

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

# Lecture 14: Cluster analysis

Lecturer: Ksenia Kasianova  
[xeniakasianova@gmail.com](mailto:xeniakasianova@gmail.com)

February 26, 2024

# Intro

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

## Plan

- 1) Hierarchical clustering
- 2) Time series clustering
- 3) Partition clustering
- 4) Metrics for assessing clustering quality

# Intro

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

## What is clustering?

- Organizing data into clusters such that there is

- high intra-cluster similarity

- low inter-cluster similarity

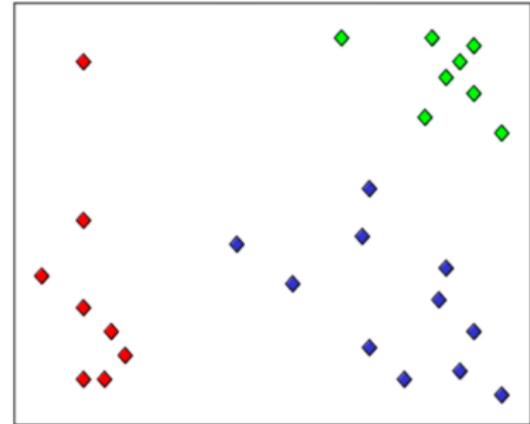
-Informally, finding natural groupings among objects.

Why do we want to do that?

Any REAL application?

- Clustering methods are unsupervised learning techniques

- We do not have a teacher that provides examples with their labels



# Intro

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

Desirable Properties of a Clustering Algorithm:

- Scalability (in terms of both time and space)
- Ability to deal with different data types
- Minimal requirements for domain knowledge to determine input parameters
- Interpretability and usability

Optional:

- Incorporation of user-specified constraints

# Clustering

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

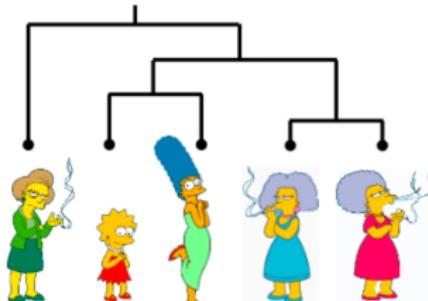
## Two Types of Clustering

- Partitional algorithms: Construct various partitions and then evaluate them by some criterion
- Hierarchical algorithms: Create a hierarchical decomposition of the set of objects using some criterion (focus of this class)

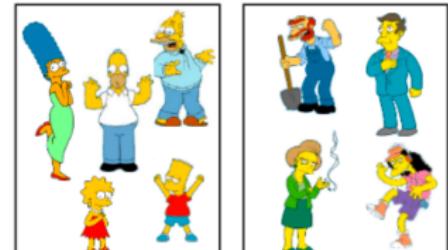
Bottom up or top down

Top down

### Hierarchical



### Partitional



# Clustering

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

## Defining Distance Measures

Definition: Let  $O_1$  and  $O_2$  be two objects from the universe of possible objects. The distance (dissimilarity) between  $O_1$  and  $O_2$  is a real number denoted by  $D(O_1, O_2)$

A few examples:

1) Euclidean distance:

$$d_{\text{euc}}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

2) Manhattan distance:

$$d_{\text{man}}(x, y) = \sum_{i=1}^n |(x_i - y_i)|$$

Correlation-based distances:

3) Pearson correlation distance

- Widely used for gene expression data analyses.

$$d_{\text{cor}}(x, y) = 1 - \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

# Clustering

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

## Data standardization

Goal: to make the variables comparable.

Generally variables are scaled to have i) standard deviation one and ii) mean zero.

The standardization of data is an approach widely used in the context of gene expression data analysis before clustering. We might also want to scale the data when the mean and/or the standard deviation of variables are largely different.

When scaling variables, the data can be transformed as follow:

$$\frac{x_i - \text{center}(x)}{\text{scale}(x)}$$

Where center ( $x$ ) can be the mean or the median of  $x$  values, and scale ( $x$ ) can be the standard deviation (SD), the interquartile range, or the MAD (median absolute deviation).

# Hierarchical clustering

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

The goal of hierarchical clustering is to create a sequence of nested partitions, which can be conveniently visualized via a tree or hierarchy of clusters, also called the cluster dendrogram.

The clusters in the hierarchy range from the fine-grained to the coarse-grained – the lowest level of the tree (the leaves) consists of each point in its own cluster, whereas the highest level (the root) consists of all points in one cluster.

Agglomerative hierarchical clustering methods work in a bottom-up manner. Starting with each of the  $n$  points in a separate cluster, they repeatedly merge the most similar pair of clusters until all points are members of the same cluster.

# Nested Partitions

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

Given a dataset  $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ , where  $\mathbf{x}_i \in \mathbb{R}^d$ , a clustering  $\mathcal{C} = \{C_1, \dots, C_k\}$  is a partition of  $\mathcal{D}$ .

A clustering  $\mathcal{A} = \{A_1, \dots, A_r\}$  is said to be nested in another clustering  $\mathcal{B} = \{B_1, \dots, B_s\}$  if and only if  $r > s$ , and for each cluster  $A_i \in \mathcal{A}$ , there exists a cluster  $B_j \in \mathcal{B}$ , such that  $A_i \subseteq B_j$ .

Hierarchical clustering yields a sequence of  $n$  nested partitions  $\mathcal{C}_1, \dots, \mathcal{C}_n$ . The clustering  $\mathcal{C}_{t-1}$  is nested in the clustering  $\mathcal{C}_t$ . The cluster dendrogram is a rooted binary tree that captures this nesting structure, with edges between cluster  $C_i \in \mathcal{C}_{t-1}$  and cluster  $C_j \in \mathcal{C}_t$  if  $C_i$  is nested in  $C_j$ , that is, if  $C_i \subset C_j$ .

