Machine Learning (IN2064)

Lecture 1 - Introduction

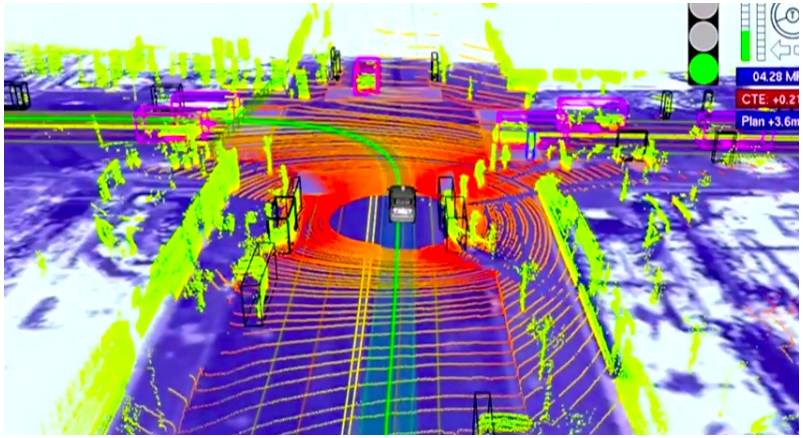
Prof. Dr. Stephan Günnemann

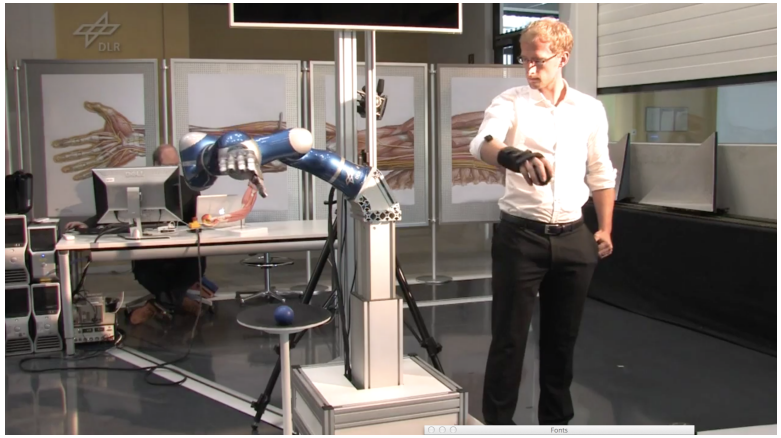
Data Analytics and Machine Learning
Technical University of Munich

