Organization + Introduction

Maschinelles Lernen -
Grundverfahren WS21/22

Prof. Gerhard Neumann
Autonome Lernende Roboter (ALR)
KIT, Institut für Anthrophomatik und Robotik

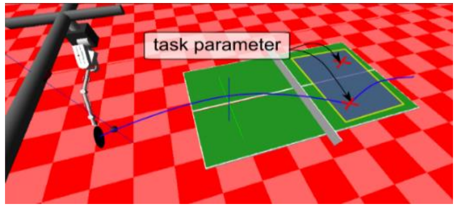
About me and ALR…

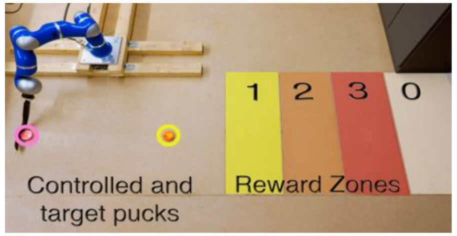
Prof. Gerhard Neumann:

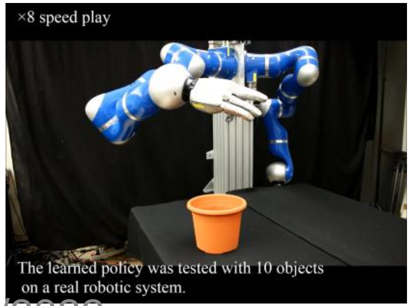
- Institut für Anthrophomatik und Robotik
- Lehrstuhl: Autonome Lernende Roboter (ALR)
- **Email: Gerhard.Neumann@kit.edu**

Research Topics: Machine Learning for Robotics

- Reinforcement Learning
- Probabilistic Machine Learning
- Deep Learning
- Interactive Learning







About me…

Timeline:

- Dissertation 2012 at the TU Graz
- 2014-2016: Junior Professor, TU Darmstadt
- 2016-2019: Professor, University of Lincoln
- 2019: Bosch Group Leader,
  “Information-theoretic Reinforcement Learning”

From 1. January 2020:

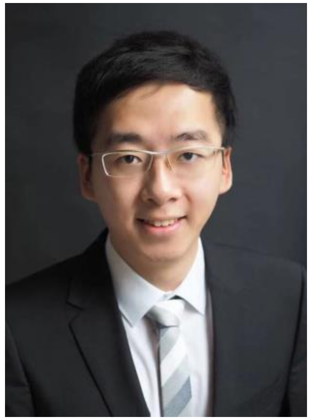
- Professorship “Autonome Lernende Roboter”, KIT

Introducing the TAs

Michael Volpp

michael.volpp@kit.edu

- Started as PhD student with Geri in 2018
- Research focus on Bayesian Meta-Learning



Ge Li

ge.li@kit.edu

- Started as PhD Student at KIT in 2020
- Research focus on Robot motion primitives and planning

Philipp Becker

philipp.becker@kit.edu

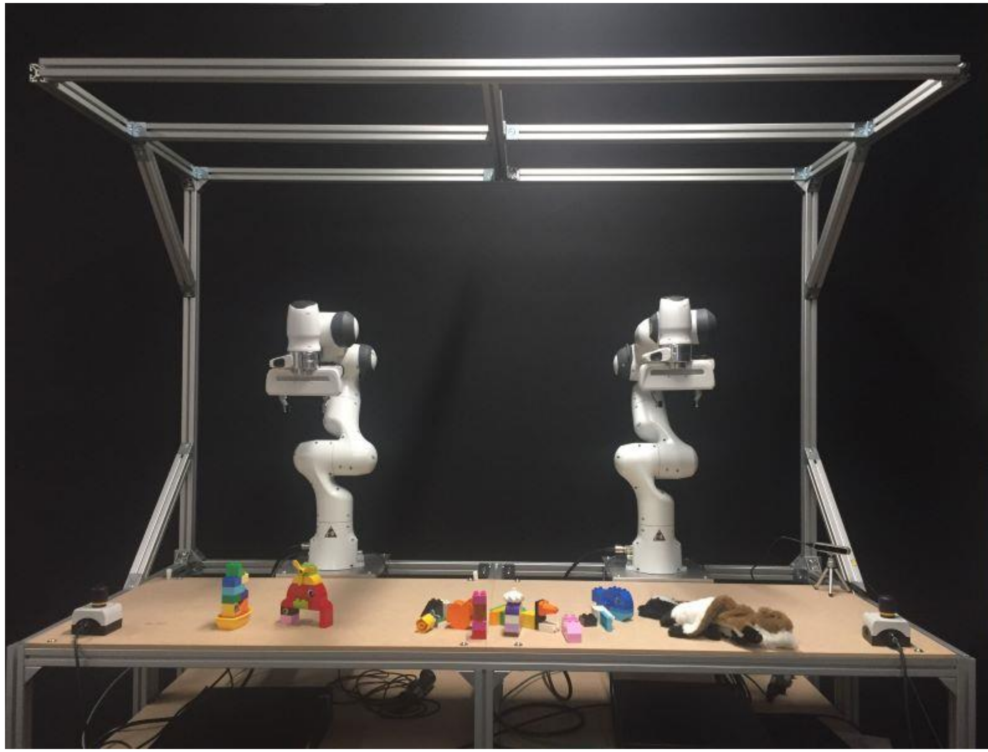
- Started as PhD Student at KIT together with Geri in beginning of 2020
- Research focus on time series and multi-modal modelling for robotics