# Dendrogram

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

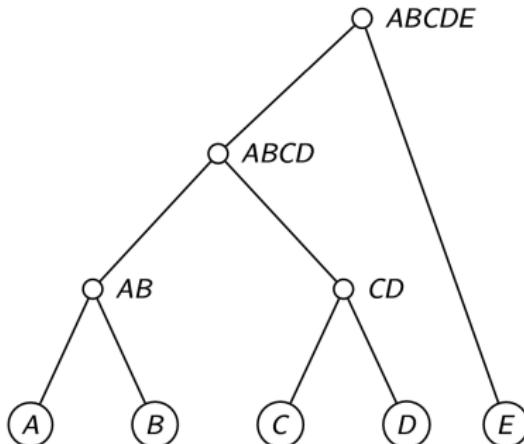
Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal



The dendrogram represents the following sequence of nested partitions:

Clustering	Clusters
$\mathcal{C}_1$	$\{A\}, \{B\}, \{C\}, \{D\}, \{E\}$
$\mathcal{C}_2$	$\{AB\}, \{C\}, \{D\}, \{E\}$
$\mathcal{C}_3$	$\{AB\}, \{CD\}, \{E\}$
$\mathcal{C}_4$	$\{ABCD\}, \{E\}$
$\mathcal{C}_5$	$\{ABCDE\}$

with  $\mathcal{C}_{t-1} \subset \mathcal{C}_t$  for  $t = 2, \dots, 5$ . We assume that  $A$  and  $B$  are merged before  $C$  and  $D$ .

# Number of Hierarchical Clusterings

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

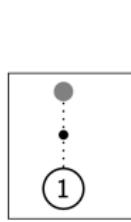
Lance–Williams  
Formula

Dendograms

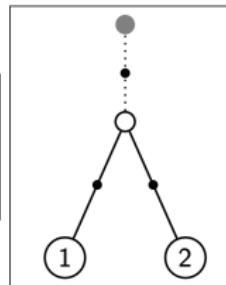
Summary  
of  
Hierarchal

The total number of different dendograms with  $n$  leaves is given as:

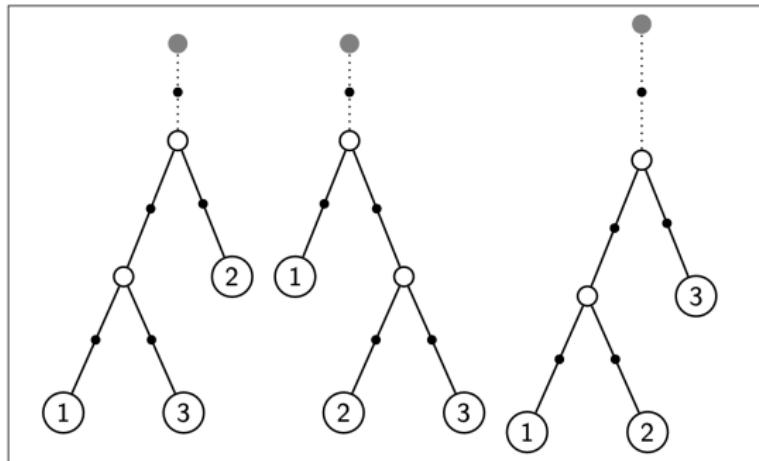
$$\prod_{m=1}^{n-1} (2m - 1) = 1 \times 3 \times 5 \times 7 \times \cdots \times (2n - 3) = (2n - 3)!!$$



(a)  $n = 1$



(b)  $n = 2$



(c)  $n = 3$

# Agglomerative Hierarchical Clustering

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerati-  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

In agglomerative hierarchical clustering, we begin with each of the  $n$  points in a separate cluster. We repeatedly merge the two closest clusters until all points are members of the same cluster.

Given a set of clusters  $\mathcal{C} = \{C_1, C_2, \dots, C_m\}$ , we find the closest pair of clusters  $C_i$  and  $C_j$  and merge them into a new cluster  $C_{ij} = C_i \cup C_j$ .

Next, we update the set of clusters by removing  $C_i$  and  $C_j$  and adding  $C_{ij}$ , as follows  $\mathcal{C} = (\mathcal{C} \setminus \{C_i, C_j\}) \cup \{C_{ij}\}$ .

We repeat the process until  $\mathcal{C}$  contains only one cluster. If specified, we can stop the merging process when there are exactly  $k$  clusters remaining.

# Agglomerative Hierarchical Clustering

Lecture 14

Ksenia  
Kasianova

Intro  
Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerati-  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

AgglomerativeClustering ( $D, k$ ) :

$\mathcal{C} \leftarrow \{C_i = \{x_i\} \mid x_i \in D\} // \text{Each point in separate cluster}$

$\Delta \leftarrow \{\delta(x_i, x_j) : x_i, x_j \in D\} // \text{Compute distance matrix}$

repeat

    Find the closest pair of clusters  $C_i, C_j \in \mathcal{C}$

$C_{ij} \leftarrow C_i \cup C_j // \text{Merge the clusters}$

$\mathcal{C} \leftarrow (\mathcal{C} \setminus \{C_i, C_j\}) \cup \{C_{ij}\} // \text{Update the clustering}$

    Update distance matrix  $\Delta$  to reflect new clustering

until  $|\mathcal{C}| = k$

# Example

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

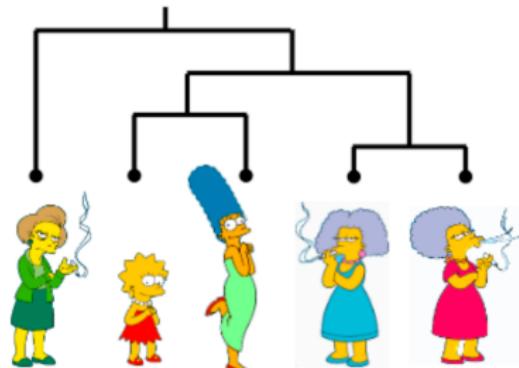
Summary  
of  
Hierarchal

**Bottom-Up (agglomerative):** Starting with each item in its own cluster, find the best pair to merge into a new cluster. Repeat until all clusters are fused together.

