

# Chapter VII: Cluster analysis

## Knowledge Discovery in Databases

Luciano Melodia M.A.

Evolutionary Data Management, Friedrich-Alexander University Erlangen-Nürnberg

Summer semester 2021



## Chapter VII: Cluster analysis

### **Cluster analysis: basic concepts.**

Partitioning methods.

Hierarchical methods.

Density-based methods.

Grid-based methods.

Evaluation of clustering.

Summary.

## What is cluster analysis?

**Cluster:** A collection of data objects within a larger set that are.

Similar (or related) to one another within the same group and,  
dissimilar (or unrelated) to the objects outside the group.

**Cluster analysis (or clustering, data segmentation, . . .).**

Define similarities among data based on the characteristics found in the data (input from user!).  
Group similar data objects into clusters.

**Unsupervised learning:**

No predefined classes.

I.e., learning by observation (vs. learning by examples: supervised).

**Typical applications:**

As a stand-alone tool to get insight into data distribution.

As a preprocessing step for other algorithms.

## Clustering for data understanding and applications

**Biology:**

**Information retrieval:**

**Land use:**

**Marketing:**

**City planning:**

**Earthquake studies:**

**Climate:**

**Economic Science:**

Thank you for your attention.  
**Any questions about the seventh chapter?**

Ask them now, or again, drop me a line:  
✉ [luciano.melodia@fau.de](mailto:luciano.melodia@fau.de).