- Feel free to contact us over the forum or mail

A bit of self-advertisment

What else do we offer?

- Interested in a Master-Thesis or Bachelor Thesis?
  - Have a look at https://alr.anthropomatik.kit.edu/
  - We always take motivated students
  - Interesting research topics: Robot Reinforcement Learning, Deep Learning, Imitation Learning, Robotics, Human-Robot Collaboration, Variational Inference
  - Use real robots (Franka Panda arms)
  - Joint supervision of PhD students and me
  - High success-rate of turning your thesis into a paper!



This Semester:

- **New Lecture:** Reinforcement Learning

Organization

Start of the lecture: 22.10.2020

- Friday: 14:00 – 15:30
- Location: Physical and Virtual
- All lectures will be recorded and put on Illias

Language:

- (Austrian) English
- Why? All the terminology / research papers are in English
- Getting used to English for these technical terms is crucial!

Exam:

- Written
- Date to be announced

Material

Lecture Material:

- Mostly slide-based
- English
- Sometimes additional lecture notes will be available (not part of exam)

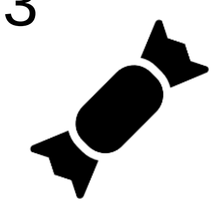
Machine Learning is very math heavy!

- Understand, not just apply!
- **Math basics:** We will recap the required math before it is used to derive the algorithms
- Math is directly applied… actually quite fun ☺

Exercises - General Info

There will be 6 Exercises

- 1 Exercise every 2 weeks
- Starting 4.11 (TBA)
- Hand in Thursday before next exercise is presented
- Solutions will be presented
- Work in groups of 3



There is a Bonus!!

- You get 0.3 bonus in the exam if you pass and have > 60% of exercise points
- There is only a joint grade of lecture and excercise

Exercises - Format

Mixture of pen and paper as well as coding

- We will use python for coding
- We will use Jupyter notebooks
- Might be a bit challenging, but you work in groups and it's a good preparation for exam

ML @ KIT

Confusion:

- Maschinelles Lernen 1 – Grundverfahren from Prof. Zöllner,
- Based on the old Machine Learning Lecture with Prof. Dillmann
- Fakultät für Wirtschaftwissenschaften, AIFB Institute
- Not applicable for Computer Science students

This lecture: Machine Learning – Fundamentals and Algorithms (old name: Maschinelles Lernen – Grundverfahren)

- New content
- Wahlfach for computer science
- More math, more theory, more programming!

ML @ KIT

Other ML lectures:

- SS: Deep Learning and Neural Networks: Prof. Waibel
- SS: Deep Learning for Computer Vision: Prof. Stiefelhagen
- WS: Optimization Methods for Machine Learning and Engineering
- SS: Cognitive Systems, Prof. Waibel and **Me**
- SS: Pattern Recognition, Prof. Beyerer

New Lecture:

- WS: Reinforcement Learning, **Me**

Ask questions!!!