The number of dendrograms with  $n$

$$\text{leafs} = (2n - 3)! / [(2^{(n-2)}) (n - 2)!]$$

	Number of Leafs	Number of Possible Dendrograms
	2	1
	3	3
	4	15
	5	105
	...	...
	10	34,459,425



# Example

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

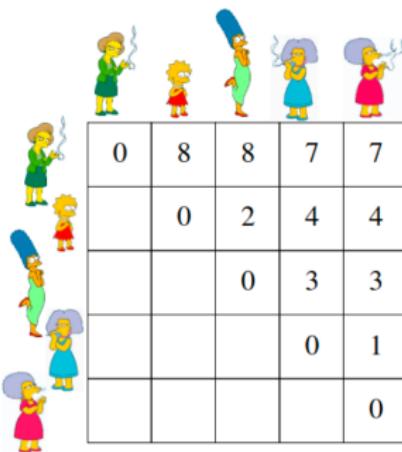
Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

We begin with a distance matrix which contains the distances between every pair of objects in our database.

$$D(\text{Simpson 1, Simpson 2}) = 8$$
$$D(\text{Simpson 1, Simpson 3}) = 1$$



# Example

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

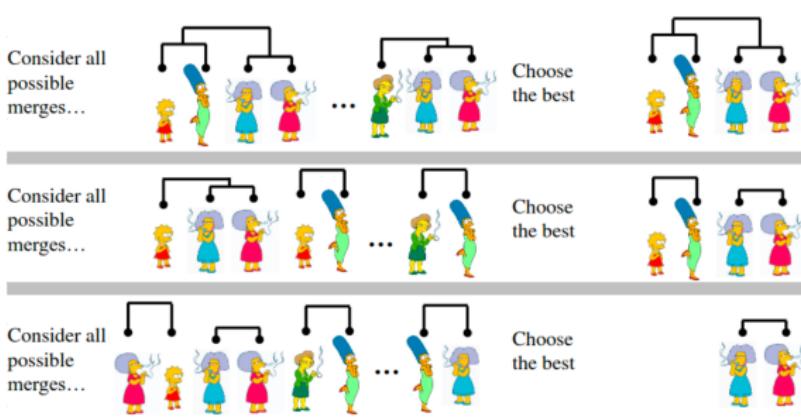
Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchical

Starting with each item in its own cluster, find the best pair to merge into a new cluster. Repeat until all clusters are fused together.



# Example

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

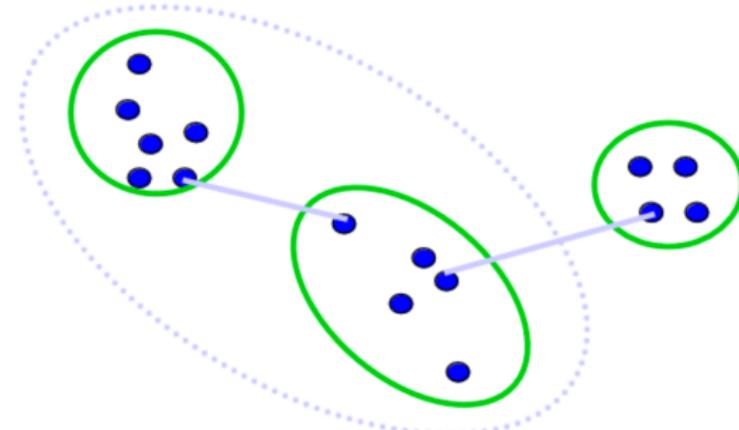
Summary  
of  
Hierarchal

But how do we compute distances between clusters rather than objects?

Computing distance between clusters:

1) Single Link

- cluster distance = distance of two closest members in each class



- Potentially long and skinny clusters

# Example

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

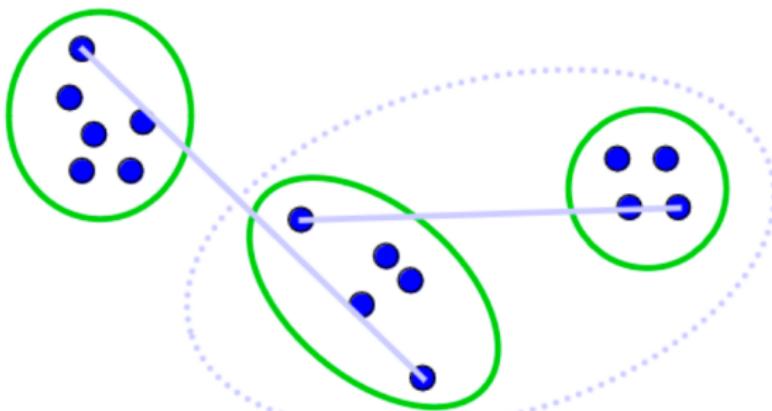
Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

## 2) Complete Link

- cluster distance = distance of two farthest members



- tight clusters

# Example

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

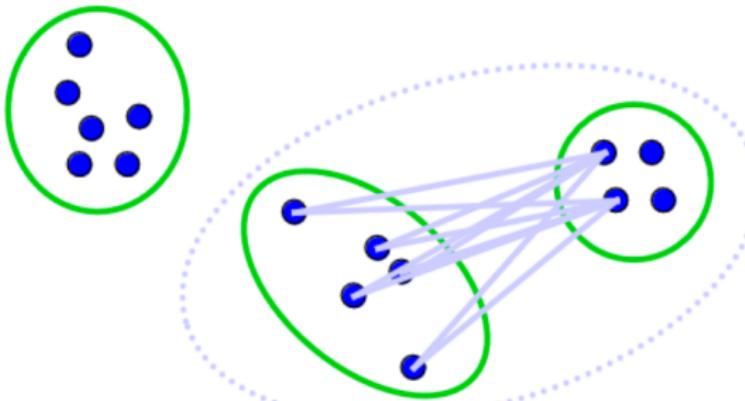
Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

