

Chapter II: Data

Knowledge Discovery in Databases

Luciano Melodia M.A.

Evolutionary Data Management, Friedrich-Alexander University Erlangen-Nürnberg

Summer semester 2021



Chapter II: Getting to know your data

This is our agenda for this lecture:

Data objects and attribute types

Basic statistical descriptions of data

Data visualization

Measuring data similarity and dissimilarity

Summary

Types of data sets

Records:

Relational records.

Data matrix, e.g. numerical matrix, crosstabs.

Document data: text documents, **term-trequence vectors**.

Transaction data.

	Item1	Item2	Item3
Group1	0.8	0.1	0.1
Group2	0.1	0.8	0.1
Group3	0.1	0.1	0.8
Group4	0.34	0.33	0.33

Graph and network:

World wide web.

Social of information networks.

Molecular structures.

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diapers, Milk
4	Beer, Bread, Diapers, Milk
5	Coke, Diapers, Milk

Types of data sets

Ordered data:

- Video data: sequences of images.
- Temporal data: time series.
- Sequential data: transaction sequences.
- Genetic sequence data.

Spatial, image and multimedia:

- Spatial data: maps.
- Image data.
- Video data.

Thank you for your attention.

Any questions about the second chapter?

Ask them now, or again, drop me a line:

✉ `luciano.melodia@fau.de`.