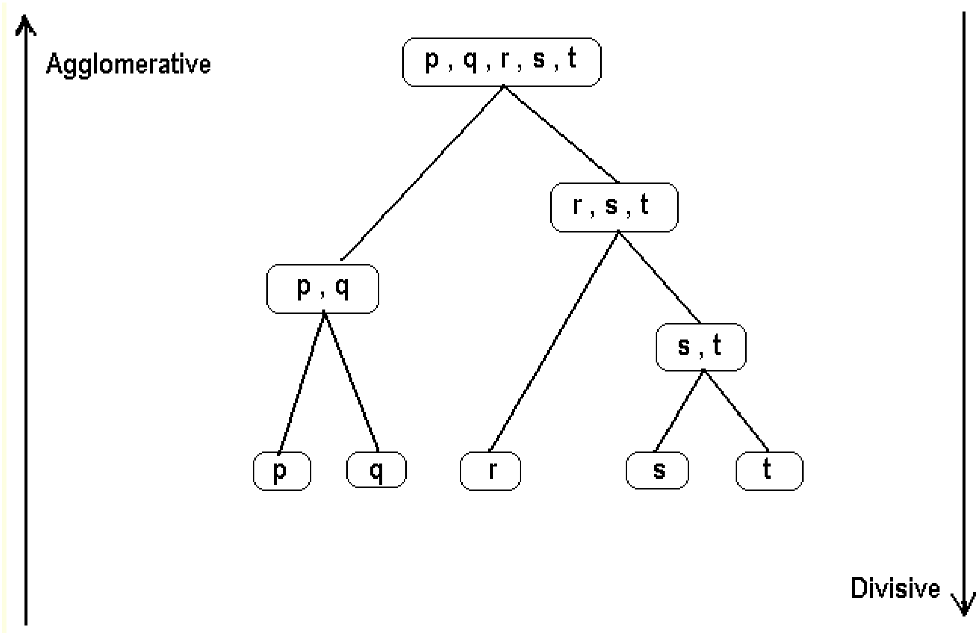
Hierarchical Structure

- Method of hierarchical clustering
  - Agglomerative (bottom-up)
  - Divisive (top-down)



Distance Measures (between objects)

- Euclidean (distance)
  - p = (p_1, p_2,..., p_n) and q = (q_1, q_2,..., q_n)

Distance

$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+\cdots+(q_n-p_n)^2} = \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

Distance Measures (between objects) [cont.]

- Jaccard's coefficients (for binary data)
  - Each object has *n* binary attributes.

Similarity

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}.$$

M_{11}: # of attributes where A and B both have a value of 1

M_{01}: # of attributes where the value of A is 0 and the value of B is 1

M_{10}: # of attributes where the value of A is 1 and the value of B is 0

  - For example, A = (1, 1, 0, 0) and B = (1, 1, 1, 0)
    - $M_{01} = 1$, $M_{10} = 0$ and $M_{11} = 2$
    - J(A, B) = 2/(1 + 0 + 2) = 2/3

Distance Measures (between objects) [cont.]

- Matching coefficients (for binary data)
  - Each object has *n* binary attributes.

Similarity

$$M(A,B) = \frac{M_{00} + M_{11}}{n}$$

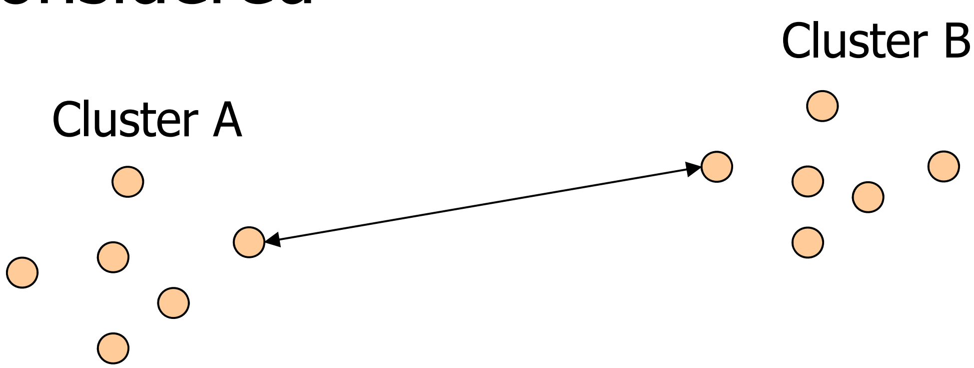
M_{00}: # of attributes where A and B both have a value of 0