### 3) Average Link

- cluster distance = average distance of all pairs



- the most widely used measure

- robust against noise

# Distance between Clusters

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance-Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

A typical distance between two points is the Euclidean distance or  $L_2$ -norm

$$\|\mathbf{x} - \mathbf{y}\|_2 = \left( \sum_{i=1}^d (x_i - y_i)^2 \right)^{1/2}$$

Single Link: The minimum distance between a point in  $C_i$  and a point in  $C_j$

$$\delta(C_i, C_j) = \min \{ \|\mathbf{x} - \mathbf{y}\| \mid \mathbf{x} \in C_i, \mathbf{y} \in C_j \}$$

Complete Link: The maximum distance between points in the two clusters:

$$\delta(C_i, C_j) = \max \{ \|\mathbf{x} - \mathbf{y}\| \mid \mathbf{x} \in C_i, \mathbf{y} \in C_j \}$$

Group Average: The average pairwise distance between points in  $C_i$  and  $C_j$ :

$$\delta(C_i, C_j) = \frac{\sum_{\mathbf{x} \in C_i} \sum_{\mathbf{y} \in C_j} \|\mathbf{x} - \mathbf{y}\|}{n_i \cdot n_j}$$

# Distance between Clusters

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

Mean Distance: The distance between two clusters is defined as the distance between the means or centroids of the two clusters:

$$\delta(C_i, C_j) = \|\mu_i - \mu_j\|$$

Minimum Variance or Ward's Method: The distance between two clusters is defined as the increase in the sum of squared errors (SSE) when the two clusters are merged, where the SSE for a given cluster  $C_i$  is given as

$$\delta(C_i, C_j) = \Delta SSE_{ij} = SSE_{ij} - SSE_i - SSE_j$$

where  $SSE_i = \sum_{x \in C_i} \|x - \mu_i\|^2$ . After simplification, we get:

$$\delta(C_i, C_j) = \left( \frac{n_i n_j}{n_i + n_j} \right) \|\mu_i - \mu_j\|^2$$

Ward's measure is therefore a weighted version of the mean distance measure.

# Distance between Clusters

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

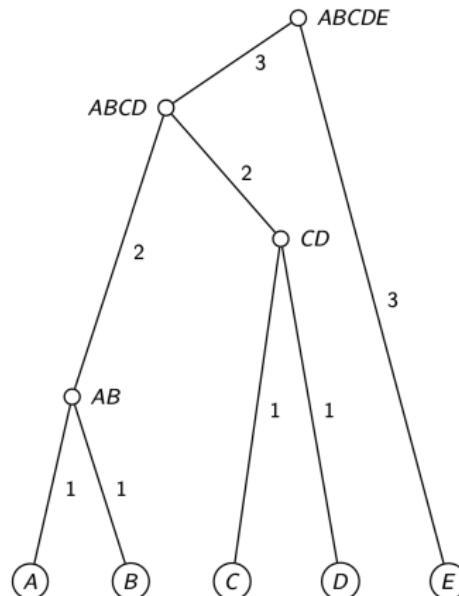
## 1) Single Link Agglomerative Clustering

$\delta$	E
ABCD	(3)

$\delta$	CD	E
AB	(2)	3
CD		3

$\delta$	C	D	E
AB	3	2	3
C		(1)	3
D			5

$\delta$	B	C	D	E
A	(1)	3	2	4
B		3	2	3
C			1	3
D				5



# Distance between Clusters

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchical

$$\begin{matrix} & 1 & 2 & 3 & 4 & 5 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \left[ \begin{matrix} 0 \\ 2 & 0 \\ 6 & 3 & 0 \\ 10 & 9 & 7 & 0 \\ 9 & 8 & 5 & 4 & 0 \end{matrix} \right] \end{matrix}$$



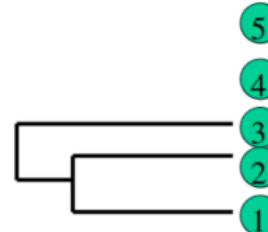
$$\begin{matrix} & (1,2) & 3 & 4 & 5 \\ \begin{matrix} (1,2) \\ 3 \\ 4 \\ 5 \end{matrix} & \left[ \begin{matrix} 0 \\ 3 & 0 \\ 9 & 7 & 0 \\ 8 & 5 & 4 & 0 \end{matrix} \right] \end{matrix}$$



$$\begin{matrix} & (1,2,3) & 4 & 5 \\ \begin{matrix} (1,2,3) \\ 4 \\ 5 \end{matrix} & \left[ \begin{matrix} 0 \\ 7 & 0 \\ 5 & 4 & 0 \end{matrix} \right] \end{matrix}$$

$$d_{(1,2,3),4} = \min\{d_{(1,2),4}, d_{3,4}\} = \min\{9, 7\} = 7$$

$$d_{(1,2,3),5} = \min\{d_{(1,2),5}, d_{3,5}\} = \min\{8, 5\} = 5$$



# Lance–Williams Formula

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Willia-  
Formula

Dendograms

Summary  
of  
Hierarchal

Whenever two clusters  $C_i$  and  $C_j$  are merged into  $C_{ij}$ , we need to update the distance matrix by recomputing the distances from the newly created cluster  $C_{ij}$  to all other clusters  $C_r (r \neq i \text{ and } r \neq j)$ .

The Lance–Williams formula provides a general equation to recompute the distances for all of the cluster proximity measures

$$\begin{aligned}\delta(C_{ij}, C_r) = & \alpha_i \cdot \delta(C_i, C_r) + \alpha_j \cdot \delta(C_j, C_r) + \\ & \beta \cdot \delta(C_i, C_j) + \gamma \cdot |\delta(C_i, C_r) - \delta(C_j, C_r)|\end{aligned}$$

The coefficients  $\alpha_i, \alpha_j, \beta$ , and  $\gamma$  differ from one measure to another.