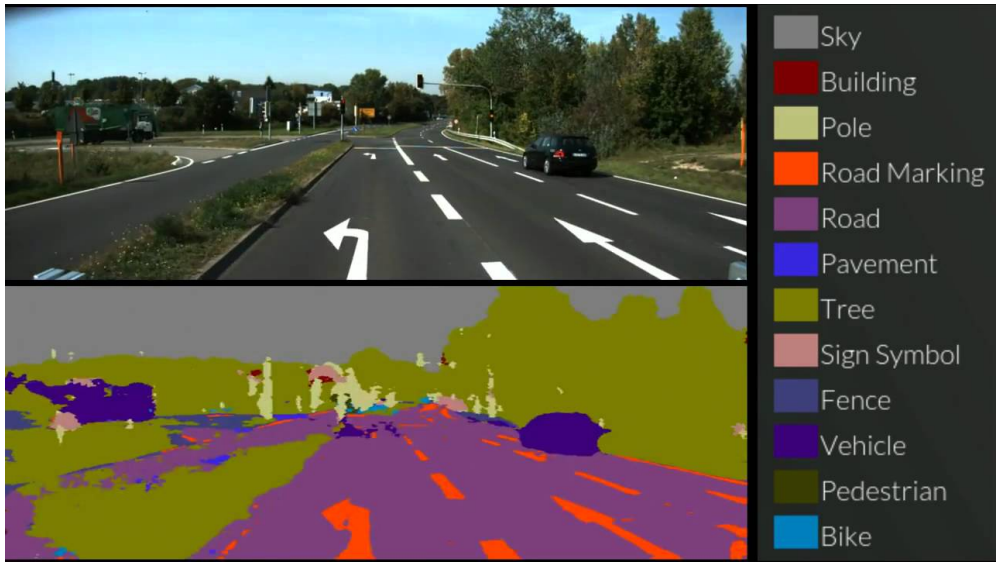
www.daml.in.tum.de

Winter term 2020/2021

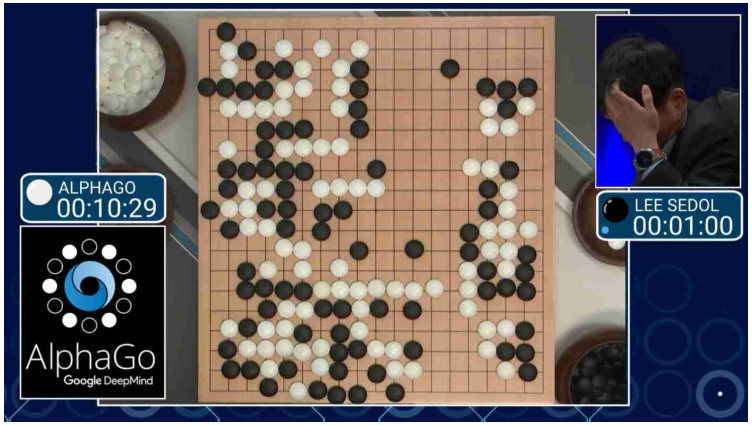
Self-driving cars and robotics

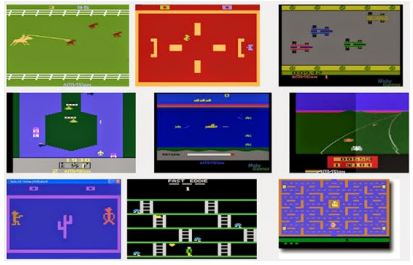


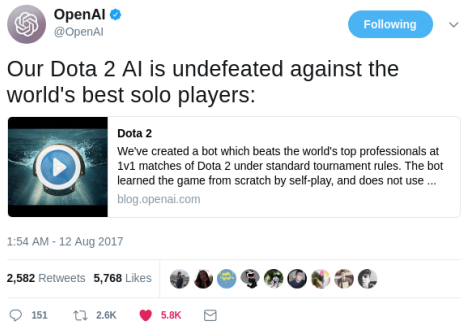


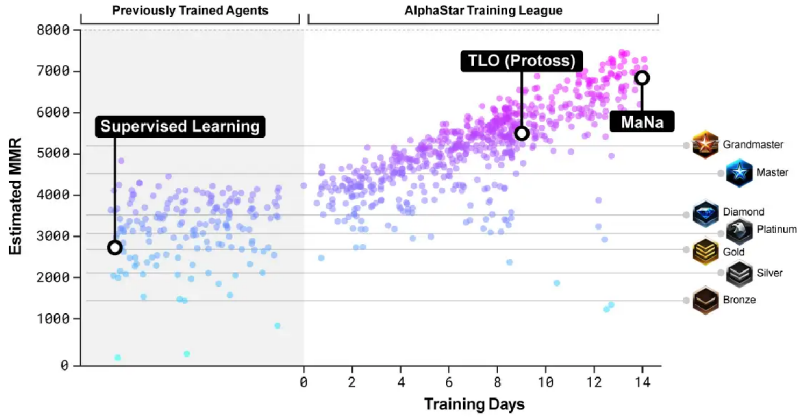


Game playing

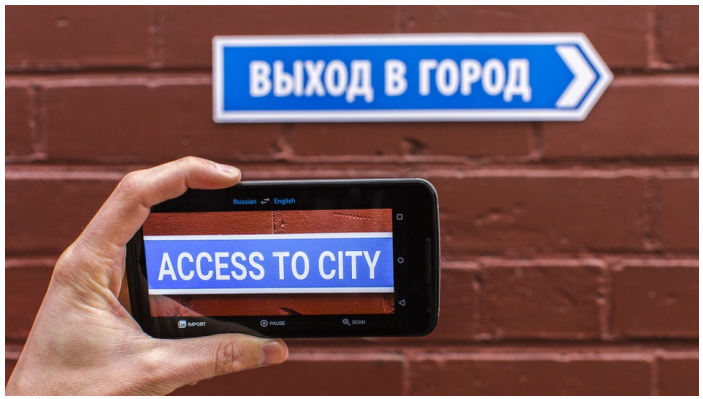