M_{11}: # of attributes where A and B both have a value of 1

  - For example, A = (1, 1, 0, 0), B = (1, 1, 1, 0)
    - M(A, B) = (1 + 2) /4 = 3/4

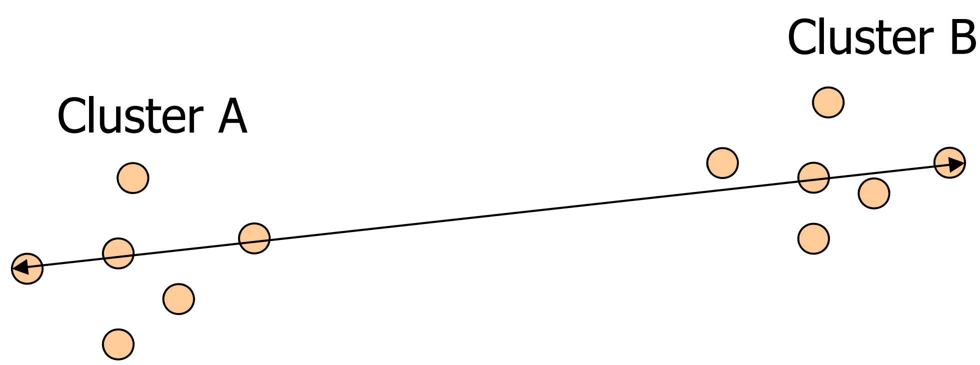
Single Linkage

- 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



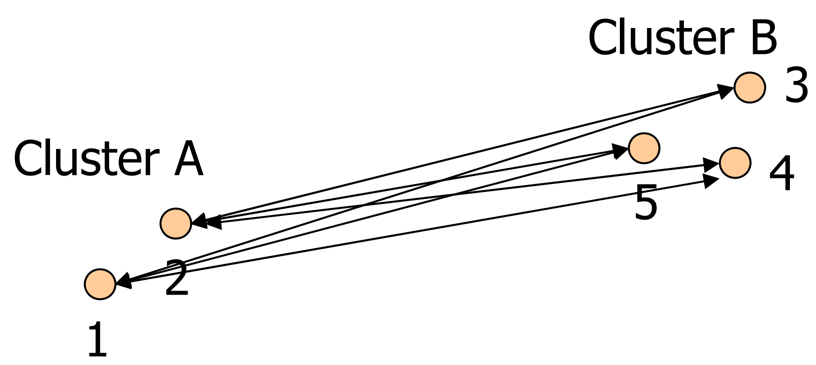
Complete Linkage

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



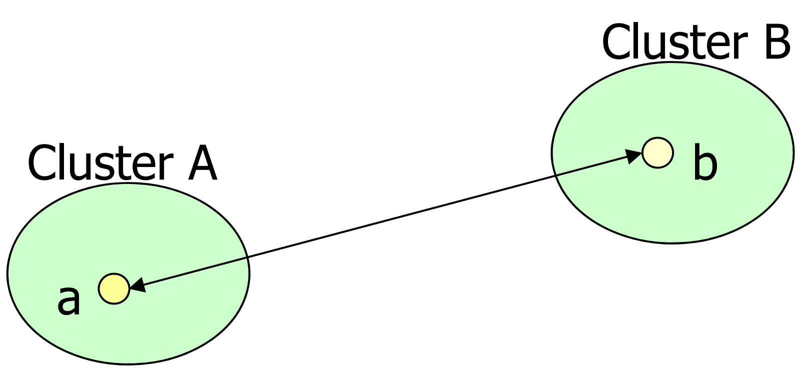
Group Average Clustering

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



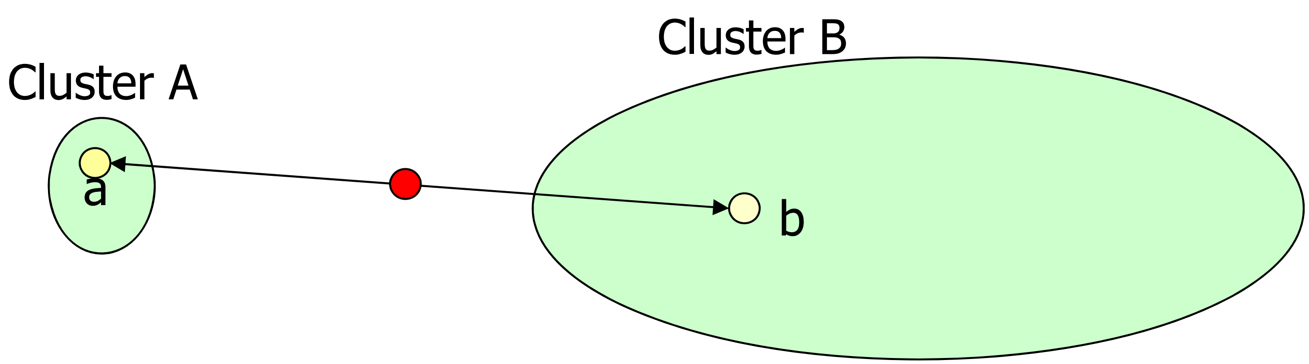
Centroid Clustering

- 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.



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 closed 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 centres is used as the centre of the newly combined group



McQuitty's Method

- $Dist(C_i, C_j)$ – distance between cluster C_i and cluster C_j

- Suppose we have three clusters C_i , C_j and C_k
- Then, C_i and C_j are merged to form a larger cluster C_p i.e., $C_p = C_i \cup C_j$
- $Dist(C_p, C_k) = (Dist(C_i, C_k) + Dist(C_j, C_k)) / 2$

Ward's Method

- Distance between 2 clusters is defined to be the information loss of the final cluster merged from 2 clusters.
- Information loss: Error sum-of-squares (ESS).

E.g., 10 objects: {6, 5, 6, 2, 2, 2, 2, 0, 0, 0}.

Treating the objects as one group: Mean of the objects = 2.5
$ESS_{\text{one group}} = (6 - 2.5)^2 + (5 - 2.5)^2 + ... + (0 - 2.5)^2 = 50.5$

Treating the objects as four groups: {0,0,0}, {2,2,2,2}, {5}, {6,6}