# Lance–Williams Formula

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Willia-  
Formula

Dendograms

Summary  
of  
Hierarchal

Measure	$\alpha_i$	$\alpha_j$	$\beta$	$\gamma$
Single link	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
Complete link	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
Group average	$\frac{n_i}{n_i+n_j}$	$\frac{n_j}{n_i+n_j}$	0	0
Mean distance	$\frac{n_i}{n_i+n_j}$	$\frac{n_j}{n_i+n_j}$	$\frac{-n_i \cdot n_j}{(n_i+n_j)^2}$	0
Ward's measure	$\frac{n_i+n_r}{n_i+n_j+n_r}$	$\frac{n_j+n_r}{n_i+n_j+n_r}$	$\frac{-n_r}{n_i+n_j+n_r}$	0

# Lance–Williams Formula

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Willia-  
Formula

Dendograms

Summary  
of  
Hierarchal

Single link: Arithmetical trick to find the minimum.

$$\delta(C_{ij}, C_r) = \frac{\delta(C_i, C_r)}{2} + \frac{\delta(C_j, C_r)}{2} - \frac{|\delta(C_i, C_r) - \delta(C_j, C_r)|}{2}$$

Complete link: Arithmetical trick to find the maximum.

$$\delta(C_{ij}, C_r) = \frac{\delta(C_i, C_r)}{2} + \frac{\delta(C_j, C_r)}{2} + \frac{|\delta(C_i, C_r) - \delta(C_j, C_r)|}{2}$$

Group average: Weight the distance by the cluster size.

$$\delta(C_{ij}, C_r) = \frac{n_i}{n_i + n_j} \cdot \delta(C_i, C_r) + \frac{n_j}{n_i + n_j} \cdot \delta(C_j, C_r)$$

# Lance–Williams Formula

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Willia-  
Formula

Dendograms

Summary  
of  
Hierarchal

Mean distance: The new centroid is in the line defined by  $\mu_i$  and  $\mu_j$ , and its distance to  $\mu_r$  has to be adjusted by  $\frac{n_i \cdot n_j}{(n_i + n_j)^2}$ .

$$\delta(C_{ij}, C_r) = \frac{n_i}{n_i + n_j} \cdot \delta(C_i, C_r) + \frac{n_j}{n_i + n_j} \cdot \delta(C_j, C_r) + \frac{-n_i \cdot n_j}{(n_i + n_j)^2} \cdot \delta(C_i, C_j)$$

Ward's measure: The  $\Delta SSE$  of the new cluster is a weighted sum of the  $\Delta SSEs$  of the original clusters, adjusted by the fact that  $n_r$  was considered twice.

$$\delta(C_{ij}, C_r) = \frac{n_i + n_r}{n_i + n_j + n_r} \cdot \delta(C_i, C_r) + \frac{n_j + n_r}{n_i + n_j + n_r} \cdot \delta(C_j, C_r) + \frac{-n_r}{n_i + n_j + n_r} \cdot \delta(C_i, C_j)$$

# Dendrograms

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

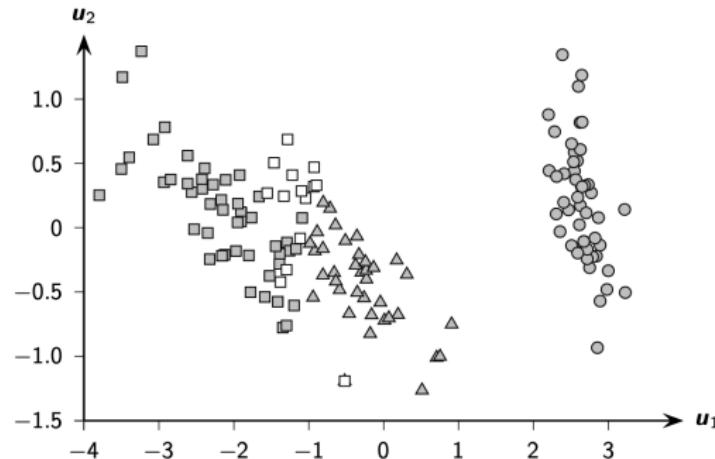
Distance  
between  
Clusters

Lance–Williams  
Formula

Dendrogram

Summary  
of  
Hierarchical

## Iris Dataset: Complete Link Clustering



Contingency Table:

	iris-setosa	iris-virginica	iris-versicolor
$C_1$ (circle)	50	0	0
$C_2$ (triangle)	0	1	36
$C_3$ (square)	0	49	14

# Dendrograms

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

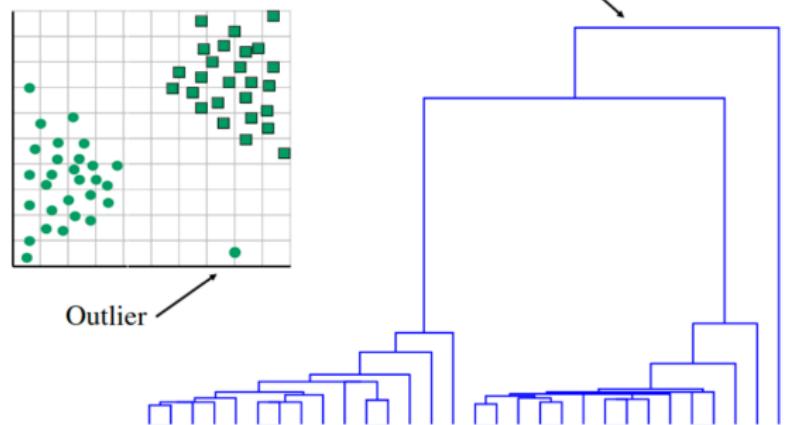
Lance–Williams  
Formula

Dendrogram

Summary  
of  
Hierarchical

One potential use of a dendrogram is to detect outliers

The single isolated branch is suggestive of a data point that is very different to all others



# Summary of Hierarchical Clustering Methods

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

- No need to specify the number of clusters in advance.
- Hierarchical structure maps nicely onto human intuition for some domains
- They do not scale well: time complexity of at least  $O(n^2)$ , where  $n$  is the number of total objects.
- Like any heuristic search algorithms, local optima are a problem.
- Interpretation of results is (very) subjective.