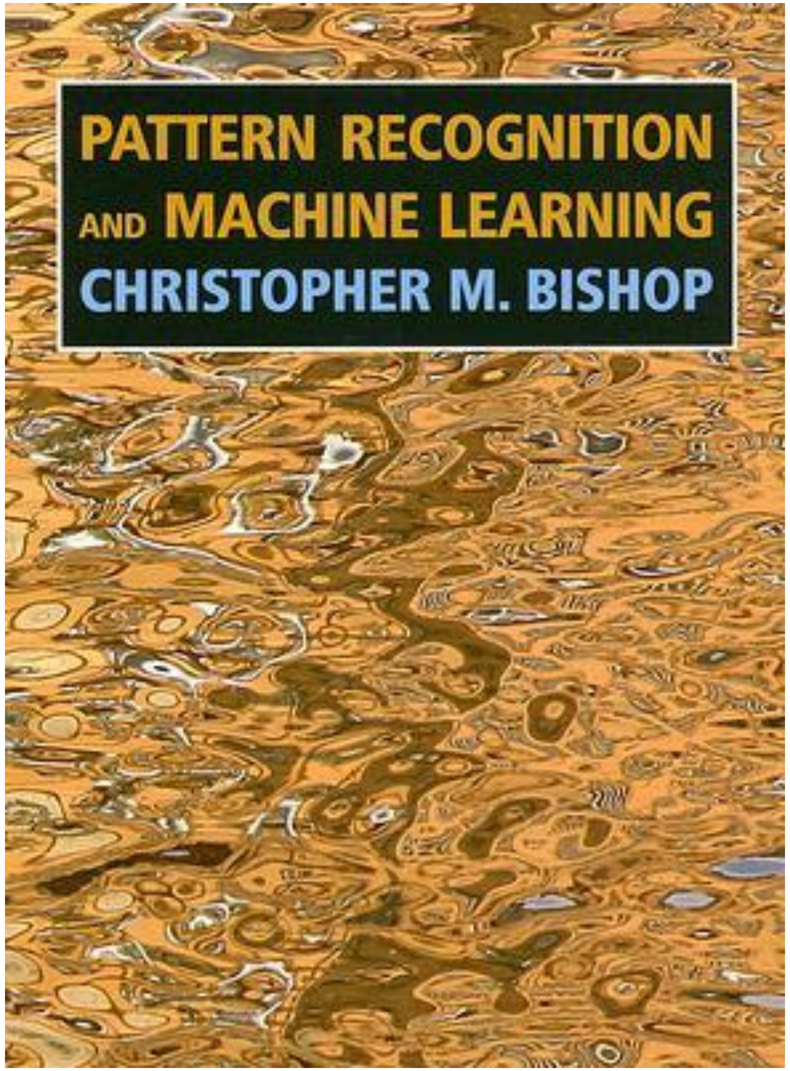
Even though I am Austrian, I am actually a nice guy…

- If it is not clear… tell me!!
- If it is too fast… tell me!!
- If you can not understand „austrian english“ … tell me!

Additional Reading

Pattern Recognition and Machine Learning

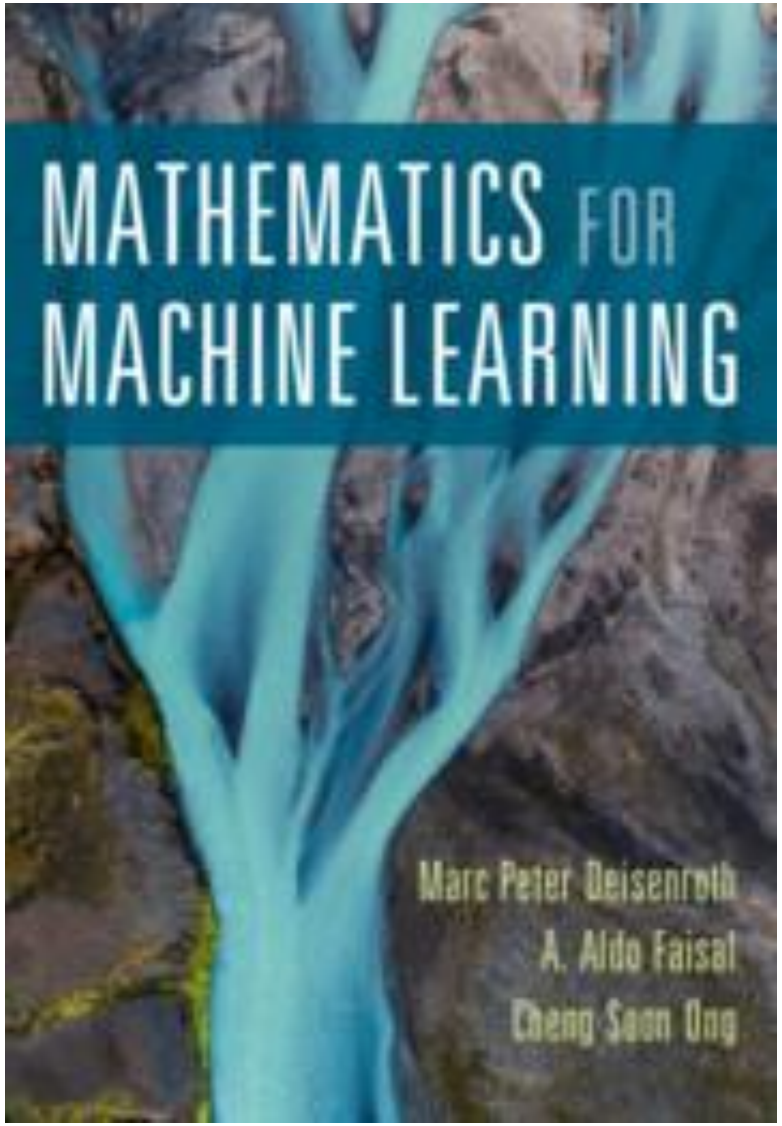
- Christopher Bishop, Springer, 2006
- Very nicely explained fundamentals in classification and regression
- PDF online