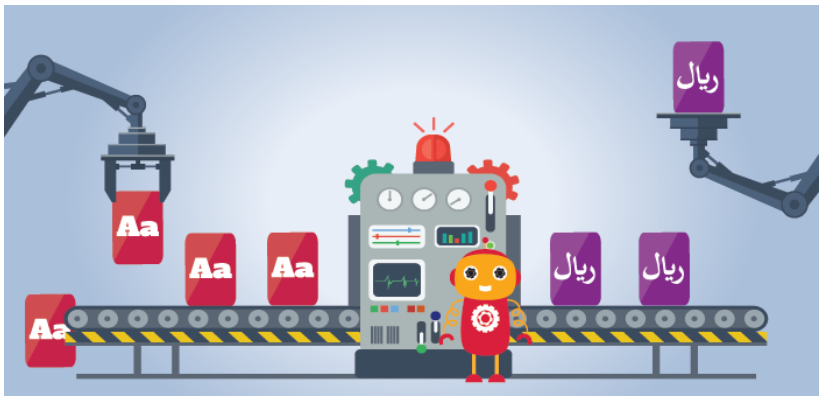


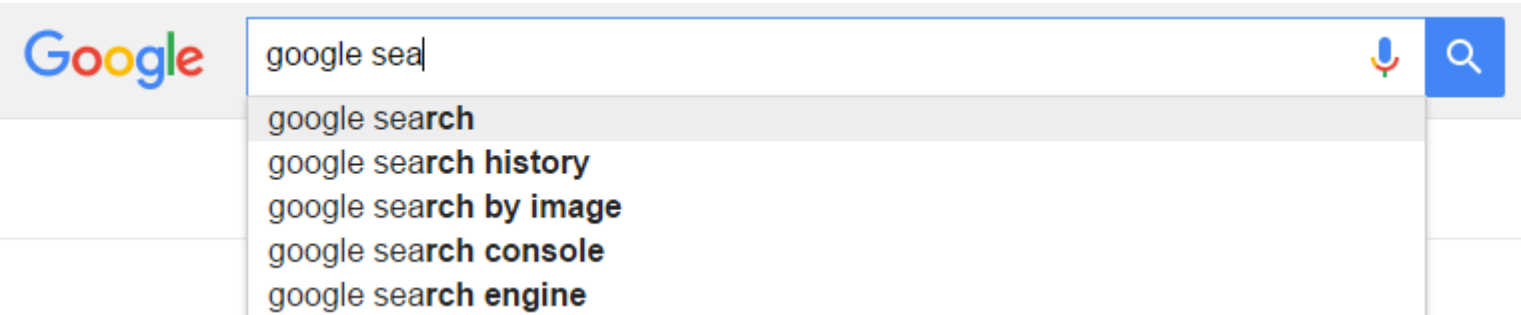




Natural language processing

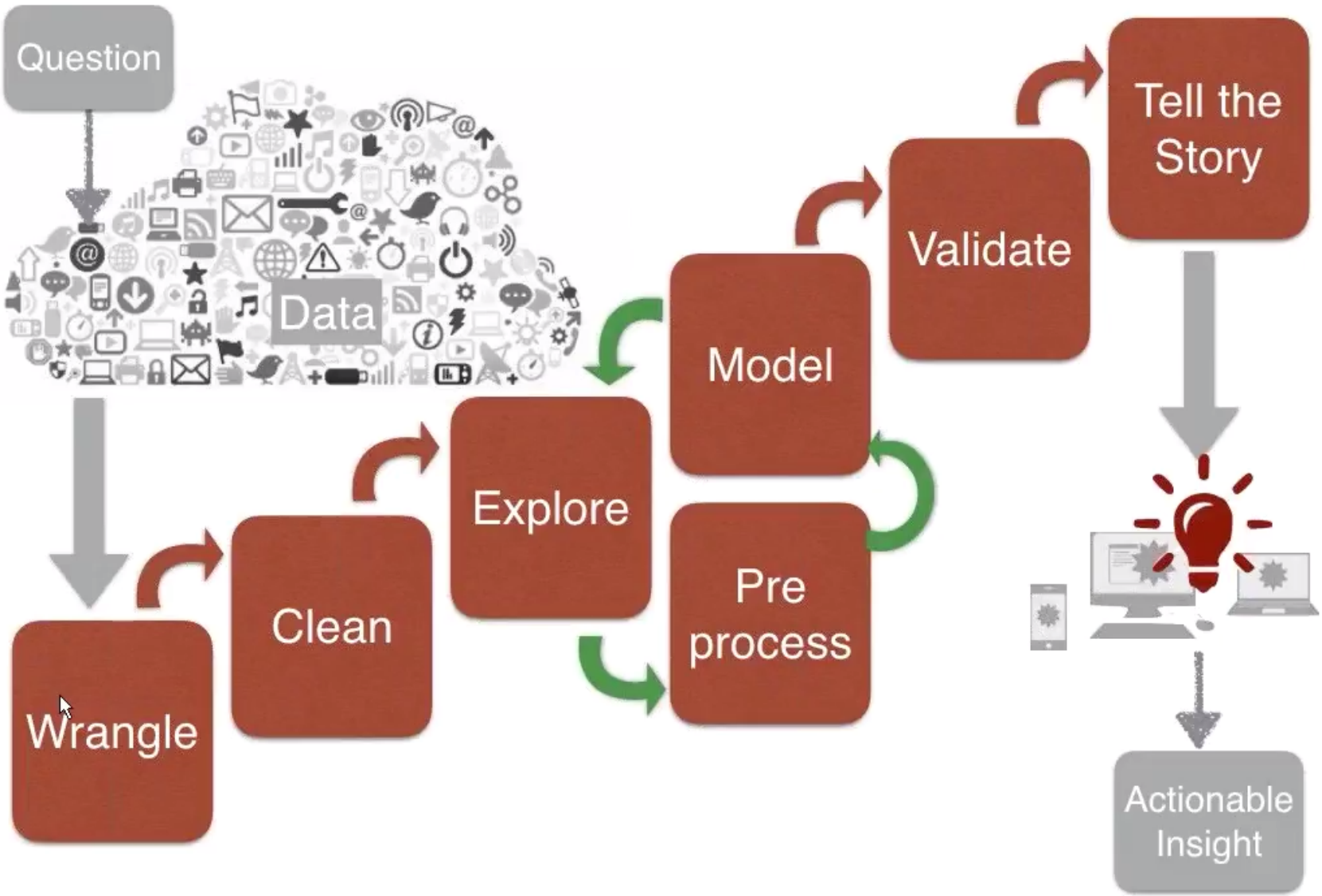




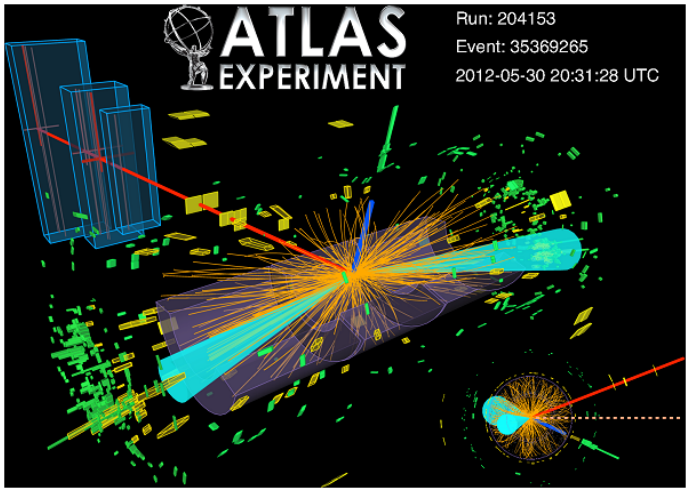


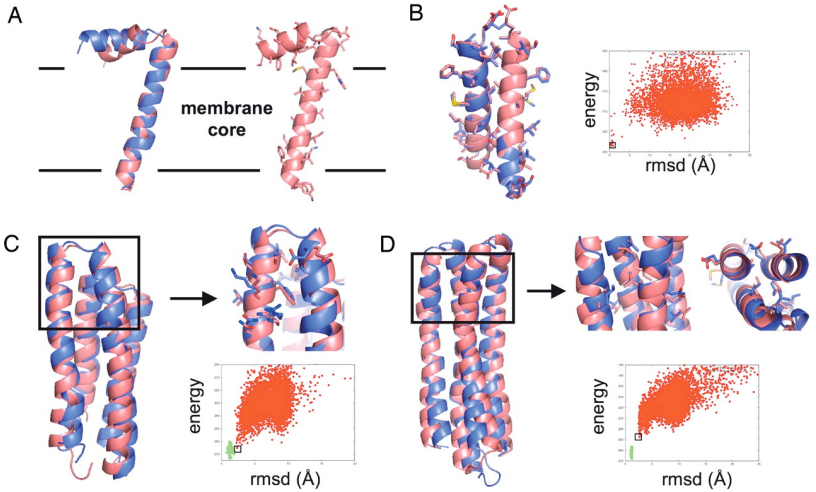
Web, ads and recommendations

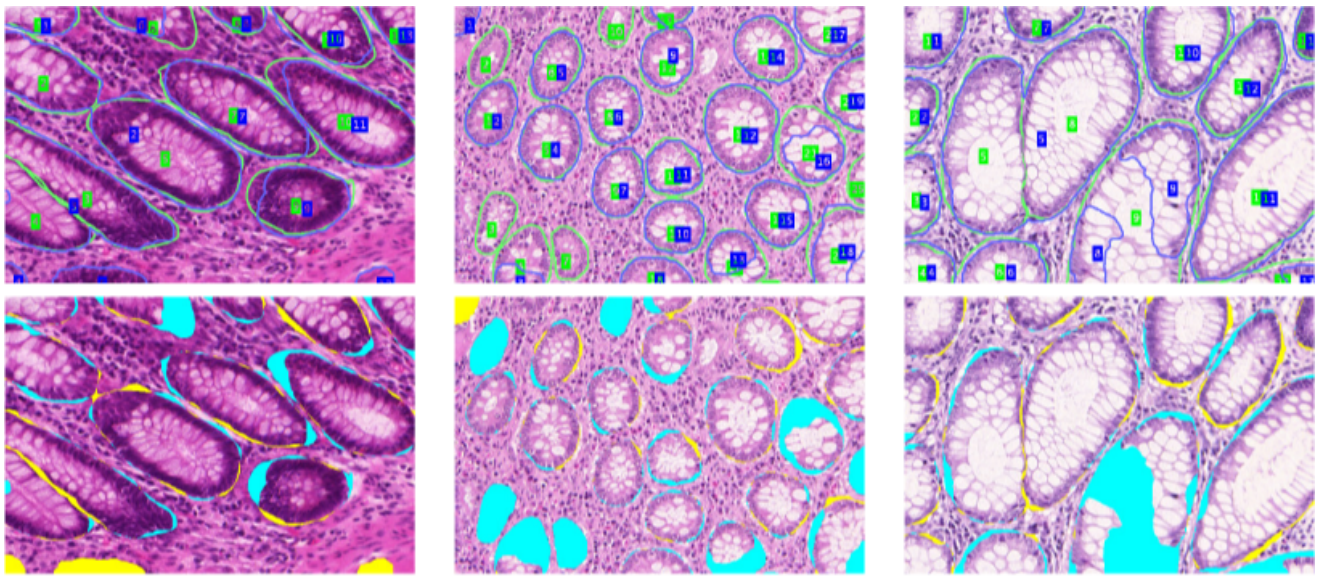
Data science



