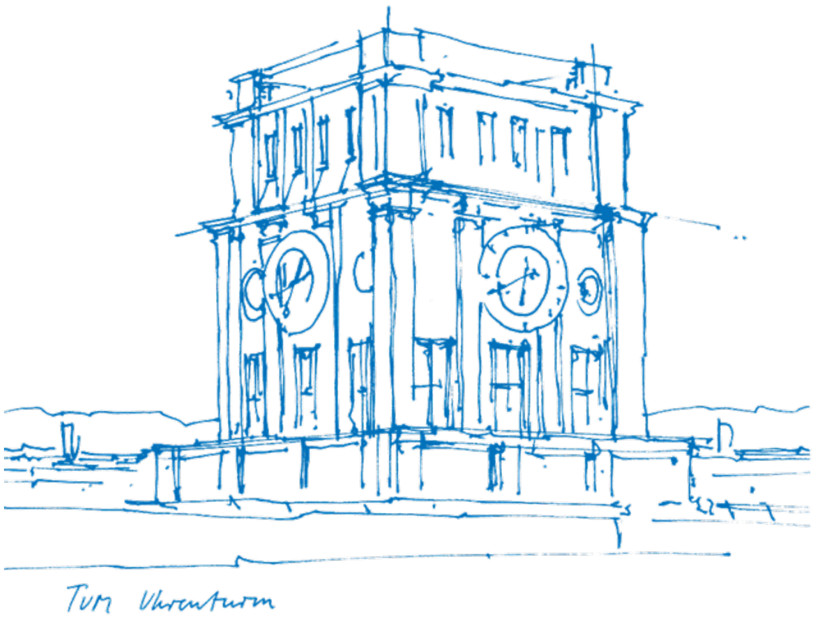
MLCMS, Lecture 3: Representation of data

Felix Dietrich

2022-12-08: Organizational issues



Organizational issues

Groups, Moodle, Reports

- Did every group upload their reports for the last exercise (deadline today, 23:59)?
- You will receive the points for the last exercise until the QnA session.
- Let me know if you have questions about my feedback.
- **Important:** there is a poll on Moodle to catch issues with the course early on. Be sure to fill it out!

Recap / Outlook

Lecture 1: Modeling crowd dynamics

- Modeling approaches, verification and validation

Lecture 2: Simulation software

- Introduction to the Vadere software, SIR models

Lecture 3: Representation of data (today!)

- Principal Component Analysis, Diffusion Maps, neural networks

Lecture 4: Dynamical systems and bifurcation theory

- Introduction to the theory and examples

Lecture 5: Extracting dynamical systems from data

- Function approximation, vector fields, time-delay embedding

Lecture 6: Future directions of machine learning

- Challenges in data science, master’s thesis topics, final projects

Today: Machine Learning!

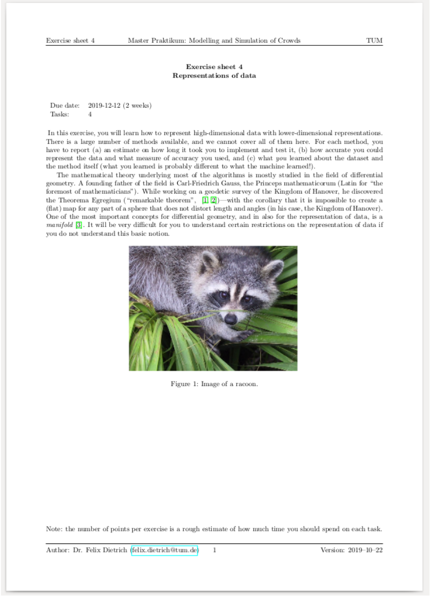
Representation of data

1. Video 2: Linear manifold learning: Principal component analysis
2. Video 3: Non-linear manifold learning: Diffusion Maps
3. Video 4: Non-linear distribution learning: Variational auto-encoders
4. Exercise sheet

Representation of data

Exercise sheet

You can find the exercise sheet on Moodle



Exercise sheet 4 Master Praktikum: Modelling and Simulation of Crowds TUM

Exercise sheet 4
Representations of data

Due date: 2019-12-12 (2 weeks)
Tasks: 4

In this exercise, you will learn how to represent high-dimensional data with lower-dimensional representations. There is a large number of methods available, and we cannot cover all of them here. For each method, you have to report (a) an estimate on how long it took you to implement and test it, (b) how accurate you could represent the data and what measure of accuracy you used, and (c) what *you* learned about the dataset and the method itself (what you learned is probably different to what the machine learned!).

