Lecture 10: Hierarchical Agglomerative Clustering

Introduction to Machine Learning

Sophie Robert

L3 MIASHS — Semestre 2

2022-2023

Introduction

Cluster linkage criteria

Agglomerative clustering

Dendograms and interpretation

Number of cluster selection

Advantages and limits

1 Introduction

2 Cluster linkage criteria

3 Agglomerative clustering

4 Dendograms and interpretation

5 Number of cluster selection

6 Advantages and limits

Definition: hierarchical clustering

Hierarchical clustering

Hierarchical clustering* is a clustering method that seeks to **build a hierarchy of clusters** (clusters within clusters).

Definition: hierarchical clustering

Hierarchical clustering

Hierarchical clustering* is a clustering method that seeks to **build a hierarchy of clusters** (clusters within clusters).

Two approaches:

- **Bottom-up approach (agglomerative*)**: Each observation starts in its own cluster, and **pairs of clusters are merged as one moves up the hierarchy**. We need to decide when to **merge** clusters.

Definition: hierarchical clustering

Hierarchical clustering

Hierarchical clustering* is a clustering method that seeks to **build a hierarchy of clusters** (clusters within clusters).

Two approaches:

- **Bottom-up approach (agglomerative*)**: Each observation starts in its own cluster, and **pairs of clusters are merged as one moves up the hierarchy**. We need to decide when to **merge** clusters.
- **Top-down approach (divisive*)**: All observations start in one cluster, and **splits are performed recursively as one moves down the hierarchy**. We need to decide when to **split** clusters.

Cluster linkage

To decide when to combine (agglomerative) or when to split (divisive), we need:

Cluster linkage

To decide when to combine (agglomerative) or when to split (divisive), we need:

- A measure of dissimilarity between *observations*

Cluster linkage

To decide when to combine (agglomerative) or when to split (divisive), we need:

- A measure of dissimilarity between *observations*
- A measure of dissimilarity between *sets* of observations.

Cluster linkage

Linkage criterion

A **linkage criterion*** is a function of the pairwise distances of observations to measure dissimilarity of sets.

Cluster linkage

Linkage criterion

A **linkage criterion*** is a function of the pairwise distances of observations to measure dissimilarity of sets.

Among the most popular,

1. Ward
2. Complete
3. Average
4. Single

Cluster linkage: Ward

Ward linkage

The **Ward linkage criterion** minimizes the sum of squared differences within all clusters: it measures how much variance is added to the clustering algorithm if the clusters were to be merged.

Cluster linkage: Ward

Ward linkage

The **Ward linkage criterion** minimizes the sum of squared differences within all clusters: it measures how much variance is added to the clustering algorithm if the clusters were to be merged.

$$W(A,B) = \sum_{x \in A \cup B} ||x - \mu_{A \cup B}||^2 - \sum_{x \in A} ||x - \mu_A||^2 - \sum_{x \in B} ||x - \mu_B||^2$$
$$= \frac{|A| \times |B|}{|A \cup B|} ||\mu_A - \mu_B||^2$$

Cluster linkage: Complete

Complete linkage

The **complete linkage criterion** minimizes the maximum distance between observations of pairs of clusters.

Cluster linkage: Complete

Complete linkage

The **complete linkage criterion** minimizes the maximum distance between observations of pairs of clusters.

Given a dissimilarity d and two sets A and B,

$$C(A, B) = max_{A \in A, b \in B} d(a, b)$$

Cluster linkage: Single

Single linkage

The **single linkage criterion** minimizes the distance between the closest observations of pairs of clusters.

Cluster linkage: Single

Single linkage

The **single linkage criterion** minimizes the distance between the closest observations of pairs of clusters.

Given a dissimilarity d and two sets A and B,

$$S(A, B) = min_{a \in A, b \in B} d(a, b)$$

Cluster linkage: Average

Average linkage

The **average linkage criterion** minimizes the average of the distances between all observations of pairs of clusters.

Cluster linkage: Average

Average linkage

The **average linkage criterion** minimizes the average of the distances between all observations of pairs of clusters.

Given a dissimilarity d and two sets A and B,

$$A(A, B) = \frac{1}{|A| \times |B|} \sum_{a \in A} \sum_{b \in B} d(a, b)$$

Cluster linkage

Question

Compute the linkage criterion of *ward*, *complete*, *average*, *single* between the following sets of $\mathbb{R}^2$ vectors using *Manhattan distance*:

- $A = \{[1, 2], [2, 3], [4, 5]\}$
- $B = \{[3, 1], [4, 5], [1, 5]\}$

Algorithm for agglomerative clustering

For a selected linkage criterion L and dissimilarity d,

Algorithm for agglomerative clustering

For a selected linkage criterion L and dissimilarity d,

1. Initially put all individuals in their own cluster

Algorithm for agglomerative clustering

For a selected linkage criterion L and dissimilarity d,

1. Initially put all individuals in their own cluster
2. Merge together closest elements

Algorithm for agglomerative clustering

For a selected linkage criterion L and dissimilarity d,

1. Initially put all individuals in their own cluster
2. Merge together closest elements
3. Until every elements are in the same cluster: Recursively merge sets of elements that minimize the linkage criterion

Algorithm for agglomerative clustering

For a selected linkage criterion L and dissimilarity d,

1. Initially put all individuals in their own cluster
2. Merge together closest elements
3. Until every elements are in the same cluster: Recursively merge sets of elements that minimize the linkage criterion

Clusters are assigned by cutting the algorithm at a selected value for the Linkage criterion.

Example

Question

Using single linkage criterion and manhattan distance, perform Hierarchical Agglomerative Clustering on the following dataset. Test different cut-off values for final clusters.