Physics, biology and medicine







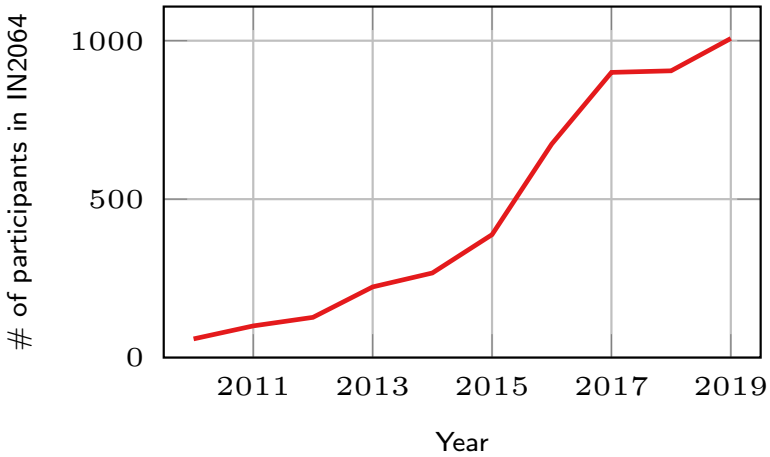
(d) benign (e) benign (f) malignant

What unites all these technologies?

- Computer vision
- Natural language processing
- Recommender systems
- Computational advertising
- Robotics
- Artificial intelligence
- Data science
- Bioinformatics
- Many other fields

