METHODS FOR MULTIVARIATE DATA ANALYSIS: TOOLS FOR DATA MINING

S. Lakshmivarahan
School of Computer Science
University of Oklahoma

BASIC TECHNIQUES

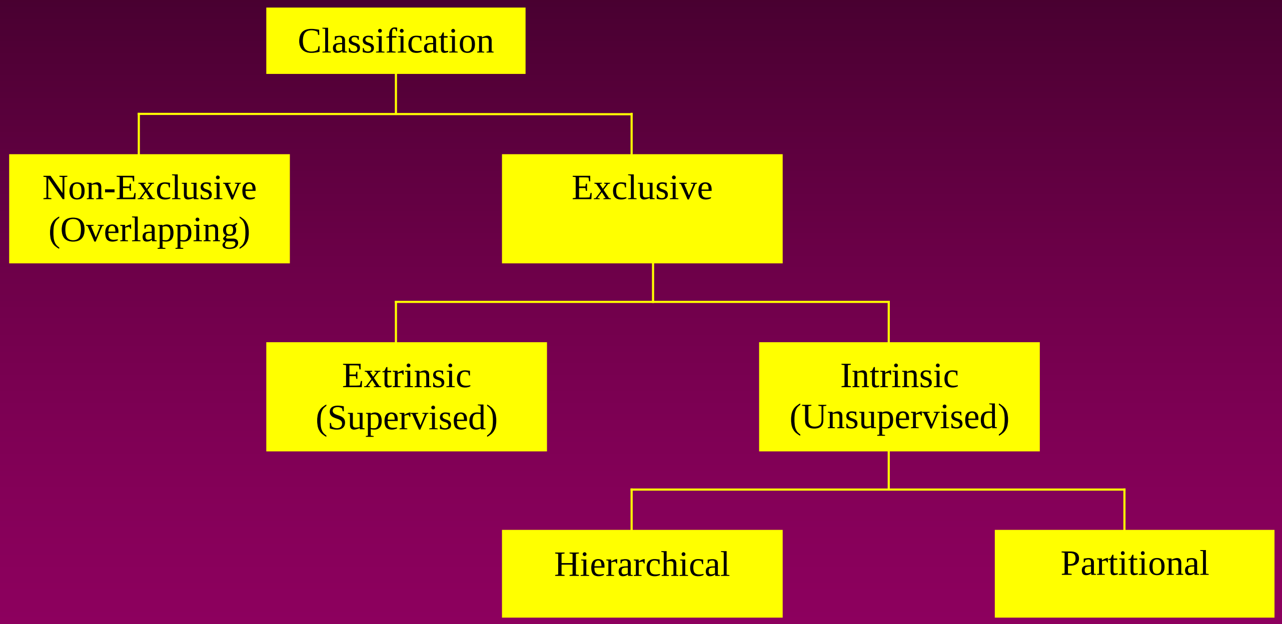
- Clustering
- Principle component analysis (PCA)
- Time series analysis
- Singular spectrum analysis
- Disriminant analysis
- Pattern recognition and learning

Clustering: An Overview

- Introduction
- Basic definitions
- Framework for Cluster Analysis
- Scales for attributes
- Standardizing the Data matrix
- Resemblance coefficients for Quantitative attributes
- Hierarchical Clustering
- Partitional Clustering

Cluster Analysis is a Classification Technique

- Multivariable data analysis



Anil K. Jain, Richard C. Dubes, 1988, "Algorithms For Clustering Data", Prentice Hall, NJ.

Approaches to carry out Cluster Analysis Algorithms

- Agglomerative *vs* Divisive
- Serial *vs* Simultaneous- objects
- Monothetic vs Polythetic - attributes
- Matrix Theory vs Graph Theory

Basic Definition

- Data Matrix

Objects

		1	2	……………………	n
	1	X_{11}	X_{12}	……………………	X_{1n}
	2	X_{21}	X_{22}	……………………	X_{2n}
X=	3	X_{31}	X_{32}	……………………	X_{3n}
	.	.	.		.
	.	.	.		.
	.	.	.		.
	m	X_{m1}	X_{m2}	……………………	X_{mn}

Basic Definition

- Data Matrix
- Objects
- Attributes

Note: Classify objects – Q analysis

Classify attributes – R analysis

Framework for Cluster Analysis

- Obtain the Data Matrix

Objects

		1	2	……………………	n
	1	X_{11}	X_{12}	……………………	X_{1n}
	2	X_{21}	X_{22}	……………………	X_{2n}
X=	3	X_{31}	X_{32}	……………………	X_{3n}
	.	.	.		.
	.	.	.		.
	.	.	.		.
	m	X_{m1}	X_{m2}	……………………	X_{mn}

X_{*j} : Refers to the jth object

X_{i*} : Refers to the ith attributes across the n objects.

Framework for Cluster Analysis

- Obtain the Data Matrix

Example:

	1	2	3	4	5
X_1	10	20	30	30	5
X_2	5	20	10	15	10

Framework for Cluster Analysis

- Obtain the Data Matrix
- Standardize the Data Matrix
- Compute the resemblance matrix
  - Resemblance Coefficient

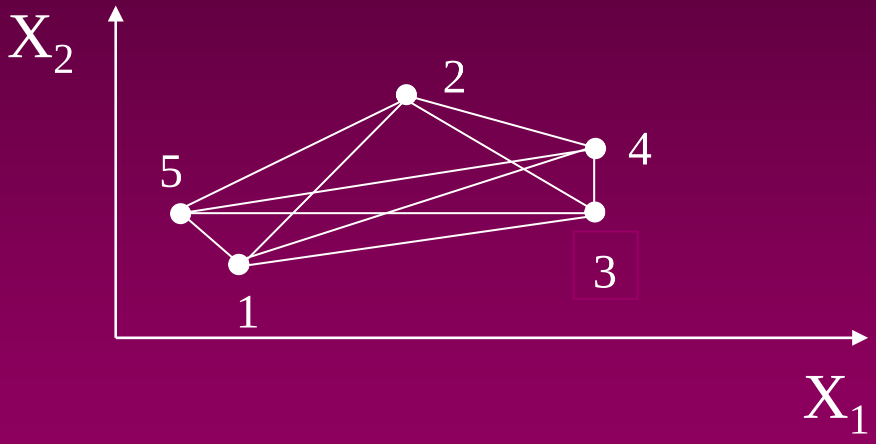
Resemblance Matrix

S=

	1	2	……………………	n-1	n
1	x	x	……………………	x	x
2	S_{21}	x	……………………	x	x
3	S_{31}	S_{32}	x ………………	x	x
.	.	.		.	.
.	.	.		.	.
.	.	.		x	.
n	S_{n1}	S_{n2}	……………………	$S_{n,n-1}$	x

Example:

	1	2	3	4	5
X_1	10	20	30	30	5
X_2	5	20	10	15	10



Euclidean distance $e_{12}=[(10-20)^2 + (5-20)^2]^{1/2} = 18.03$

Example: Resemblance Matrix

Resemblance Matrix

		1	2	3	4	5
	1	x	x	x	X	x
	2	18.03	x	x	X	x
S=	3	20.6	14.1	x	X	x
	4	22.4	11.2	5	X	x
	5	7.07	18.03	25	25.5	x

Framework for Cluster Analysis

- Obtain the Data Matrix
- Standardize the Data Matrix
- Compute the resemblance matrix
  - Resemblance Coefficient
- Execute the clustering method
  - What is the clustering method?

Example:

	1	2	3	4	5
X_1	10	20	30	30	5
X_2	5	20	10	15	10

UPGMA (unweighted pair-group method using arithmetic average) , Euclidean distance:

$$e_{AB} = \frac{1}{|A||B|} \sum_{\substack{i \in A \\ j \in B}} e_{ij}$$

Example

a) Merge 3 and 4 to form one cluster (34) => 1, 2, 5, (34)

$e_{(34)1} = \frac{1}{2}\,[e_{31} + e_{41}] = \frac{1}{2}\,[20.6 + 22.4] = 21.5$

$e_{(34)2} = \frac{1}{2}\,[e_{32} + e_{42}] = \frac{1}{2}\,[14.1 + 11.2] = 12.7$

$e_{(34)5} = \frac{1}{2}\,[e_{35} + e_{45}] = \frac{1}{2}\,[25 + 25.5] = 25.3$

S=

	1	2	5	(34)
1	x	x	x	x
2	18.03	x	x	x
5	7.07	18.03	x	x
(34)	21.5	12.7	25.3	x

Example

b) Merge 1 and 5 to form one cluster (15) => (15), (34), 2

$e_{(15)2} = \frac{1}{2}\,[e_{12} + e_{52}] = \frac{1}{2}\,[18.03 + 18.03] = 18.03$

$e_{(15)(34)} = \frac{1}{4}\,[e_{13} + e_{14} + e_{35} + e_{45}] = \frac{1}{4}\,[20.6 + 22.4 + 25 + 25.5] = 23.4$

S=

	2	(15)	(34)
2	x	x	x
(15)	18.03	x	x
(34)	12.7	23.4	x

Example

c) Merge 2 and (34) to form (234) cluster => (15), (234)

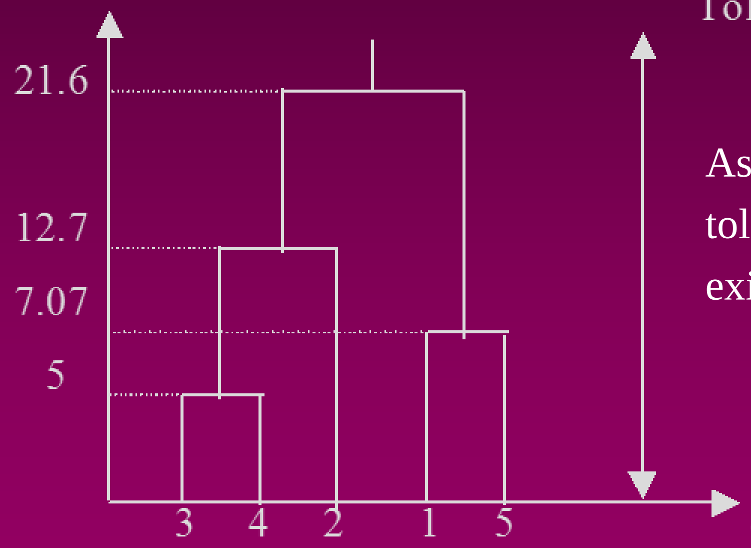
$$e_{(15)(234)} = \frac{1}{6}\left[e_{12} + e_{13} + e_{14} + e_{52} + e_{53} + e_{54}\right] = 21.6$$

S=

	(15)	(234)
(15)	x	x
(234)	21.6	x

Example

d) The last step is to combine these in a single cluster (12345), and from that the tree can be drawn.