Additional Reading

Mathematics for Machine Learning

- Marc Deisenroth, Aldo Faisal and Cheng Ong
- Cambridge Press 2020
- Available as PDF online



A new friend…

The matrix cookbook

- Great collection of matrix identities
- We will use a few of them

Agenda for today

Lets take it easy...

- Introduction in Machine Learning

Introduction in
Machine Learning

What is learning?

- *"Learning denotes* ***changes in a system*** *that ... enable a system to do the same task … **more efficiently** the next time."* - Herbert Simon *(Nobel Prize in Economics)*
- *"Learning is constructing or* ***modifying representations*** *of what is being* ***experienced****."* - Ryszard Michalski *(ML pioneer)*
- *"Learning is making* ***useful changes*** *in our minds."* - Marvin Minsky (MIT)

What is Machine Learning?

Algorithms that can **improve** their **performance** using **training data**

- Typically we have a large number of **parameters**
- Learned from data

Useful if:

- **No expert knowledge available:** industrial/manufacturing control, mass spectrometer analysis, drug design, astronomic discovery
- **Black-box expert knowledge:** face/handwriting/speech recognition, driving a car, flying a plane
- **Fast changing phenomena:** credit scoring, financial modeling, diagnosis, fraud detection
- **Customization/personalization:** personalized news reader, movie/book recommendation