$ESS_{\text{four groups}} = ESS_{\text{group1}} + ESS_{\text{group2}} + ESS_{\text{group3}} + ESS_{\text{group4}} = 0$

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
  - Monothetic – divide the data on the basis of the possession of a single specified attribute

Polythetic Approach

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

$D(4, A) = 24.7$ | $D(4, B) = 10.0$ | $\Delta_4 = -14.7$

$D(5, A) = 25.3$ | $D(5, B) = 11.7$ | $\Delta_5 = -13.6$

$D(6, A) = 34.3$ | $D(6, B) = 10.0$ | $\Delta_6 = -24.3$

$D(7, A) = 38.0$ | $D(7, B) = 13.0$ | $\Delta_7 = -25.0$

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

$A = \{1, 3, 2\}$

$B = \{4, 5, 6, 7\}$

Monothetic

It is usually used when the data consists of **binary** variables.

	A	B	C
1	0	1	1
2	1	1	0
3	1	1	1
4	1	1	0
5	0	0	1

B \ A	1	0
1	a=3	b=1
0	c=0	d=1

Chi-Square Measure

$$\chi_{AB}{}^2 = \frac{(ad - bc)^2 N}{(a+b)(a+c)(b+d)(c+d)}$$

$$= \frac{(3-0)^2 \cdot 5}{4 \cdot 3 \cdot 2 \cdot 1}$$

= 1.875

B \ A	1	0
1	a=3	b=1
0	c=0	d=1

etic

Attr.	AB	AC	BC
a	3	1	2
b	1	2	1
c	0	2	2
d	1	0	0
N	5	5	5
χ^2	1.87	2.22	0.83

Chi-Square Measure

It is usually used when the data consists of **binary** variables.

	A	B	C
1	0	1	1
2	1	1	0
3	1	1	1
4	1	1	0
5	0	0	1

For attribute A,

$$\chi_{AB}^2 + \chi_{AC}^2 = 4.09$$

For attribute B,

$$\chi_{AB}^2 + \chi_{BC}^2 = 2.70$$

For attribute C,

$$\chi_{AC}^2 + \chi_{BC}^2 = 3.05$$

We choose attribute A for dividing the data into two groups. {2, 3, 4}, and {1, 5}