As we move higher, more tolerant and allow differences to exist while proclaiming similarities


Data Matrix
Cluster
Analysis
Process
Clusters
Input
Output

Scales for Attributes

Let X be an attribute and A and B be two objects whose scores on the attribute X are X_A and X_B.

- Nominal scale

$X_A = X_B$ or $X_A \neq X_B$

Example: binary variables takes two values - true/false values,

gender takes two values - Male/Female.

colors of a rainbow – 7 values,VIBGYOR

Scales for Attributes

Let X be an attribute and A and B be two objects whose scores on the attribute X are X_A and X_B.

- Nominal scale
- Ordinal scale

  $X_A = X_B$, $X_A > X_B$, or $X_A < X_B$.

  Example: rating on a scale of 1 to 10,

  grades in a course: A, B, C, D, and F.

Scales for Attributes

Let X be an attribute and A and B be two objects whose scores on the attribute X are X_A and X_B.

- Nominal scale
- Ordinal scale
- Interval scale

If $X_A > X_B$, one can say A is $X_A - X_B$ units difference than B.

Example: when $X_A=10^oC$ and $X_B=35\,^oC$,one can say A is cooler than B by 25^oC.

Scales for Attributes

Let X be an attribute and A and B be two objects whose scores on the attribute X are X_A and X_B.

- Nominal scale
- Ordinal scale
- Interval scale
- Ratio scale

If $X_A > X_B$, then one can say that A is $\frac{X_A}{X_B}$ times superior to B.

Example: Salary.

Scales for Attributes

Let X be an attribute and A and B be two objects whose scores on the attribute X are X_A and X_B.

- Nominal scale
- Ordinal scale
- Interval scale
- Ratio scale

  - Quantitative / Qualitative attributes

Standardizing the Data Matrix

$$
X = \begin{array}{c|cccc|}
 & 1 & 2 & \ldots\ldots\ldots & n \\
\hline
1 & X_{11} & X_{12} & \ldots\ldots\ldots & X_{1n} \\
2 & X_{21} & X_{22} & \ldots\ldots\ldots & X_{2n} \\
3 & X_{31} & X_{32} & \ldots\ldots\ldots & X_{3n} \\
. & . & . & & . \\
. & . & . & & . \\
. & . & . & & . \\
m & X_{m1} & X_{m2} & \ldots\ldots\ldots & X_{mn} \\
\hline
\end{array}
$$

Data Matrix

$$
Z = \begin{array}{c|cccc|}
 & 1 & 2 & \ldots\ldots\ldots & n \\
\hline
1 & Z_{11} & Z_{12} & \ldots\ldots\ldots & Z_{1n} \\
2 & Z_{21} & Z_{22} & \ldots\ldots\ldots & Z_{2n} \\
3 & Z_{31} & Z_{32} & \ldots\ldots\ldots & Z_{3n} \\
. & . & . & & . \\
. & . & . & & . \\
. & . & . & & . \\
m & Z_{m1} & Z_{m2} & \ldots\ldots\ldots & Z_{mn} \\
\hline
\end{array}
$$

Standardized Data Matrix

Standardizing the Data Matrix

Let X_{ij} be the score of object j on the attribute i,

X_{i*} the i^{th} row of X in the data matrix X,

X_{*j} the j^{th} column of X in the data matrix X,

Define

$\overline{X}_{i*} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$ as the i^{th} row average, and

$\overline{X}_{*j} = \frac{1}{m}\sum_{i=1}^{m} X_{ij}$ as the i^{th} column average.

Standardizing the Data Matrix

Method 1

$$Z_{ij} = \frac{X_{ij} - \overline{X}_{i*}}{S_{i*}}$$

$-2.0 \leq Z_{ij} \leq +2.0$

where $$S_{i*} = \left[\frac{1}{n-1} \sum_{j=1}^{n} \left(X_{ij} - \overline{X}_{i*} \right)^2 \right]^{\frac{1}{2}}$$

is the sample standard deviation of the i^{th} row.

Standardizing the Data Matrix

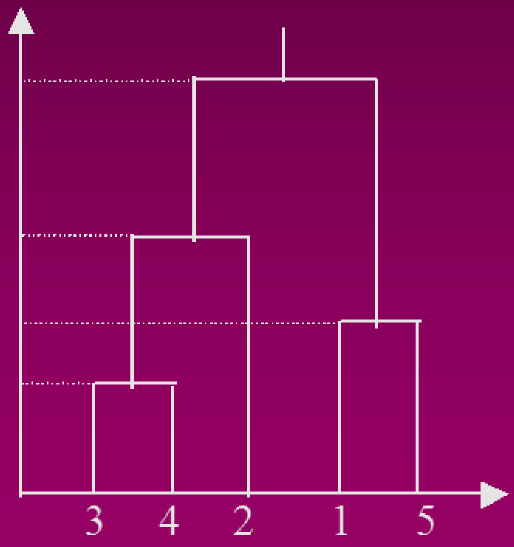
Example 1:

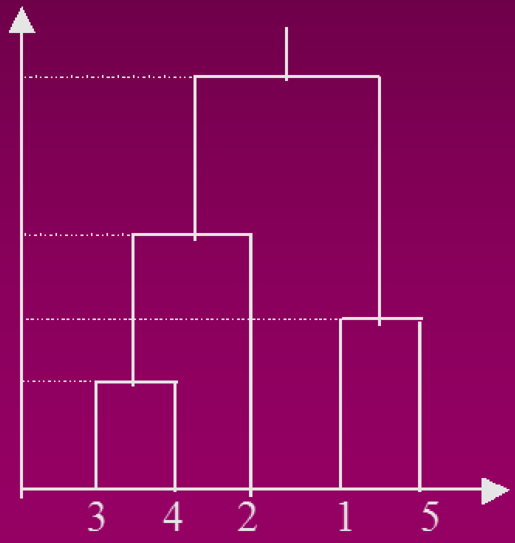
Data Matrix

X=	1	2	3	4	5	$\overline{X}$	S
1	10	20	30	30	5	19	9.34
2	5	20	10	15	10	12	5.70

Standardized Data Matrix

Z=	1	2	3	4	5
1	-0.96	0.11	1.18	1.18	-1.5
2	-1.23	1.4	-0.35	0.35	-0.35





Standardizing the Data Matrix

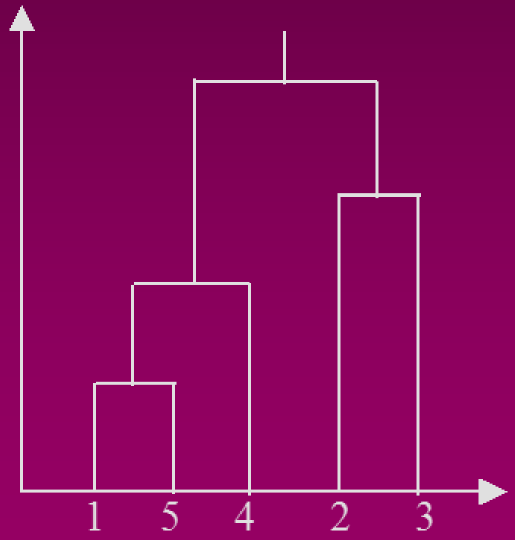
Example 2:

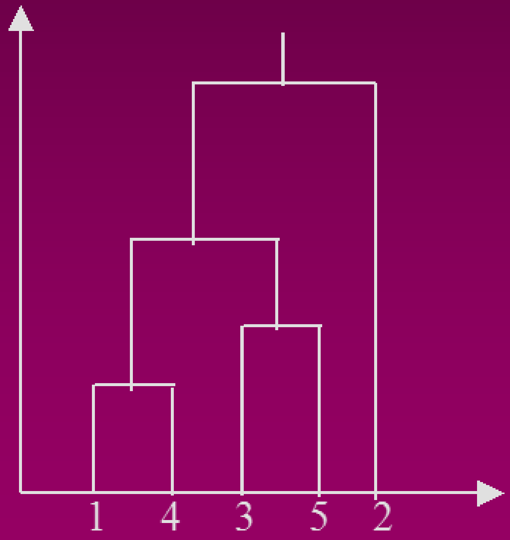
Data Matrix

X=	1	2	3	4	5	$\overline{X}$	S
1	20	24	21	19	23	21.4	2.07
2	19	6	21	24	18	15.8	6.94

Standardized Data Matrix

Z=	1	2	3	4	5
1	-0.68	1.25	-0.19	-1.16	0.77
2	0.46	-1.41	-0.55	1.18	0.32





Standardizing the Data Matrix

Other Standardizing techniques:

1. Transformation

   $Z_{ij} = \log(X_{ij})$, or , $Z_{ij} = \sqrt{X_{ij}}$ …. etc.