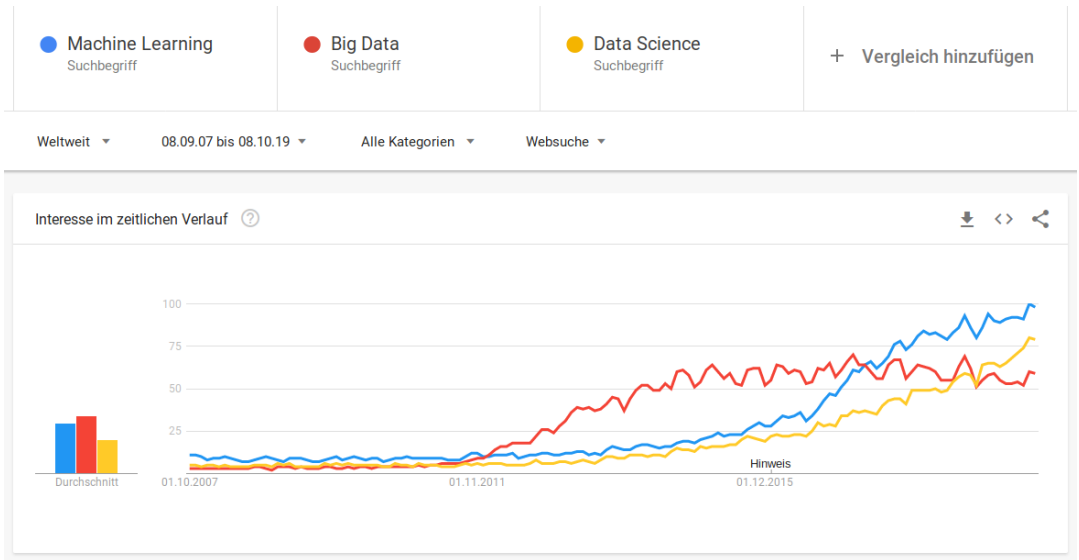
} All are using Machine learning

Hot topic


# of participants in IN2064
1000
500
0
2011
2013
2015
2017
2019
Year


Science
AAAS
SPECIAL ISSUE
ARTIFICIAL INTELLIGENCE


nature
At last — a computer program that can beat a champion Go player
ALL SYSTEMS GO

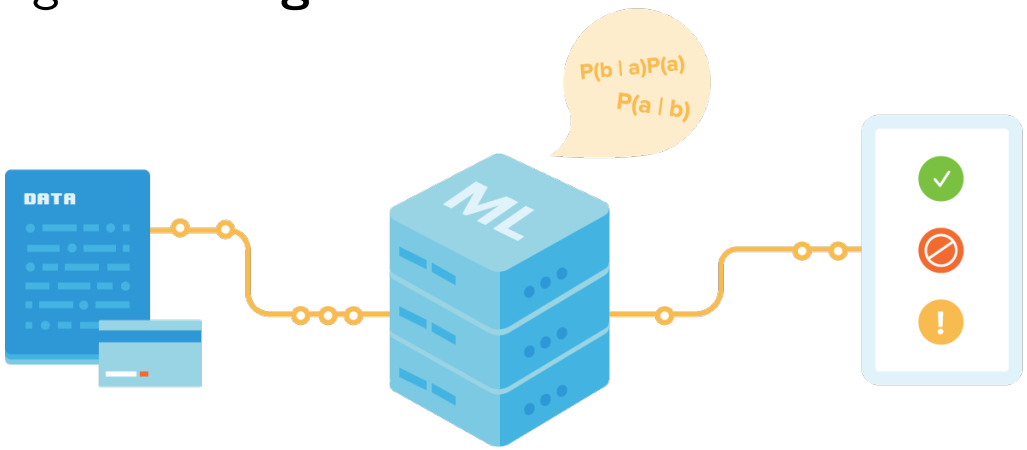

Machine Learning
Suchbegriff
Big Data
Suchbegriff
Data Science
Suchbegriff
Vergleich hinzufügen
Weltweit
08.09.07 bis 08.10.19
Alle Kategorien
Websuche
Interesse im zeitlichen Verlauf
100
75
50
25
Hinweis
Durchschnitt
01.10.2007
01.11.2011
01.12.2015

What is Machine Learning?

Simple example - classify transactions into legitimate and fraudulent.

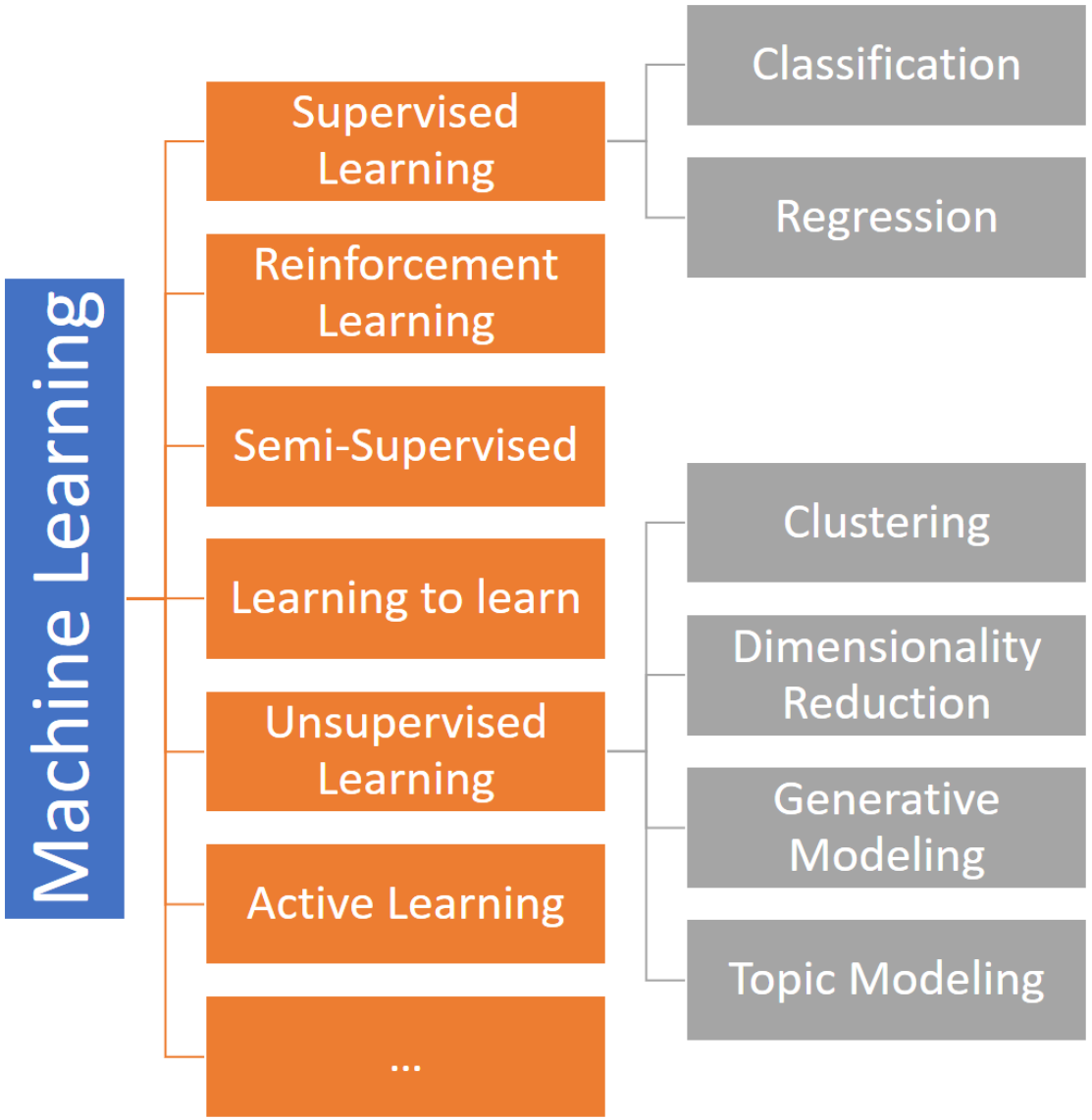
Rule-based approaches - **rules** handcrafted by human experts