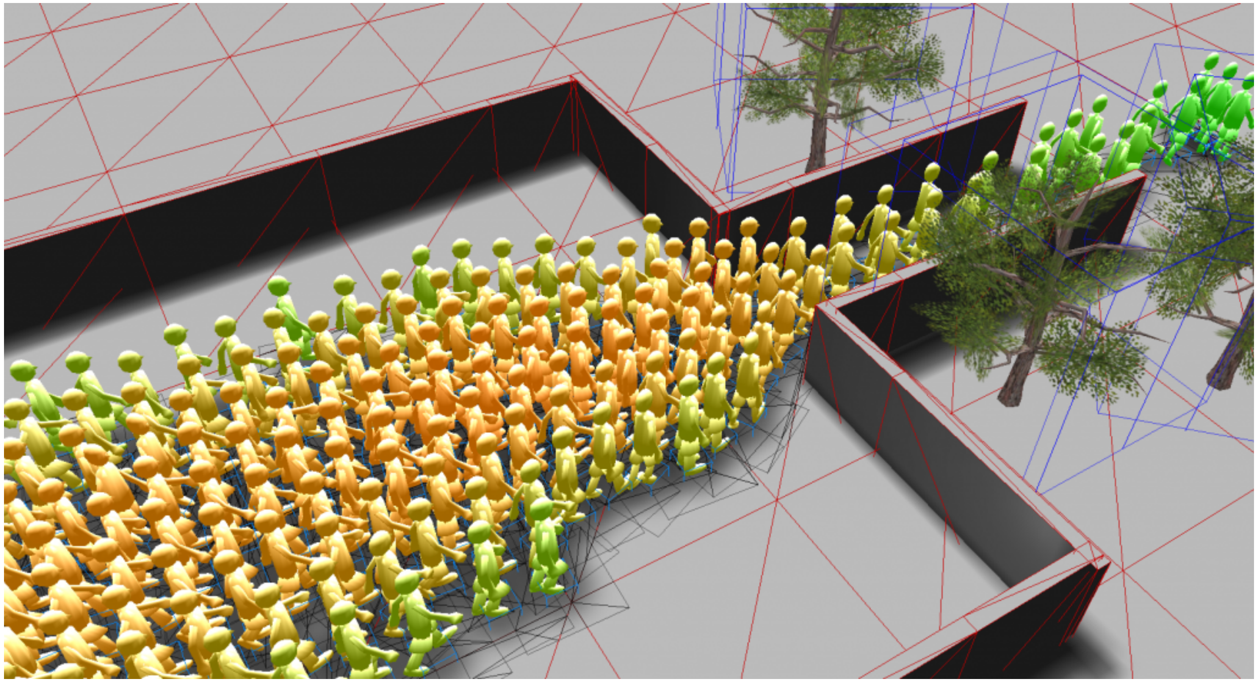
The mathematical theory underlying most of the algorithms is mostly studied in the field of differential geometry. A founding father of the field is Carl-Friedrich Gauss, the Princeps mathematicorum (Latin for "the foremost of mathematicians"). While working on a geodetic survey of the Kingdom of Hanover, he discovered the Theorema Egregium ("remarkable theorem", [1, 2])—with the corollary that it is impossible to create a (flat) map for any part of a sphere that does not distort length and angles (in his case, the Kingdom of Hanover). One of the most important concepts for differential geometry, and in also for the representation of data, is a *manifold* [3]. It will be very difficult for you to understand certain restrictions on the representation of data if you do not understand this basic notion.

Figure 1: Image of a racoon.

Note: the number of points per exercise is a rough estimate of how much time you should spend on each task.

Author: Dr. Felix Dietrich (felix.dietrich@tum.de) 1 Version: 2019–10–22

Questions?



Homework: finish the exercise & upload report until 2022-12-08.
Homework: Fill out the poll on Moodle
For questions / appointments: please ask via email or chat, `felix.dietrich@tum.de`.