2. Removing outliers.

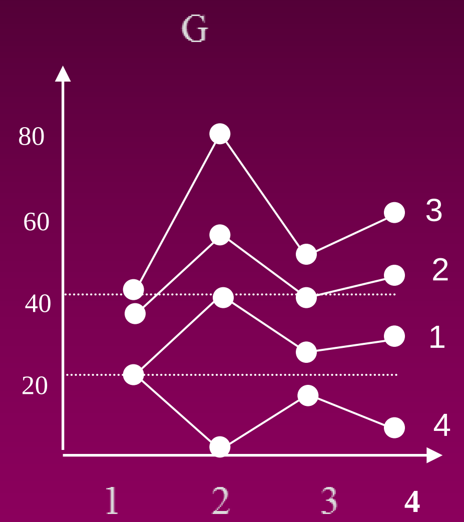
Resemblance Coefficients for Quantitative attributes

$$\binom{N}{2} = \frac{1}{2} N(N-1)$$

Resemblance Coefficients for Quantitative attributes

$$\binom{N}{2} = \frac{1}{2}N(N-1)$$

		1	2	3	4	Mean	STD
X=	1	20	35	40	20	28.75	10.31
	2	40	55	80	0	43.75	33.51
	3	25	40	50	15	32.5	15.55
	4	30	45	60	10	36.25	21.36
Mean		28.75	43.75	57.5	11.25		
STD		8.54	8.54	17.07	8.54		



X is a data matrix consisting of 4 objects and 4 attributes. The graph G on the right depicts these objects. Notice that Object 2=object 1+15 (addition), Object 3=object 1*2 (multiplication), and Object 4 is a mirror image of Object 1 with respect to 20.

Resemblance Coefficients for Quantitative attributes

1. Euclidean Distance Coefficient, e_{jk}
   (dissimilarity coefficient)

$$e_{jk} = \left[\sum_{i=1}^{m} (X_{ij} - X_{ik})^2 \right]^{\frac{1}{2}} \quad , 0 \leq e_{jk} \leq \infty$$

$$e_{jk} = \langle X_{*j} - X_{*k}, X_{*j} - X_{*k} \rangle^{\frac{1}{2}}$$

$$= [(X_{*j} - X_{*k})^T (X_{*j} - X_{*k})]^{\frac{1}{2}}$$

Resemblance Coefficients for Quantitative attributes

2. Average Euclidean Distance Coefficient, d_{jk}

(dissimilarity coefficient)

$$d_{jk} = \left[\frac{1}{m}\sum_{i=1}^{m}(X_{ij} - X_{ik})^2\right]^{\frac{1}{2}} \qquad , 0 \le d_{jk} \le \infty$$

$$d_{jk} = [\,\frac{1}{m}\,(X_{*j} - X_{*k})^{T}\,(X_{*j} - X_{*k})]^{\frac{1}{2}}$$

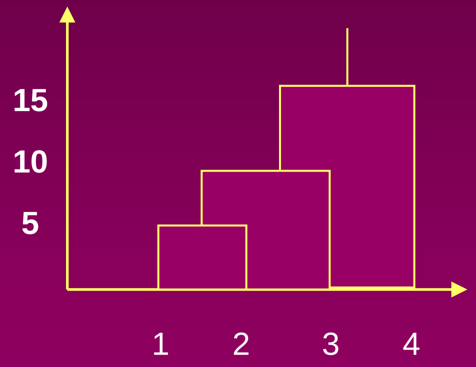
Resemblance Coefficients for Quantitative attributes

2. Average Euclidean Distance Coefficient, d_{jk}

Example (UPGMA):

S=

x	x	x	x
15	x	x	x
29.7	15.6	x	x
22.9	35.7	51.3	x



Resemblance Coefficients for Quantitative attributes

3. The Coefficient of Shape Difference, Z_{jk}
   (dissimilarity coefficient)

$$Z_{jk} = \left[\frac{m}{m-1}\left(d_{jk}^{2} - q_{jk}^{2}\right)\right]^{\frac{1}{2}} \quad , 0 \leq Z_{jk} \leq \infty$$

where $$qij = \frac{1}{m^{2}}\left(\sum_{i=1}^{m} X_{ij} - \sum_{i=1}^{m} X_{ik}\right)^{2}$$

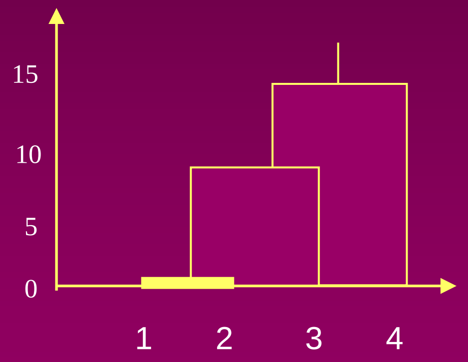
Resemblance Coefficients for Quantitative attributes

3. The Coefficient of Shape Difference, Z_{jk}

Example (UPGMA):

S=

x	x	x	x
0	x	x	x
8.54	8.54	x	x
17.1	17.1	25.6	x



Resemblance Coefficients for Quantitative attributes

4. The Cosine Coefficient, C_{jk}
(similarity coefficient)

$$C_{jk} = \frac{\sum_{i=1}^{m} X_{ij} X_{ik}}{\left(\sum_{i=1}^{m} X_{ij}^{2}\right)^{\frac{1}{2}} \left(\sum_{i=1}^{m} X_{ik}^{2}\right)^{\frac{1}{2}}} \quad , -1.0 \le C_{jk} \le 1.0$$

$$C_{jk} = \mathrm{COS}\,\alpha = \frac{X_{*j}^{T} X_{*k}}{\|X_{*j}\| \, \|X_{*k}\|}$$

where α is the angle between vectors j and k.

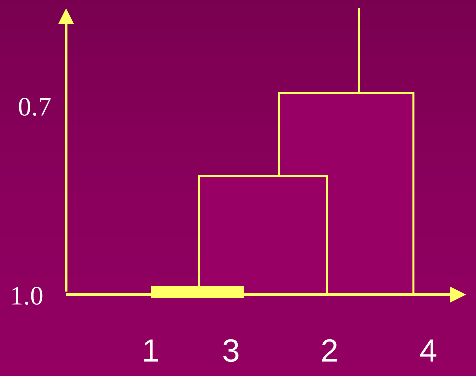
Resemblance Coefficients for Quantitative attributes

4. The Cosine Coefficient, C_{jk}

Example (UPGMA):

S=

x	x	x	x
0.996	x	x	x
1.0	0.996	x	x
0.672	0.732	0.672	x



Resemblance Coefficients for Quantitative attributes

5. The Correlation Coefficient, r_{jk}
(similarity coefficient)

$$r_{jk} = \frac{\sum_{i=1}^{m} X_{ij}X_{ik} - \frac{1}{m}\left(\sum_{i=1}^{m} X_{ij}\right)\left(\sum_{i=1}^{m} X_{ik}\right)}{\left[\sum_{i=1}^{m} X_{ij}{}^2 - \frac{1}{m}\left(\sum_{i=1}^{m} X_{ij}\right)^2\right]^{\frac{1}{2}}\left[\sum_{i=1}^{m} X_{ik}{}^2 - \frac{1}{m}\left(\sum_{i=1}^{m} X_{ik}\right)^2\right]^{\frac{1}{2}}} \quad , -1.0 \le C_{jk} \le 1.0$$

$$r_{jk} = \frac{\sum_{i=1}^{m} \left(X_{ij} - \overline{X}_{*j}\right)\left(X_{ik} - \overline{X}_{*k}\right)}{\left[\sum_{i=1}^{m} \left(X_{ij} - \overline{X}_{*j}\right)^2\right]^{\frac{1}{2}}\left[\sum_{i=1}^{m} \left(X_{ik} - \overline{X}_{*k}\right)^2\right]^{\frac{1}{2}}}$$

$$= \frac{COV\ (X_{*j}, X_{*k})}{\left[VAR\ (X_{*j})\ VAR\ (X_{*k})\right]^{\frac{1}{2}}}$$

= Cosine of the angle between the centered vectors

$(X_{1j} - \overline{X}_{*j}, X_{2j} - \overline{X}_{*j}, \ldots\ldots, X_{mj} - \overline{X}_{*j})^T$, and

$(X_{1k} - \overline{X}_{*k}, X_{2k} - \overline{X}_{*k}, \ldots\ldots, X_{mk} - \overline{X}_{*k})^T$.