# Time series clustering

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

In this work Dynamic Time Warping (DTW) and Soft-DTW (S-DTW) are used to measure the similarity of time series, because they often used especially for time series.

1) DTW is a measuring method between two time series, which aligns them in time and calculating the minimum distance between corresponding points.

The distance between two time series is computed as the sum of the distances between corresponding points, after aligning them by warping the time axis non-linearly.

$$\text{DTW}(x, y) = \min_p \sqrt{\sum_{(i,j) \in p} d(x_i, y_j)^2},$$

where  $p = [p_0, \dots, p_l]$  is a path that satisfies the following properties:  
- path is a list of index pairs  $p_l = (i_l, j_l)$  with  $0 \leq i_l < a, 0 \leq j_l < b$   
-  $p_0 = (0, 0)$  and  $p_l = (a - 1, b - 1)$   
- for all  $l > 0$ ,  $p_l = (i_l, j_l)$  is related to  $p_{l-1} = (i_{l-1}, j_{l-1})$  as follows:

- $i_{l-1} \leq i_l \leq i_{l-1} + 1$
- $j_{l-1} \leq j_l \leq j_{l-1} + 1$

The warping path is the set of corresponding points that minimize the distance between the two time series.

It's a nonparametric measure which means it doesn't need any assumption on the data distribution.

# Time series clustering

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

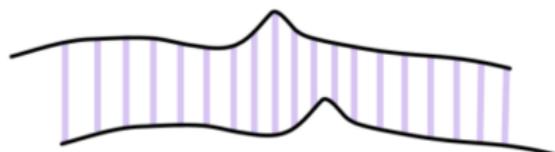
Dendograms

Summary  
of  
Hierarchal

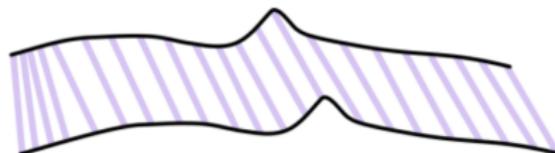
2) S-DTW is a variant of DTW that allows for a "soft" alignment between the time series. It tries to find the best alignment between two time series, while also allowing for small deviations from the optimal alignment.

The method is based on a probability distribution over the warping path and it relaxes the hard constraints of DTW. This means that it is more robust to noise and can handle time series with different lengths more effectively.

**Euclidian distance**



**Dynamic time warping**



# Time series clustering

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

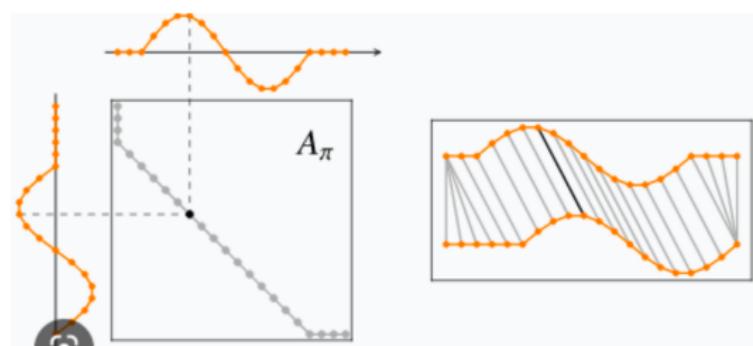
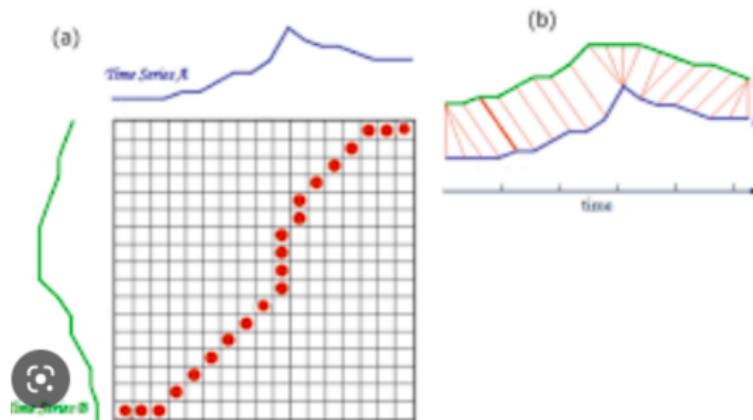
Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal



# Partitional Clustering

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchical

- Nonhierarchical, each instance is placed in exactly one of K non-overlapping clusters.
- Since the output is only one set of clusters the user has to specify the desired number of clusters K.



# Partitional Clustering

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

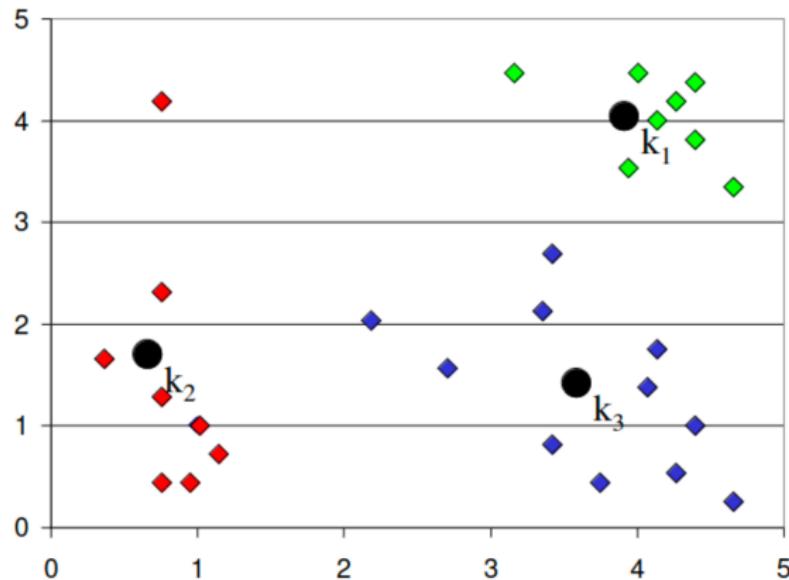
Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

## K-means Clustering:



Re-assign and move centers, until no objects changed membership.