Machine learning - **learning from data**



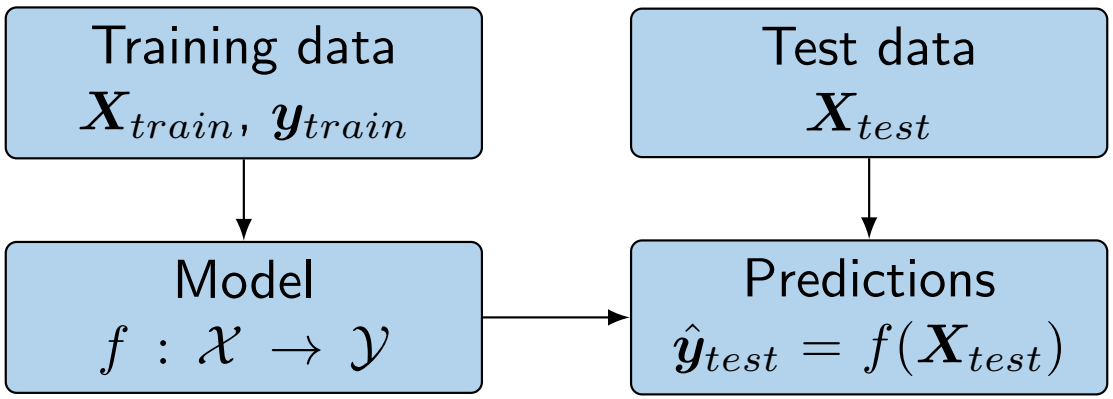
Figures adapted from https://siftscience.com/sift-edu/prevent-fraud

Types of ML problems



Supervised learning

- Given training samples $\boldsymbol{X}_{train} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\} \subseteq \mathcal{X}$ with corresponding targets $\boldsymbol{y}_{train} = \{y_1, ..., y_N\} \subseteq \mathcal{Y}$
- Find a function f that generalizes this relationship, i.e. $f(\boldsymbol{x}_i) \approx y_i$.
- Using f, make predictions $\hat{\boldsymbol{y}}_{test}$ for the test data $\boldsymbol{X}_{test}$.



Supervised learning: Classification

If the targets y_i represent categories, the problem is called classification.

Examples

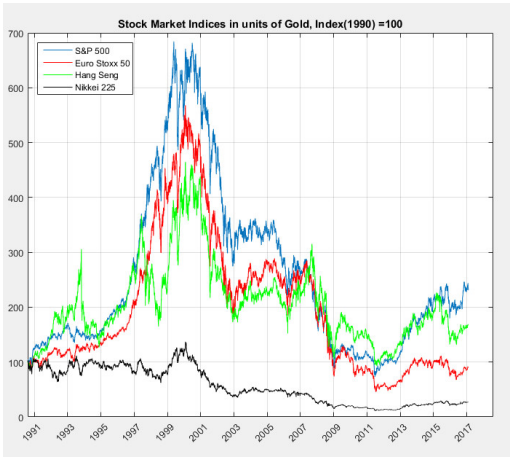
- Handwritten digit recognition
- Transaction classification (fraud, valid)
- Object classification (cat, dog, hotdog, ...)
- Cancer detection

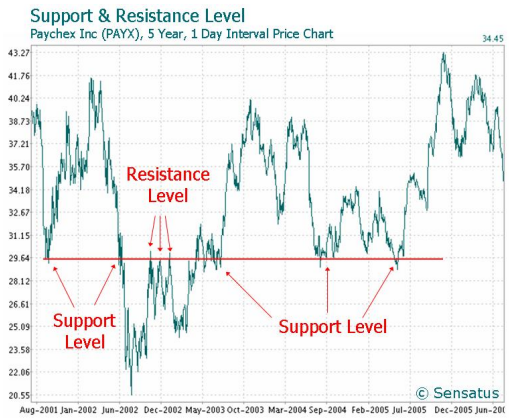
Supervised learning: Regression

If the targets y_i represent continuous numbers, the problem is called regression.