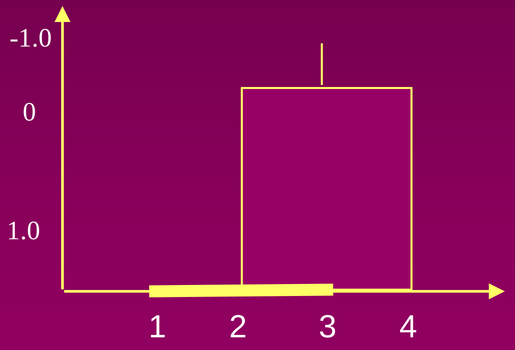
Resemblance Coefficients for Quantitative attributes

5. The Correlation Coefficient, r_{jk}

Example (UPGMA):

S=

x	x	x	x
1.0	x	x	x
1.0	1.0	x	x
-1.0	-1.0	-1.0	x



Resemblance Coefficients for Quantitative attributes

Coefficient		Range	Insensitive To	
			Addition	Multiplication
Dissimilarity	e_{jk}	$0.0 \leq e_{jk} \leq \infty$	No	No
	d_{jk}	$0.0 \leq d_{jk} \leq \infty$	No	No
	a_{jk}	$0.0 \leq a_{jk} \leq 1.0$	No	No
	b_{jk}	$0.0 \leq b_{jk} \leq 1.0$	No	No
	z_{jk}	$0.0 \leq z_{jk} \leq \infty$	Yes	No
Similarity	c_{jk}	$-1.0 \leq c_{jk} \leq 1.0$	No	Yes
	r_{jk}	$-1.0 \leq r_{jk} \leq 1.0$	Yes	Yes

Hierarchical Clustering

- Agglomerative vs divisive hierarchical algorithm
- The Basic framework for the Agglomerative Algorithms

$$
\begin{array}{ccccccc}
 & & 1 & 2 & \cdots\cdots & n-1 & n \\
 & 1 & x & x & \cdots\cdots & x & x \\
 & 2 & S_{21} & x & \cdots\cdots & x & x \\
S= & 3 & S_{31} & S_{32} & x \ \cdots\cdots & x & x \\
 & . & . & . & & . & . \\
 & . & . & . & & . & . \\
 & . & . & . & & x & . \\
 & n & S_{n1} & S_{n2} & \cdots\cdots & S_{n,n-1} & x
\end{array}
$$

Hierarchical Clustering

- Agglomerative vs divisive hierarchical algorithm
- The Basic framework for the Agglomerative Algorithms

1 **Begin with n clusters each with one object. Let the clusters be labeled 1 through n.**

2 **Search the resemblance matrix for the most similar pair of clusters. Let p and q be the two similar clusters, with S_{pq} as their similarity measure with p>q.**

3 **Reduce the number of clusters by 1 by merging the two clusters p and q. Label the new cluster q and update the resemblance matrix objects to reflect the revised similarity between the new cluster q and other existing clusters other than p. Delete the row and column of S that corresponds to the cluster p.**

4 **Perform step 2 and step 3 a total of (n-1) times. At each stage record the elements of each cluster and keep track of all similarity measures at each stage to have a complete record.**

Hierarchical Clustering

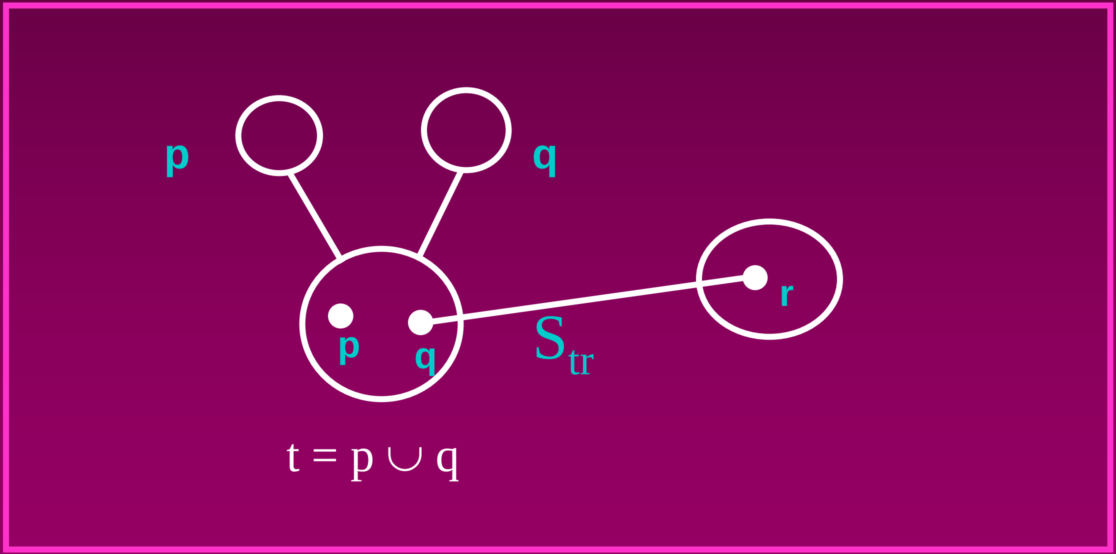
1. Single Linkage Method (SLINK)

$$S_{tr} = \min \{ S_{ij} \mid i \in t = p \cup q, \; j \in r \}$$

Hierarchical Clustering

1. Single Linkage Method (SLINK)

$$S_{tr} = \min \{ S_{ij} \mid i \in t = p \cup q, \; j \in r \}$$



1. Single Linkage Method (SLINK)

Example

X is the data matrix

X		1	2	3	4	5
	x_1	10	20	30	30	5
	x_2	5	20	10	15	10

STEP 1: Compute resemblance matrix, S, using Euclidean distance, e_{ij}

S=	1	2	3	4	5
1	x	x	x	x	x
2	18.03	x	x	x	x
3	20.6	14.1	x	x	x
4	22.4	11.2	5	x	x
5	7.07	18.03	25	25.5	x

Merge (3) and (4) to get (34)

1. Single Linkage Method (SLINK)

Example

STEP 2: Update S

$e_{(34)1} = \min\{e_{31}, e_{41}\} = \min\{20.6, 22.4\} = 20.6$

$e_{(34)2} = \min\{e_{32}, e_{42}\} = \min\{14.1, 11.2\} = 11.2$

$e_{(34)5} = \min\{e_{35}, e_{45}\} = \min\{25, 25.5\} = 25$

S=

	1	2	5	(34)
1	x	x	x	x
2	18.03	x	x	x
5	7.07	18.03	x	x
(34)	20.6	11.2	25	x

Merge (1) and (5) to get (15)

1. Single Linkage Method (SLINK)

Example

STEP 3: Update S

$e_{(15)2} = \min\{e_{12}, e_{52}\} = \min\{18.03, 18.03\} = 18.03$

$e_{(15)(34)} = \min\{e_{13}, e_{14}, e_{53}, e_{54}\} = \min\{20.6, 22.4, 25, 25.5\} = 20.6$

S=

	2	(34)	(15)
2	x	x	x
(34)	11.2	x	x
(15)	18.03	20.6	x

Merge (2) and (34) to get (234)

1. Single Linkage Method (SLINK)

Example

STEP 4: Update S

$$e_{(15)(234)} = \min\{ e_{12}, e_{13}, e_{14}, e_{52}, e_{53}, e_{54}\}$$
$$= \min\{18.03, 20.6, 22.4, 18.03, 25, 25.5\} = 18.03$$

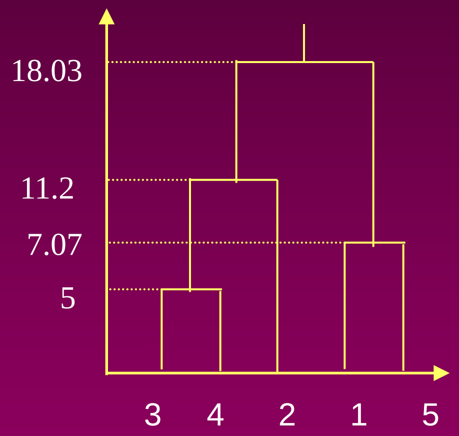
S=

	(15)	(234)
(15)	x	x
(234)	18.03	x

Merge (15) and (234) to get (12345)

1. Single Linkage Method (SLINK)

Example



Hierarchical Clustering

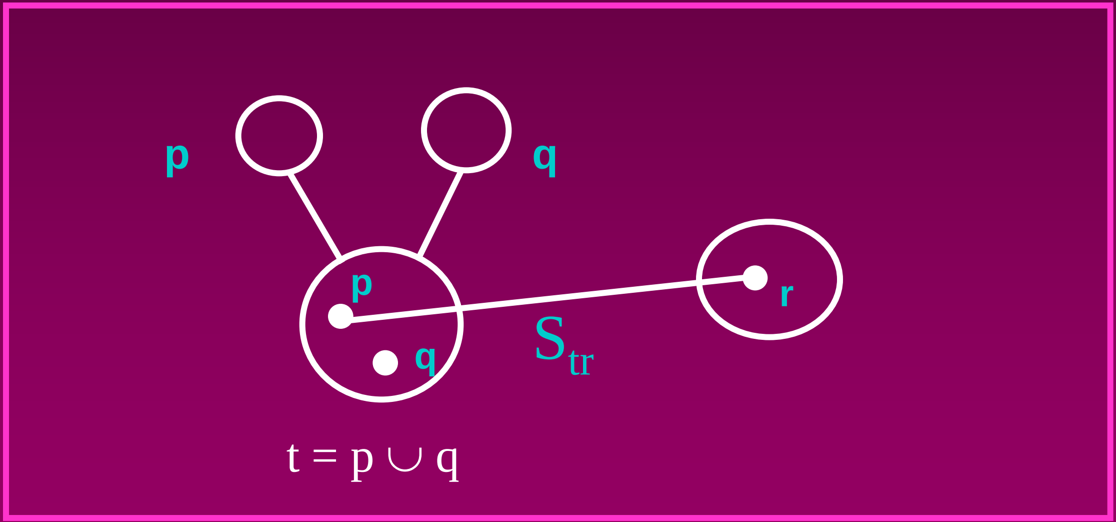
2. Complete Linkage Method (CLINK)

$$S_{tr} = \max \{ S_{ij} \mid i \in t = p \cup q, \quad j \in r \}$$

Hierarchical Clustering

2. Complete Linkage Method (CLINK)

$$S_{tr} = \max \{ S_{ij} \mid i \in t = p \cup q, \;\; j \in r \}$$



2. Complete Linkage Method (CLINK)

Example

X is the data matrix

X		1	2	3	4	5
	x_1	10	20	30	30	5
	x_2	5	20	10	15	10

STEP 1: Compute resemblance matrix, S, using Euclidean distance, e_{ij}

S=	1	2	3	4	5
1	x	x	x	x	x
2	18.03	x	x	x	x
3	20.6	14.1	x	x	x
4	22.4	11.2	5	x	x
5	7.07	18.03	25	25.5	x