# Partitional Clustering

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

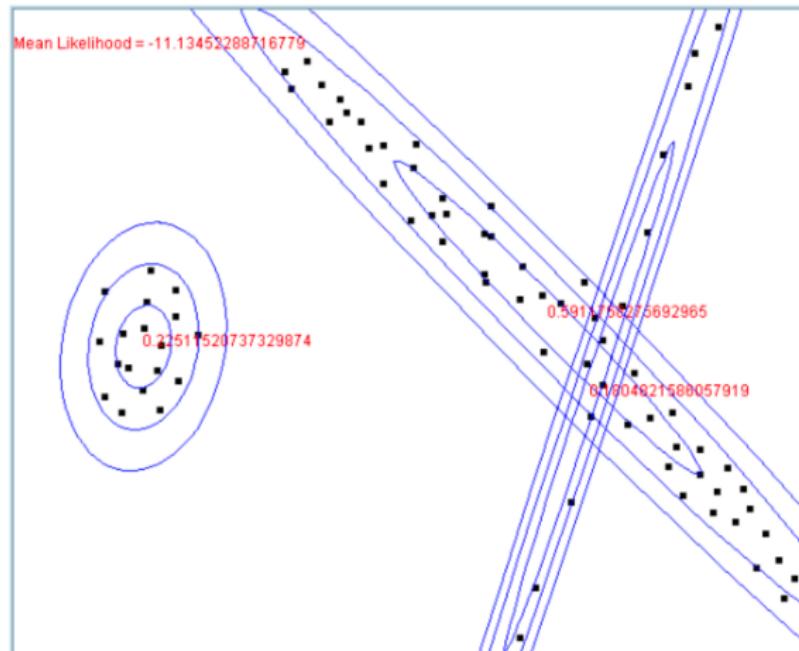
Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchical

## Gaussian mixture clustering



# Clustering methods: Comparison

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

	Hierarchical	K-means	GMM
Running time	naively, $O(N^3)$	fastest (each iteration is linear)	fast (each iteration is linear)
Assumptions	requires a similarity / distance measure	strong assumptions	strongest assumptions
Input parameters	none	$K$ (number of clusters)	$K$ (number of clusters)
Clusters	subjective (only a tree is returned)	exactly $K$ clusters	exactly $K$ clusters

# Clustering quality

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

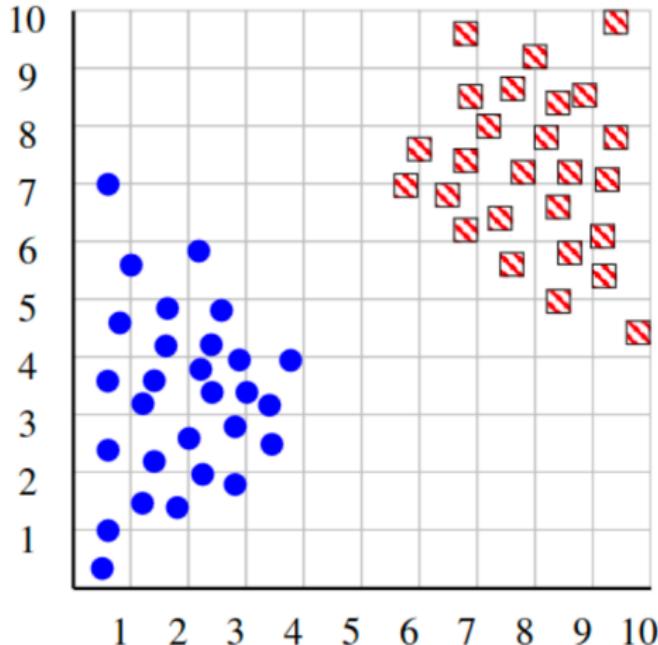
Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

How can we tell the right number of clusters?

In general, this is a unsolved problem. However there are many approximate methods.  
In the next few slides we will see an example.



# Clustering quality

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

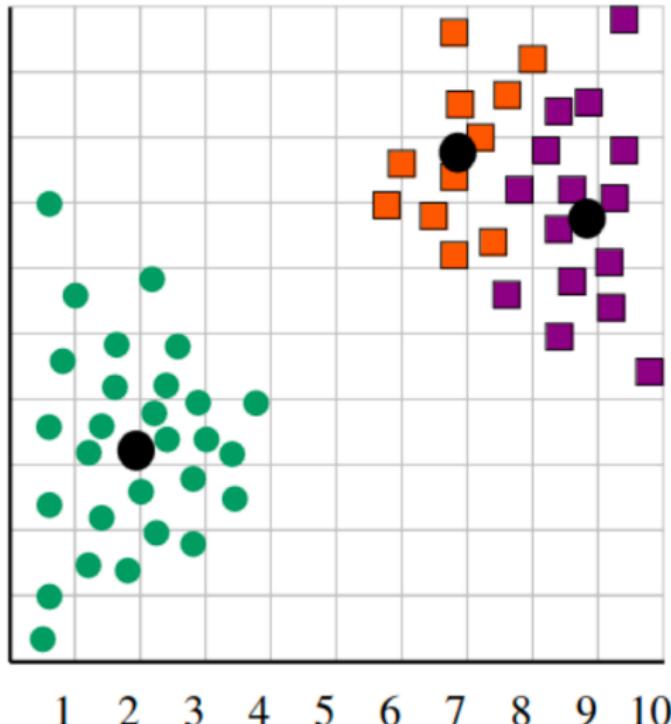
Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