Sepal length	Sepal width	Petal length
5	2	1
4	3	1
6	5	4
7	1	5

Lecture 10: Hierarchical Agglomerative Clustering

Sophie Robert

Introduction

Cluster linkage criteria

Agglomerative clustering

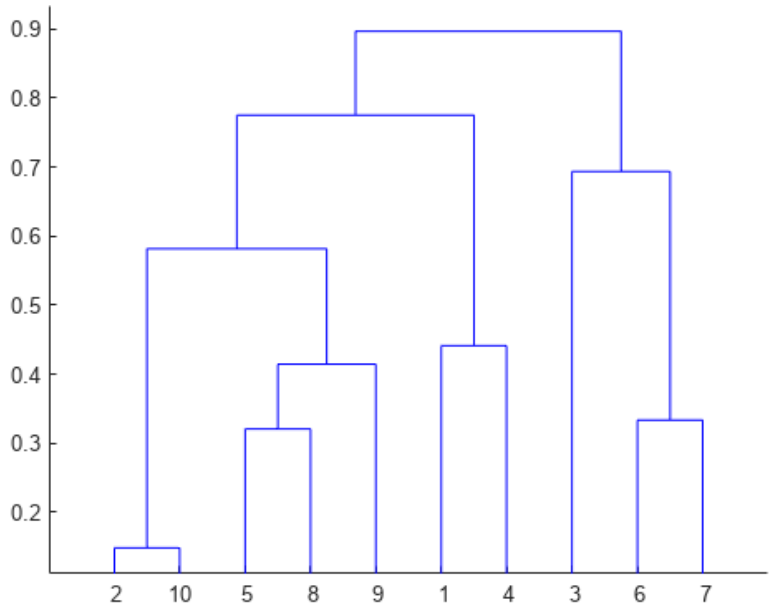
Dendograms and interpretation

Number of cluster selection

Advantages and limits

Dendrograms

A **dendrogram*** is a diagram representing a tree. In the case of hierarchical clustering, it consists in representing graphically the clusters and their linkage criterion.



Interpretation of results

The dendrogram is used to infer information regarding the clustering process:

Interpretation of results

The dendrogram is used to infer information regarding the clustering process:

- Easily visualize elements grouped together

Interpretation of results

The dendrogram is used to infer information regarding the clustering process:

- Easily visualize elements grouped together
- Easily select the number of clusters by setting the cut-off

Interpretation of results

Lecture 10: Hierarchical Agglomerative Clustering

Sophie Robert

Introduction

Cluster linkage criteria

Agglomerative clustering

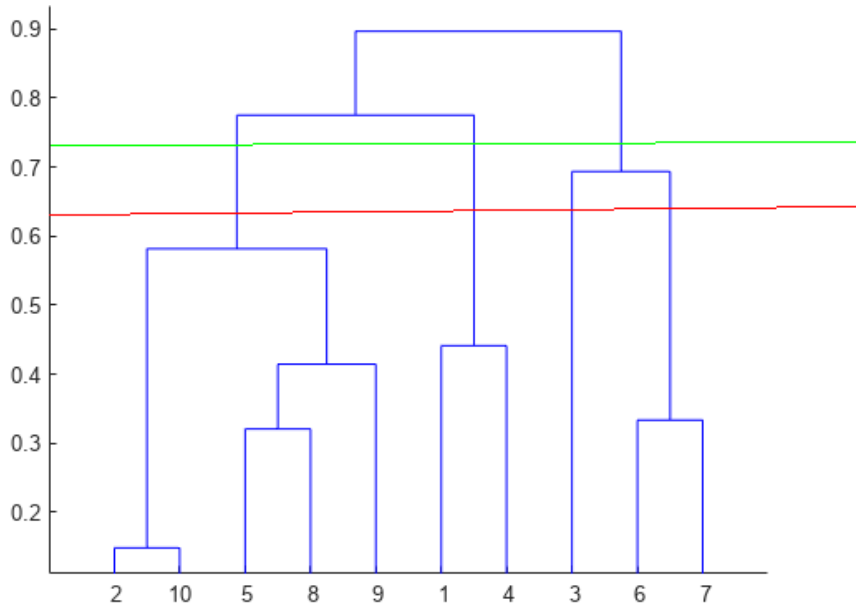
Dendograms and interpretation

Number of cluster selection

Advantages and limits

The dendrogram is used to infer information regarding the clustering process:

- Easily visualize elements grouped together
- Easily select the number of clusters by setting the cut-off



Selecting the right number of clusters

Selecting the right number of clusters is again a problem...

Selecting the right number of clusters

Selecting the right number of clusters is again a problem...

- The dendogram can show splits that "make sense"
- Plot clustering quality scores *vs* the number of clusters

Advantages and limits

Lecture 10: Hierarchical Agglomerative Clustering

Sophie Robert

Introduction

Cluster linkage criteria

Agglomerative clustering

Dendograms and interpretation

Number of cluster selection

Advantages and limits

Lecture 10: Hierarchical Agglomerative Clustering

Sophie Robert

Introduction

Cluster linkage criteria

Agglomerative clustering

Dendograms and interpretation

Number of cluster selection

Advantages and limits

Advantages

Advantages and limits

Advantages

- Easy to interprete even for people outside our field
- No need to perform the computations again when changing the number of clusters

Advantages and limits

Advantages

- Easy to interprete even for people outside our field
- No need to perform the computations again when changing the number of clusters

Limits

Advantages and limits

Advantages

- Easy to interprete even for people outside our field
- No need to perform the computations again when changing the number of clusters

Limits

- Number of clusters has to be inferred

Questions

Lecture 10: Hierarchical Agglomerative Clustering

Sophie Robert

Introduction

Cluster linkage criteria

Agglomerative clustering

Dendograms and interpretation

Number of cluster selection

Advantages and limits

Questions ?