Merge (3) and (4) to get (34)

2. Complete Linkage Method (CLINK)

Example

STEP 2: Update S

$e_{(34)1} = \max\{e_{31}, e_{41}\} = \max\{20.6, 22.4\} = 22.4$

$e_{(34)2} = \max\{e_{32}, e_{42}\} = \max\{14.1, 11.2\} = 14.1$

$e_{(34)5} = \max\{e_{35}, e_{45}\} = \max\{25, 25.5\} = 25.5$

S=

	1	2	5	(34)
1	x	X	x	x
2	18.03	X	x	x
5	7.07	18.03	x	x
(34)	22.4	14.1	25.5	x

Merge (1) and (5) to get (15)

2. Complete Linkage Method (CLINK)

Example

STEP 3: Update S

$e_{(15)2} = \max\{e_{12}, e_{52}\} = \max\{18.03, 18.03\} = 18.03$

$e_{(15)(34)} = \max\{e_{13}, e_{14}, e_{53}, e_{54}\} = \max\{20.6, 22.4, 25, 25.5\} = 25.5$

S=

	2	(34)	(15)
2	x	x	x
(34)	14.1	x	x
(15)	18.03	25.5	x

Merge (2) and (34) to get (234)

2. Complete Linkage Method (CLINK)

Example

STEP 4: Update S

$e_{(15)(234)} = \max\{ e_{12}, e_{13}, e_{14}, e_{52}, e_{53}, e_{54}\}$

$= \max\{18.03, 20.6, 22.4, 18.03, 25, 25.5\} = 25.5$

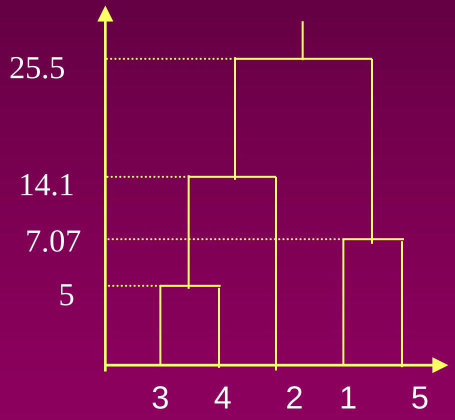
S=

	(15)	(234)
(15)	x	x
(234)	25.5	x

Merge (15) and (234) to get (12345)

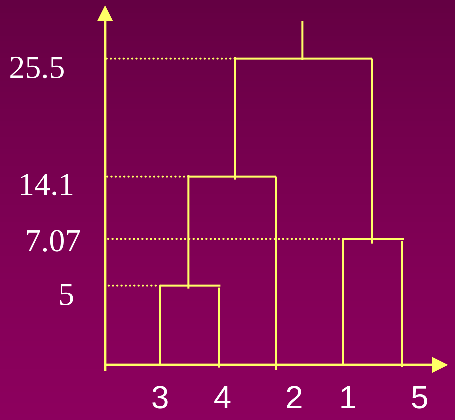
2. Complete Linkage Method (CLINK)

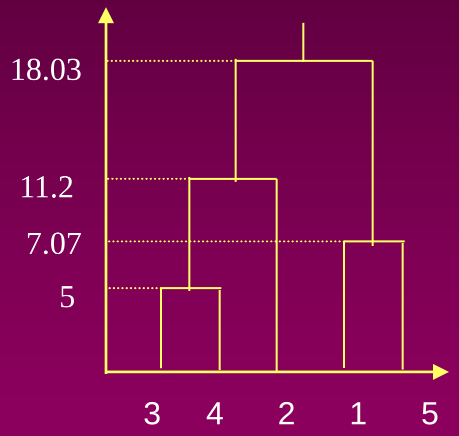
Example



2. Complete Linkage Method (CLINK)

Example





Hierarchical Clustering

3. Ward's Minimum Variance Clustering Method

At each step it makes whichever merger of two clusters that will result in the smallest increase in the value of variance, E. The value of E at the beginning is zero, E=0.

3. Ward's Minimum Variance Clustering Method

Example

X is the data matrix

X		1	2	3	4	5
	x_1	10	20	30	30	5
	x_2	5	20	10	15	10

$$E = \underbrace{(10-10)^2+(5-5)^2}_{\text{cluster 1}} + \underbrace{(20-20)^2+(20-20)^2}_{\text{cluster 2}} + \underbrace{(30-30)^2+(10-10)^2}_{\text{cluster 3}}$$

$$+ \underbrace{(30-30)^2+(15-15)^2}_{\text{cluster 4}} + \underbrace{(5-5)^2+(10-10)^2}_{\text{cluster 5}} = 0.0$$

3. Ward's Minimum Variance Clustering Method

Example

STEP 1 Compute E for all possible mergers,

Possible Mergers				E
(12)	3	4	5	162.5
(13)	2	4	5	212.5
(14)	2	3	5	250
(15)	2	3	4	25
(23)	1	4	5	100
(24)	1	3	5	62.5
(25)	1	3	4	162.5
(34)	1	2	5	12.5
(35)	1	2	4	312.5
(45)	1	2	3	325

merging (3) and (4),
gives 1 , 2 , (34) , 5
at the value of E = 12.5

3. Ward’s Minimum Variance Clustering Method

Example

To show how E computed, let’s take the first one: (12), 3, 4, 5 .

First, we must calculate the mean for (12). It is

$$\frac{10+20}{2}=15 \quad , \quad \frac{5+20}{2}=125$$

For the first possible merger the value of E is

3. Ward's Minimum Variance Clustering Method

Example

$$E = \underbrace{(10-15)^2+(5-12.5)^2 + (20-15)^2+(20-12.5)^2}_{\text{cluster (12)}}$$

$$+ \underbrace{(30-30)^2+(10-10)^2}_{\text{cluster 3}} + \underbrace{(30-30)^2+(15-15)^2}_{\text{cluster 4}}$$

$$+ \underbrace{(5-5)^2+(10-10)^2}_{\text{cluster 5}} = 162.5$$

3. Ward's Minimum Variance Clustering Method

Example

STEP 2 With Ward's method, objects merge at previous clustering steps are never unmerged. Thus, at the beginning of step 2 there are six possible mergers of two clusters.

Possible Mergers			E
(34)	(12)	5	175.0
(34)	(15)	2	37.5
(34)	(25)	1	175.0
(134)	2	5	316.7
(234)	1	5	116.7
(345)	1	2	433.3

merging (1) and (5), gives 2, (34) , (15) at the value of E = 37.5

3. Ward's Minimum Variance Clustering Method

Example

The set of clusters 2, (34) , (15) is chosen because it gives the smallest value of E;

$E = \underbrace{(20-20)^2+(20-20)^2}_{\text{cluster 2}}$

$+ \underbrace{(30-30)^2+(10-12.5)^2 + (30-30)^2+(15-12.5)^2}_{\text{cluster (34)}}$

$+ \underbrace{(10-7.5)^2+(5-7.5)^2 + (5-7.5)^2+(10-7.5)^2}_{\text{cluster 5}} = 37.5$

3. Ward's Minimum Variance Clustering Method

Example

STEP 3 Compute E for all possible mergers.

Possible Mergers		E
(234)	(15)	141.7
(125)	(34)	245.9
(1345)	2	568.6

merging (2) and (34), gives (15) , (234) at the value of E = 141.7

3. Ward's Minimum Variance Clustering Method

Example

<u>STEP 4</u> Only one other merger is possible, that's (12345). The cluster mean is:

$$\frac{10+20+30+30+5}{5}=19 \quad , \qquad \frac{5+20+10+15+10}{5}=12$$

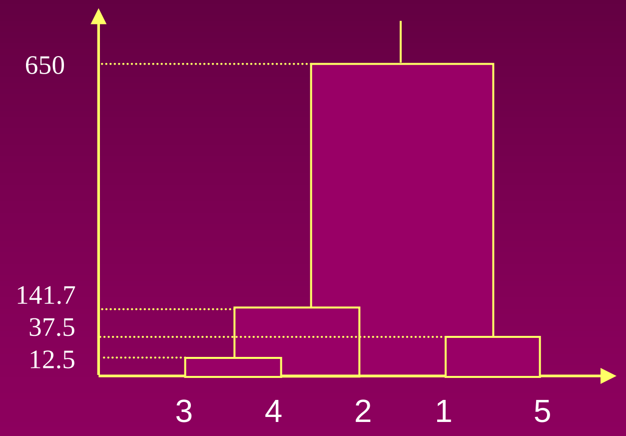
The value of E is

$$E=(10-19)^2+(5-12)^2+(20-19)^2+(20-12)^2+(30-19)^2+(10-12)^2+$$
$$(30-19)^2+(15-12)^2+(5-19)^2+(10-12)^2=650$$

merging (15) and (234), gives (12345) at the value of E = 650.