When  $k = 3$ , the objective function is 133.6



# Clustering quality

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

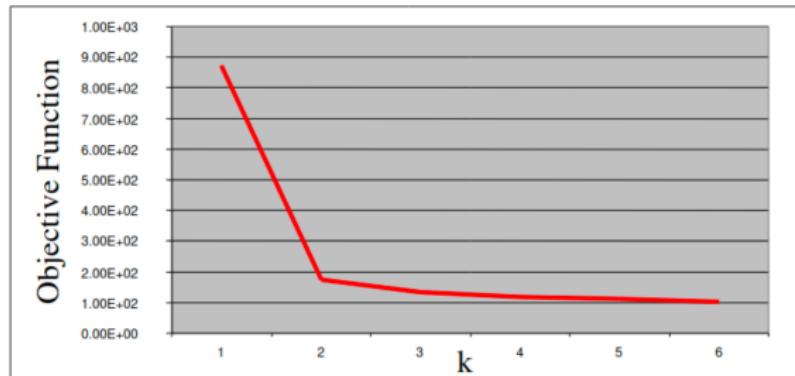
Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

We can plot the objective function values for  $k$  equals 1 to 6...



The abrupt change at  $k = 2$ , is highly suggestive of two clusters in the data. This technique for determining the number of clusters is known as "knee finding" or "elbow finding".

# Clustering quality

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

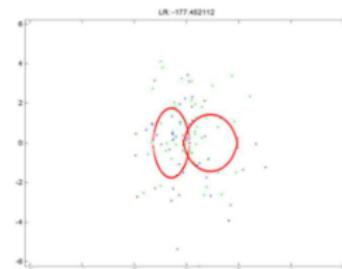
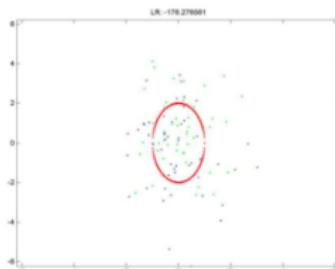
Dendograms

Summary  
of  
Hierarchal

## Cross validation

- We can also use cross validation to determine the correct number of classes
- Recall that GMMs is a generative model. We can compute the likelihood of the left out data to determine which model (number of clusters) is more accurate

$$p(x_1 \cdots x_n | \theta) = \prod_{j=1}^n \left( \sum_{i=1}^k p(x_j | C=i) w_i \right)$$



# Clustering quality

## Lecture 14

### Silhouette score

The Silhouette score is a measure of the similarity of an object to its own cluster compared to other clusters. In this work it is used to evaluate the performance of a clustering algorithm and to choose the number of clusters in a dataset.

$$S = \frac{b - a}{\max(a, b)},$$

where:

- $a(x)$  - mean intra-cluster distance for each sample
- $b(x)$  - mean nearest-cluster distance for each sample

The Silhouette score of the cluster is calculated as follows:

$$S_j = \frac{1}{n} \sum_{i=1}^n s(i)$$

where  $n$  is the number of elements in the cluster  $S_j$ .

So, the final Silhouette score can be calculated in the following way:

$$S = \frac{1}{k} \sum_{j=1}^k S_j$$

where  $k$  is the number of clusters.

The score is limited between -1 for incorrect clustering and +1 for highly dense clustering. Scores around zero indicate overlapping clusters. The score is higher when clusters are dense and well separated, which relates to a standard concept of a cluster.

# Clustering quality

## Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

## Dunn index

Another popular metric is Dunn index. It can be defined as the ratio between the smallest inter-cluster distance and the largest intra-clusters distance. The Dunn index of some set of clusters  $C$  can be computed as:

$$D(C) = \frac{\min_{1 \leq i < j \leq k} \delta(C_i, C_j)}{\max_{1 \leq n \leq k} \Delta_n}$$

where  $k$  is the number of clusters. The larger value of  $D(C)$  the better, so the number of clusters which maximizes this value should be chosen (Dunn, 1973).

# Clustering quality

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

## Calinski-Harabasz index

Calinski-Harabasz (CH) index is somewhat close to the Dunn index, it calculates the ratio of the sum of inter-cluster dispersion mean and of intra-cluster dispersion, where dispersion is defined as the sum of distances square. For a dataset  $E$  of size  $n_E$  with  $k$  clusters, CH can be defined as follows:

$$CH = \frac{\text{tr}(B_k)}{\text{tr}(W_k)} \times \frac{n_E - k}{k - 1}$$

where  $\text{tr}(B_k)$  is a trace of inter-clusters dispersion matrix and  $\text{tr}(W_k)$  is a trace of intrACLusters dispersion matrix:

$$W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q)(x - c_q)^T$$

$$B_k = \sum_{q=1}^k n_q (c_q - c_E)(c_q - c_E)^T$$

Where  $C_q$  is a set of point in cluster  $q$ ,  $c_E$  is a center of  $E$ ,  $n_q$  is a number of samples in cluster  $q$  (Caliński and Harabasz, 1974).

# Summary

Lecture 14

Ksenia  
Kasianova

Intro

Clustering

Hierarchical  
clustering

Nested  
Partitions

Dendrogram

Number of  
Hierarchi-  
cal  
Clusterings

Agglomerative  
Hierarchi-  
cal  
Clustering

Example

Distance  
between  
Clusters

Lance–Williams  
Formula

Dendograms

Summary  
of  
Hierarchal

- 1) Hierarchical clustering vs partition clustering
- 2) Time series clustering
- 3) Choosing the right number of clusters and assessing cluster quality