Examples

- Stock market prediction
- Demand forecasting
- User involvement measurement
- Revenue analysis



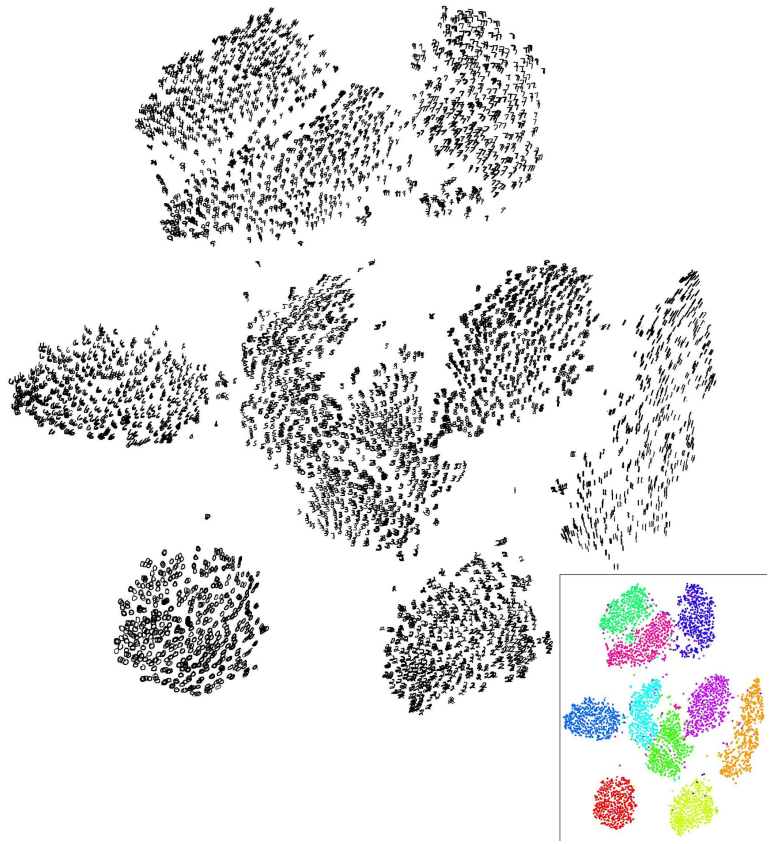


Unsupervised learning

Unsupervised learning is concerned with finding structure in unlabeled data.

Typical tasks

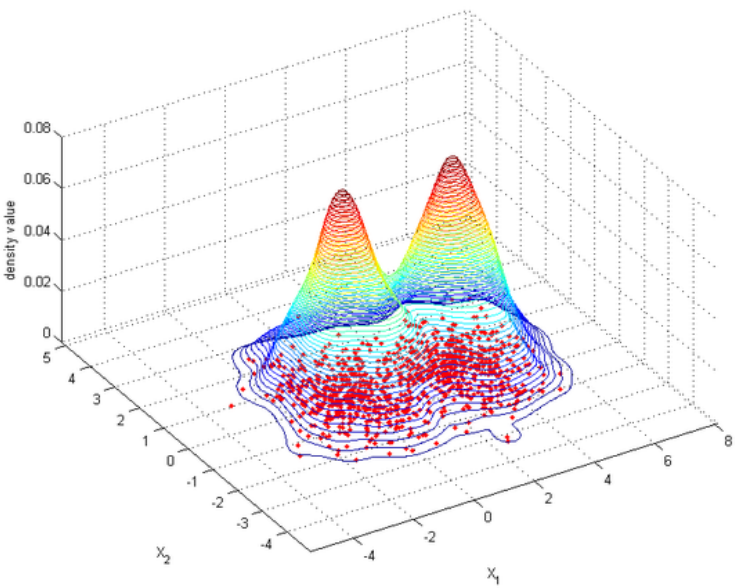
- Clustering
  - Group similar objects together
- Dimensionality reduction
  - Project down high-dimensional data