3. Ward's Minimum Variance Clustering Method

Example



Hierarchical Clustering

4. Graph Theory algorithm for Single-Linkage

Assume a dissimilarity matrix.

1 **begin with the disjoint clustering, which places each object in its own cluster.**
Find a MST on G(∞)
Repeat steps 2 and 3 until all objects are in one cluster.

2 **Merge the two clusters connected by the MST edge with the smallest weight to define the next clustering.**

3 **Replace the weight of the edge selected in STEP 2 by a weight larger than the largest similarity.**

A divisive algorithm is just as simple. Cut the edge in the MST in the order of weight, cutting the largest first. Each cut defines a new clustering, with those objects connected in the MST at any stage belonging to the same cluster.

4. Graph Theory algorithm for Single-Linkage

Example

Let S be the dissimilarity matrix.

S=

	1	2	3	4	5
1	x	x	x	X	x
2	2.3	x	x	x	x
3	3.4	2.6	x	x	x
4	1.2	1.8	4.2	x	x
5	3.7	4.6	0.7	4.4	x

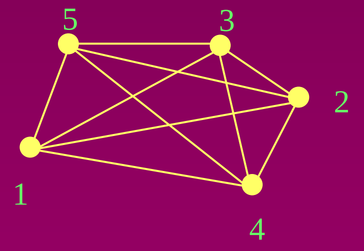
4. Graph Theory algorithm for Single-Linkage

Example

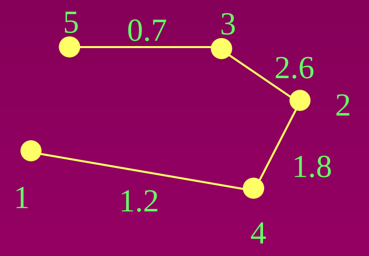
Let S be the dissimilarity matrix.

S=

	1	2	3	4	5
1	x	x	x	X	x
2	2.3	x	x	x	x
3	3.4	2.6	x	x	x
4	1.2	1.8	4.2	x	x
5	3.7	4.6	0.7	4.4	x



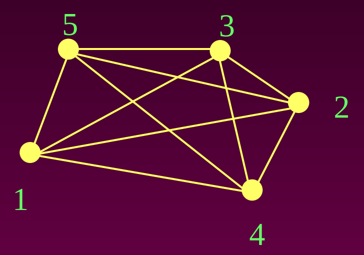
Complete Graph $G(\infty)$



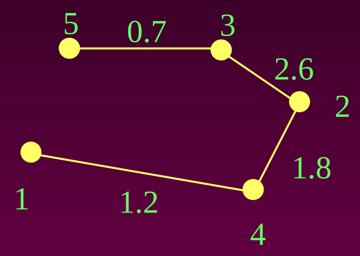
MST

4. Graph Theory algorithm for Single-Linkage

Example

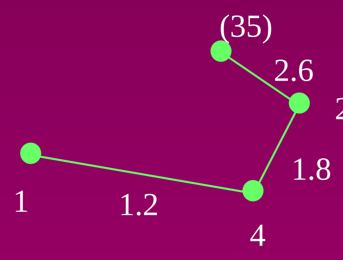


Complete Graph G(∞)

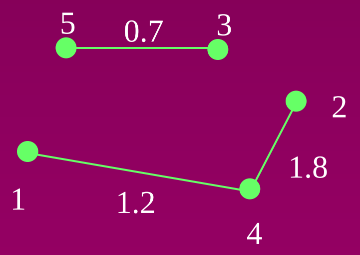


MST

Agglomerative



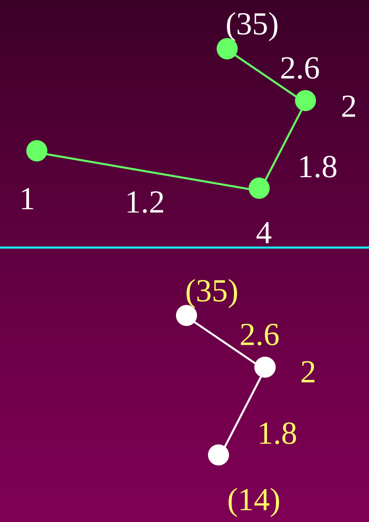
Divisive

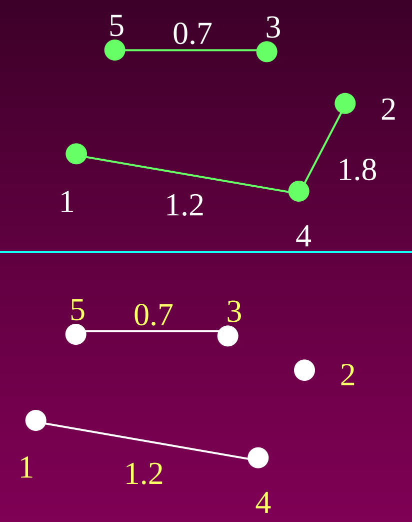


4. Graph Theory algorithm for Single-Linkage

Example

Agglomerative | **Divisive**

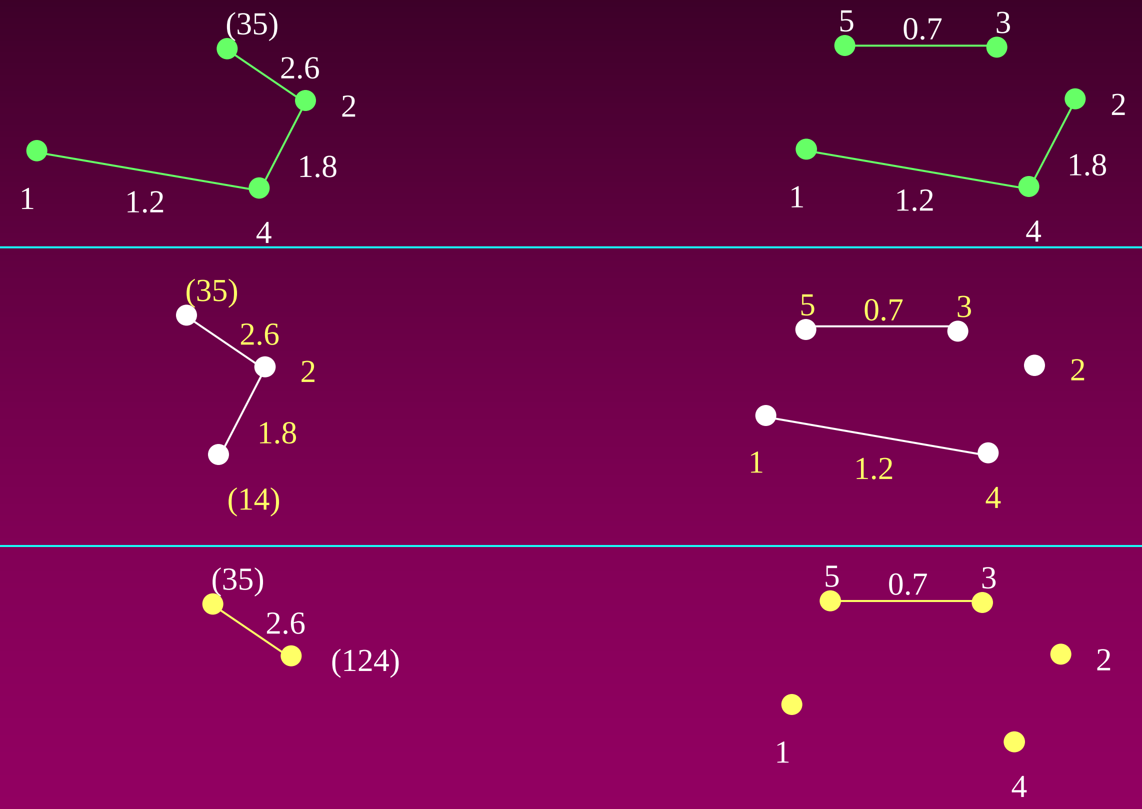




4. Graph Theory algorithm for Single-Linkage

Example

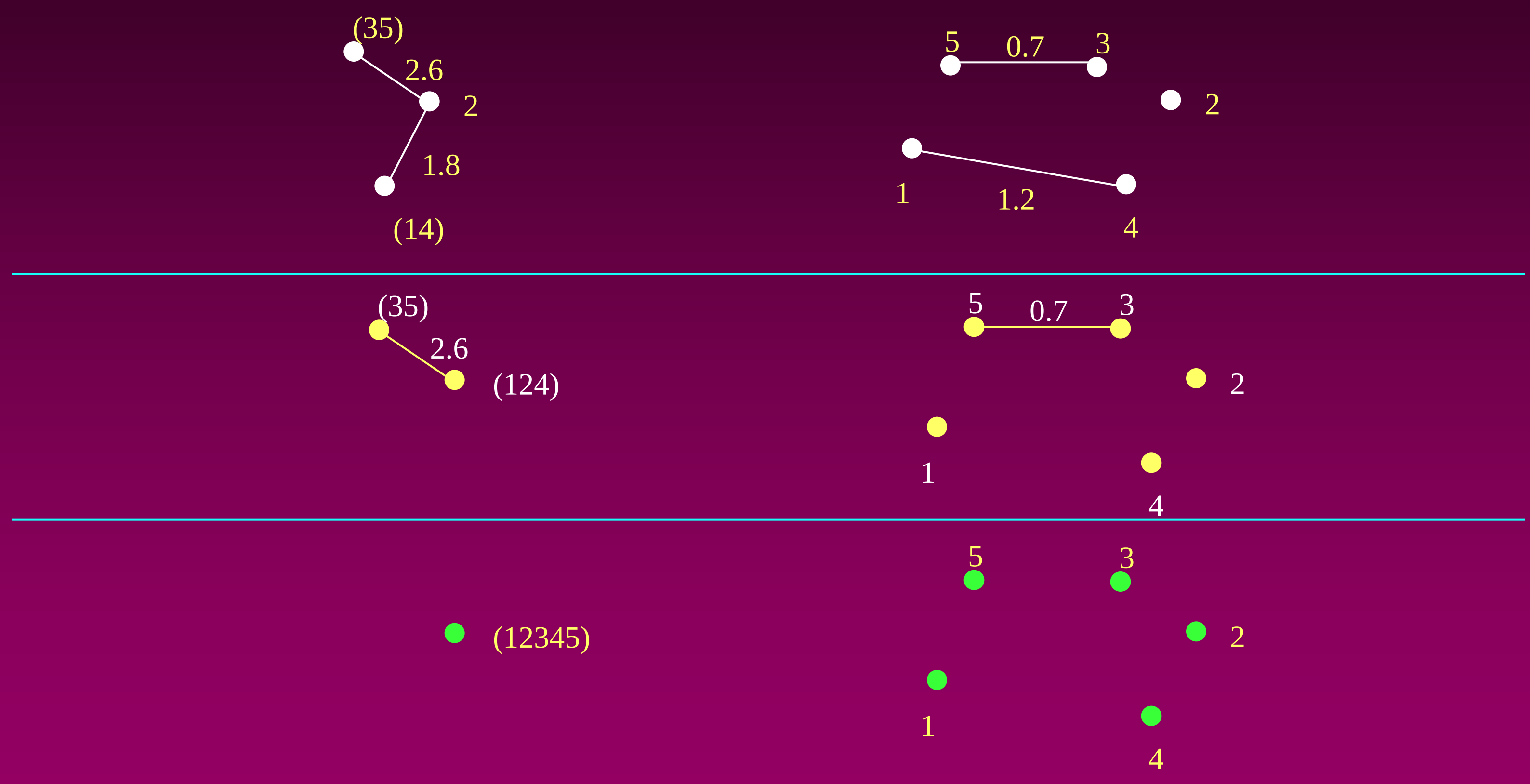
Agglomerative | **Divisive**



4. Graph Theory algorithm for Single-Linkage

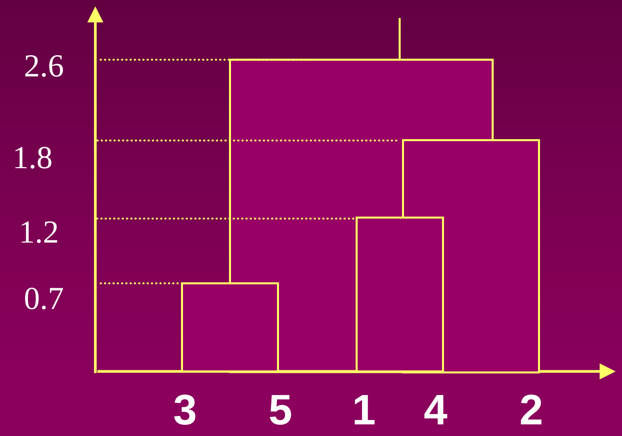
Example

Agglomerative **Divisive**



4. Graph Theory algorithm for Single-Linkage

Example



Partitional Clustering

- Statement of the problem of partitional clustering.
- The basic idea of partitional clustering method
- Initial Partition
  - Seed Points

Partitional Clustering

Seed Points:

1. Choose the first k objects in the data set.

2. Label the objects from 1 to n and choose those labeled n/k, 2n/k, ..., (k-1)n/k, and n.

3. Subjectively choose any k objects from the data set.

4. Label the objects from 1 to n and choose the objects corresponding to k different random numbers in the range [1,n].

5. Take any desired partition of the objects into k mutually exclusive clusters and compute the cluster centroids as seed points.

Partitional Clustering

- Statement of the problem of partitional clustering.
- The basic idea of partitional clustering method
- Initial Partition
  - Seed Points
  - Initial Partition

Partitional Clustering

Initial Partition:

1. Assign each object to the cluster built around the nearest seed point. This point remains stationary throughout one full pass over all objects.

2. Let each seed point to form a cluster of one member. Then assign objects one at a time to the cluster with the nearest centroid; after an object is assigned to a cluster, update the centroid so that it is the true mean vector for all the objects currently in that cluster.

3. Use hierarchical clustering to obtain an initial partition.

4. The analyst could use his judgment to sort the set of objects into an initial partition

5. The analyst could rely on some random allocation schemes.

Partitional Clustering

- Statement of the problem of partitional clustering.
- The basic idea of partitional clustering method
- Initial Partition
  - Seed Points
  - Initial Partition
- Criteria for Partitional Clustering

Partitional Clustering

Criteria for Partitional Clustering :

Let $X = [X_{*1}, X_{*2}, \ldots\ldots\ldots, X_{*n}]_{m \times n}$

The problem is to partition X into k clusters, such that

$X = C_1 \cup C_2 \cup \ldots\ldots \cup C_k$, and $C_i \cap C_j = \phi$, $1 \leq i, j \leq k$ and $i \neq j$

Let $|C_i| = n_i$ and $\sum_{i=1}^{k} n_i = n$

Partitional Clustering

*Criteria for Partitional Clustering **:***

1. Sum-of-Squared error Criterion

Criteria for Partitional Clustering :

1. Sum-of-Squared error Criterion

The centroid of cluster C_i,

$$m_i = \frac{1}{n_i} \sum_{X*j \in C_i} X*_j$$

The square-error J_i for cluster C_i is the sum of the squared Euclidean distance between each object in C_i and its cluster centroid m_i,

$$J_i = \sum_{X*j \in C_i} \left\| X*_j - m_i \right\|^2$$

$$= \sum_{X*j \in C_i} \left(X*_j - m_i\right)^T \left(X*_j - m_i\right)$$

The square-error, J_e, for the entire clustering containing k clusters is the sum of square-error of the individual clusters,

$$J_e = \sum_{i=1}^{k} J_i$$

Criteria for Partitional Clustering :

1. Sum-of-Squared error Criterion

$$J_e = \sum_{i=1}^{k} J_i$$

The objective of a partitional clustering algorithm based on the square-error criterion is to find a partition that minimizes J_e.

Partitional Clustering

Criteria for Partitional Clustering **:**

1. Sum-of-Squared error Criterion

2. Scatter matrix Criterion

2. Scatter matrix Criterion

Let m_i be the mean of the i^{th} cluster C_i ,and m be the pooled mean of all objects in X,

$$m_i = \frac{1}{n_i} \sum_{X*j \in C_i} X*_j \quad , \quad m = \frac{1}{n} \sum_{j=1}^{n} X*_j = \frac{1}{n} \sum_{j=1}^{n} n_i m_i$$

Define S_i to be the scatter matrix for the i^{th} cluster,

$$S_i = \sum_{X*j \in C_i} (X*_j - m_i)(X*_j - m_i)^T$$

The within-cluster, S_W, is the sum of scatter matrices of the individual clusters,

$$S_w = \sum_{i=1}^{k} S_i$$

The between-cluster scatter matrix, S_B , is defined as

$$S_B = \sum_{i=1}^{k} n_i (m_i - m)(m_i - m)^T$$

$S_T = S_B + S_W$

The total clusters scatter matrix, S_T , is defined as

$$S_T = \sum_{X*j \in C_i} (X*_j - m)(X*_j - m)^T$$

2. Scatter matrix Criterion

A good partition can be obtained by minimizing the trace of S_W.

$$tr(S_w) = \sum_{i=1}^{k} t_r(S_i)$$

By expansion

$$tr(S_w) = \sum_{i=1}^{k} \sum_{X*j \in C_i} (X*_j - m_i)^T (X*_j - m_i)$$

$$tr(S_W) = J_e$$

Therefore minimizing $tr(S_W)$ immediately implies that

$$tr(S_B) = \sum_{i=1}^{k} n_i (m_i - m)^T (m_i - m)$$

is maximized, and hence, the resulting partition is optimal.

Partitional Clustering

- Statement of the problem of partitional clustering.
- The basic idea of partitional clustering method
- Initial Partition
  - Seed Points
  - Initial Partition
- Criteria for Partitional Clustering
- Partitional Clustering algorithms

Partitional Clustering algorithms:

1. Frogy’s algorithms

1. Begin with any desired initial partition. Go to step 2 if beginning with a set of seed points; go to step 3 if beginning with a partition of the objects.

2. Allocate each object to the cluster with the nearest seed point. The seed points remain fixed for a full cycle through the entire set of objects.