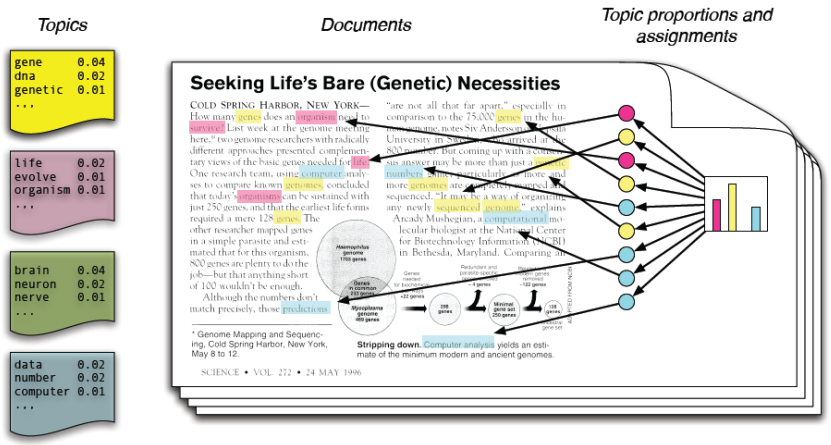


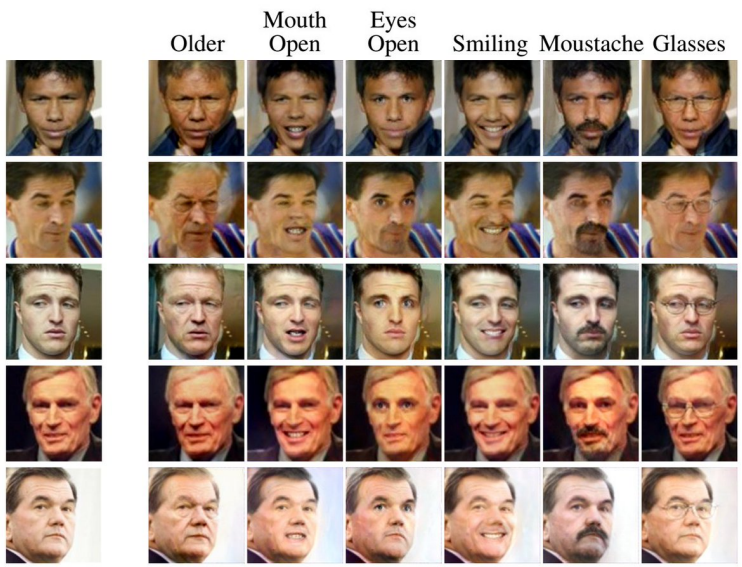
Unsupervised learning

Typical tasks - continued

- Generative modeling
  - (Controllably) generate new "realistic" data
- Topic models
  - Discover hidden semantic structures in text.







Other categories

- Reinforcement learning
  - Learning by interacting with a dynamic environment. Goal is to maximize rewards obtained by performing "desirable" actions.
- Semi-supervised learning
  - Learning to combine lots of unlabeled data with a few labeled examples for further prediction tasks.
- Active learning
  - Learn while obtaining labels by querying an oracle.
- Learning to learn (meta-learning)
  - Learning to construct better models for ML. Operates one level above the standard ML techniques.
- Learning to rank
  - What are the relevant items for a given query? (e.g. Netflix, web search, ad placement)
- And many more...

General information

Staff

- Lecturer: Prof. Dr. Stephan Günnemann
- Teaching assistants:
  Aleksandar Bojchevski, Simon Geisler, Johannes Klicpera, Marten Lienen, Oleksandr Shchur

Details

- 8 ECTS
- Language - English
- Doesn't count for Wirtschaftsinformatik (Information Systems) students
- Doesn't stack with IN2332 in your curriculum

Piazza

- Register at `https://piazza.com/tum.de/fall2020/in2064`
  Access code: `ml2020`
- All announcements will be made on Piazza.
- All course material will be uploaded to Moodle.
- You will miss important information if you don't register.

Use Piazza to ask questions - your emails will likely not be answered.

Schedule

Core content upload

Before Mon	Lecture slides, exercise sheet, lecture video, in-class exercise solution and video
Thu	Previous homework solution and video

Optional

We will hold a Q&A session where you can ask your questions in person (primarily regarding the current topic).

Wed	12:00–14:00	Online Q&A session

Schedule & Logistics

Week #	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
N - 1						Slides, lecture video, exercise sheet, in-class solution & video N (or earlier)	
N			Q&A session N				
N + 1			23:59 HW N deadline	Homework solution & video N			

- Lecture slides and exercise sheet for topic of week N are uploaded before Monday of week N.
- Submit Homework via Moodle (see sheet 1 for detailed instructions).
- Homework of week N is due on Wednesday of week $N + 1$ at 23:59.
- Exercise solutions are published on Moodle.
- Homework of week N will be discussed in video of week $N + 1$.

Grading

Exam

- Written final exam, probably in February
- Most likely remote exam
- 120 minutes
- Open book
- Bonus of 0.3 if you show sufficient effort (1 out of 4 points) for $\geq 75\%$ of HW sheets.
- Only grades in the range 1.3 - 4.0 can be improved.

Group formation

- Submit homework in groups of up to 3 people
- You can also work alone by forming a **group of 1**.
- You can only select/change groups before the first deadline, i.e. **before Wednesday, Nov. 11, 23:59**.
- After this date the groups are fixed.

Planned weekly schedule

Week	Date	Topic
1	Nov 2	Introduction, basic concepts
2	Nov 9	k-nearest neighbors, decision trees
3	Nov 16	Probabilistic inference
4	Nov 23	Linear regression
5	Nov 30	Linear classification
6	Dec 7	Optimization
7	Dec 14	Deep learning 1
8	Dec 21	Deep learning 2
9	Jan 11	Support vector machines & kernel methods
10	Jan 18	Dimensionality reduction & matrix factorization 1
11	Jan 25	Dimensionality reduction & matrix factorization 2
12	Feb 1	Clustering, mixture models
13	Feb 8	Advanced topics (fairness, privacy, robustness, ...)

Contents

- This is an introductory, theoretical Machine Learning course
  - There will be a fair amount of theory and mathematics
  - We will focus on fundamental Machine Learning concepts
  - We will mostly discuss independent (iid) data

- Next semester, we will cover (even) more advanced topics (IN2323)
  - Generative models
  - Robustness
  - Sequential data
  - Graphs & networks
    - $\rightarrow$ These are the core research topics of our group :-)

What this course is not about

- This is not a pure Deep Learning course
  - look at IN2346, IN2349 instead
- This is not a course about Big Data (Hadoop, etc.)
  - look at IN2326 instead
- This is not an applied Data Science / Business Analytics course
  - look at IN2028, IN2339 instead

Recommended reading

Our official reading recommendation:

- Christopher M. Bishop, *Pattern Recognition and Machine Learning*. Springer, Berlin, New York, 2006 (free, online version available).

but we also like:

- Kevin Murphy, *Machine Learning: A probabilistic perspective*. MIT Press, 2012.

What's next?

Brush up on your linear algebra, calculus, and probability theory knowledge.

Read

- `http://cs229.stanford.edu/section/cs229-linalg.pdf`
- `http://cs229.stanford.edu/summer2020/cs229-prob.pdf`
- Bishop [ch. 1.2.0 - 1.2.3, 2.1 - 2.3.0]
- Solve the math refresher (exercise sheet 1)