3. Compute new seed points as the centroids of the clusters of objects.

4. .Repeat steps 2and 3 until no objects change their clsuter membership at step 2.

Partitional Clustering algorithms:

2. MacQueen's *k*-Means algorithms

1. Take the first k objects in the data set as clusters of one member each.

2. Assign each of the remaining (n – k) objects to the cluster with the nearest centroid. Recompute the centroid of the gaining cluster after each assignment.

3. After all objects have been assigned in step 2, take the existing cluster centroids as fixed seed points and make one more pass through the objects assigning each object to the nearest seed point.

Partitional Clustering algorithms:

2. MacQueen's *k*-Means algorithms

Anderberg's convergent version of this method:

1. Begin with an initial partition of the objects into clusters.

2. Take each object in sequence and compute the distances to all cluster centroids; if the nearest centroid is not that of object's parent cluster, then reassign the data unit and update the centroids of the losing and gaining clusters.

3. Repeat step 2 until convergence is achieved ; that is, continue until a full cycle through the objects fails to cause any change in cluster membership.

3. Square-Error Clustering algorithms

Let $y \in C$. Decide to move object y from C_i to C_j. As result of this move, the quantities m_j, J_j, m_i, and J_i will change. Let m_j^*, J_j^*, m_i^*, and J_i^* be the value of these quantities after the move. Then

$$m_{j^*} = m_j + \frac{y - m_j}{n_j + 1} \quad , \quad J_{j^*} = J_j + \frac{n_j}{n_j + 1}\|y - m_j\|^2$$

$$m_i^* = m_i - \frac{y - m_i}{n_i - 1} \quad , \quad J_i^* = J_i - \frac{n_i}{n_i - 1}\|y - m_i\|^2$$

Therefore, the transfer of y from C_i to C_j is welcome only if

$$|J_i^* - J_i| > |J_j^* - J_j|$$

which is same as

$$\frac{n_i}{n_i - 1}\|y - m_i\|^2 > \frac{n_j}{n_j + 1}\|y - m_j\|^2$$

Partitional Clustering algorithms:

3. Square-Error Clustering algorithms

An iterative algorithm using this method can be described as follows.

1. Select an initial partition of the n objects into k clusters and compute m_i and J_e

LOOP:

2. Select a candidate for move $y \in C_i$
3. IF $n_i = 1$, go to NEXT

   ELSE compute

$$R_j = \begin{cases} \dfrac{n_j}{n_j + 1} \left\| y - m_j \right\|^2, \ j \neq i \\ \\ \dfrac{n_i}{n_i - 1} \left\| y - m_i \right\|^2, \ j = i \end{cases}$$

4. Transfer y to C_k if $R_k \leq R_j$ for all j
5. Update m_i, m_k, J_e

NEXT

6. IF J_e has not changed in n steps then STOP

   ELSE go to LOOP.

Application of Cluster Analysis in Meteorology

- Ensemble Forecasting

  is the process of introducing small perturbations to the initial conditions and examining their growth in order to determine the predictability of model forecasts [MITT95]

[MITT95] Jon Mittelstadt, “Introduction to Ensemble Forecasting”, Western Region Technical Attachement No. 95-29, Nov. 21, 1995, Salt Lake City, UT

Application of Cluster Analysis in Meteorology

- Ensemble Forecasting
- An “ensemble”

is a set of model solutions such that each solution, or “member”, is initiated with a slightly different set of initial conditions. The different members are created by introducing small errors, called “perturbations” to the initial conditions of a “control forecast”. Statistically, the ensemble mean should , over time, result in better skill than the individual members [MITT95].

[MITT95] Jon Mittelstadt, “Introduction to Ensemble Forecasting”, Western Region Technical Attachement No. 95-29, Nov. 21, 1995, Salt Lake City, UT

Application of Cluster Analysis in Meteorology

- Objective

A sequence of daily hemispheric weather maps is defined to constitute a persistent or quasi-stationary (QS) events, if the spatial correlation between any pair of maps within the sequence exceeds a given threshold P_o, say $P_o = 0.5$, and if the duration of the event so defined also exceeds a given threshold [MOGHIL88].

[MOGHIL88] K. Mo, M. Ghil, "Cluster Analysis of Multiple Planetary Flow Regimes", Journal o f Geophysical Research, Vol. 93, No. D9, pp 10927-10952, Sep. 20, 1988

Application of Cluster Analysis in Meteorology

- Models and preparation of the Data sets
  - a model that is obtained from extended integrations of a very simple, deterministic, nonlinear mode of NH flow.
  - a set of 500-mbar geopotential height maps for NH winter.

Application of Cluster Analysis in Meteorology

Criteria

1. Membership criterion.

The pattern correlation between the center of a cluster $\bar{c}$ and any element ϕ_j in the cluster should exceed a threshold r_1,

$$p(\bar{c}, \phi_j) = \sum_{v=1}^{v0} \bar{a}_v \bar{c}_v \geq r_1$$

Application of Cluster Analysis in Meteorology

Criteria

1. Membership criterion.

2. Separation criterion.

The pattern correlation between the centers of two clusters, b and c, say, should not exceed a threshold r_2,

$$p(\bar{b}, \bar{c}) = \sum_{v=1}^{v0} \bar{b}_v \bar{c}_v \leq r_2$$

Application of Cluster Analysis in Meteorology

Criteria

1. Membership criterion.

2. Separation criterion.

3. Exclusion criterion.

If a map ϕ does not correlate sufficiently well with the center $\bar{c}_k$ of any cluster,

$$p(\phi, \bar{c}_k) < r_1$$

and it does not satisfy the separation criterion for at least one cluster $\bar{c}_{k0}$, say,

$$p(\phi, \bar{c}_{k0}) > r_2$$

then ϕ belongs to the nonrecurring cluster.

Application of Cluster Analysis in Meteorology

Criteria

1. Membership criterion.

2. Separation criterion.

3. Exclusion criterion.

4. Small-anomaly criterion.

A map $\phi(x,t_n)$ belongs to the small-anomaly cluster, rather than to one of the significant clusters or to the special, nonrecurrent cluster , if its distance to the origin is less than or equals a given threshold d_0

$$d(t_n) = \left\{ \sum_{v=1}^{v0} A_v^2(t_n) \right\}^{\frac{1}{2}} \leq d_0$$

Application of Cluster Analysis in Meteorology

Criteria

1. Membership criterion.

2. Separation criterion.

3. Exclusion criterion.

4. Small-anomaly criterion.

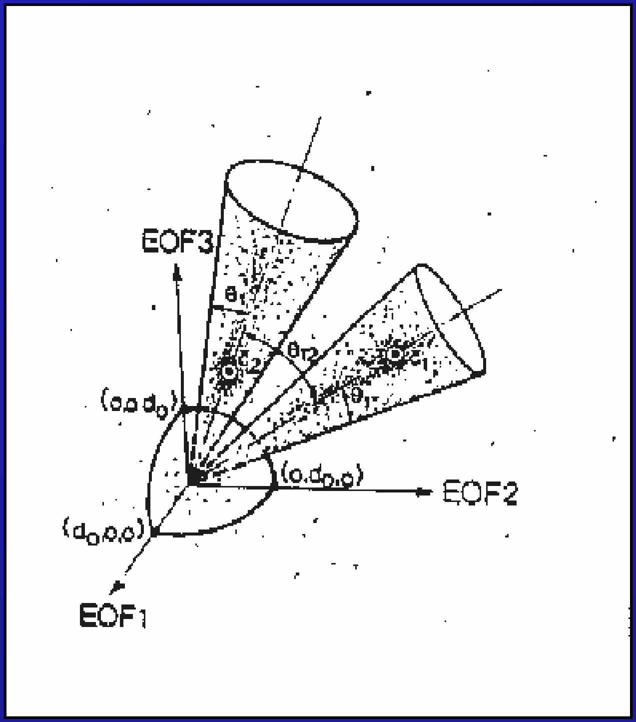
5. Small cluster criterion.

Clusters with less than L_o elements are assigned to the special, nonrecurrent cluster. For model results, L_o = 25 and for NH data, L_o = 8.

Application of Cluster Analysis in Meteorology

Application of Cluster Analysis in Meteorology

Criteria



[MOGHIL88] K. Mo, M. Ghil, "Cluster Analysis of Multiple Planetary Flow Regimes", Journal o f Geophysical Research, Vol. 93, No. D9, pp 10927-10952, Sep. 20, 1988

Application of Cluster Analysis in Meteorology

Algorithm

Seed points

Step A1. Take any map in the time series as point 1.

Step A2. Proceed through the sequence, calculating the correlations $p(\phi, \bar{c}_k)$ between any given map $\phi(x,t)$ and existing centers of cluster $\bar{c}_1, \ldots, \bar{c}_m$

IF $p(\phi, \bar{c}_k) \geq r_1$ THEN

ϕ is assigned to cluster C_k and $\bar{c}_k$ is recomputed.

IF, on the other hand, $p(\phi, \bar{c}_k) \leq r_2$ for all $\bar{c}_k$, k=1,.....,m, THEN

ϕ is allowed to form a new cluster, $\phi = \bar{c}_{m+1}$

IF the exclusion is satisfied, THEN

ϕ is assigned to the special, diffuse cluster.

Step A3. Keep centered fixed and make one pass through the data, assigning points ϕ to existing clusters if $p(\phi, \bar{c}_k) \geq r_1$ for some k, and to the diffuse cluster otherwise.