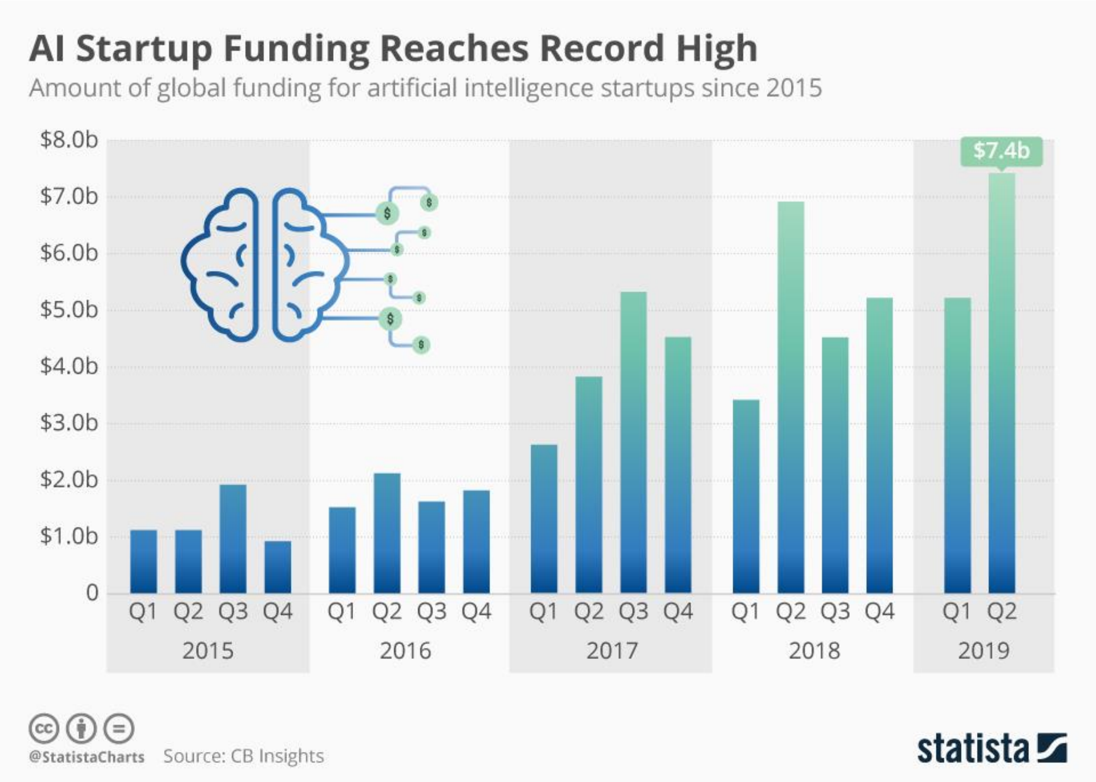
Machine Learning is a hot topic

Do we see a new era of machine learning?

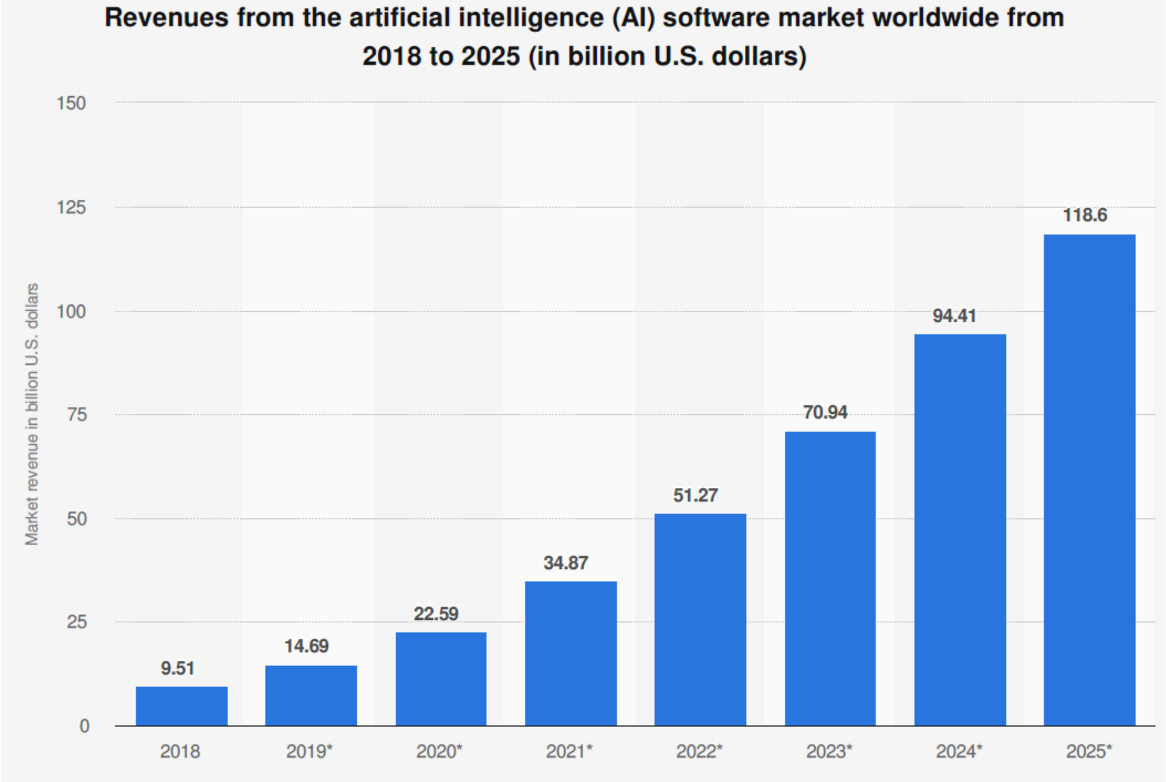
- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, Director, DARPA)
- Machine learning is the hot new thing” (John Hennessy, President, Stanford)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)

Commercial Importance

AI startup funding



Predicted Revenue of AI

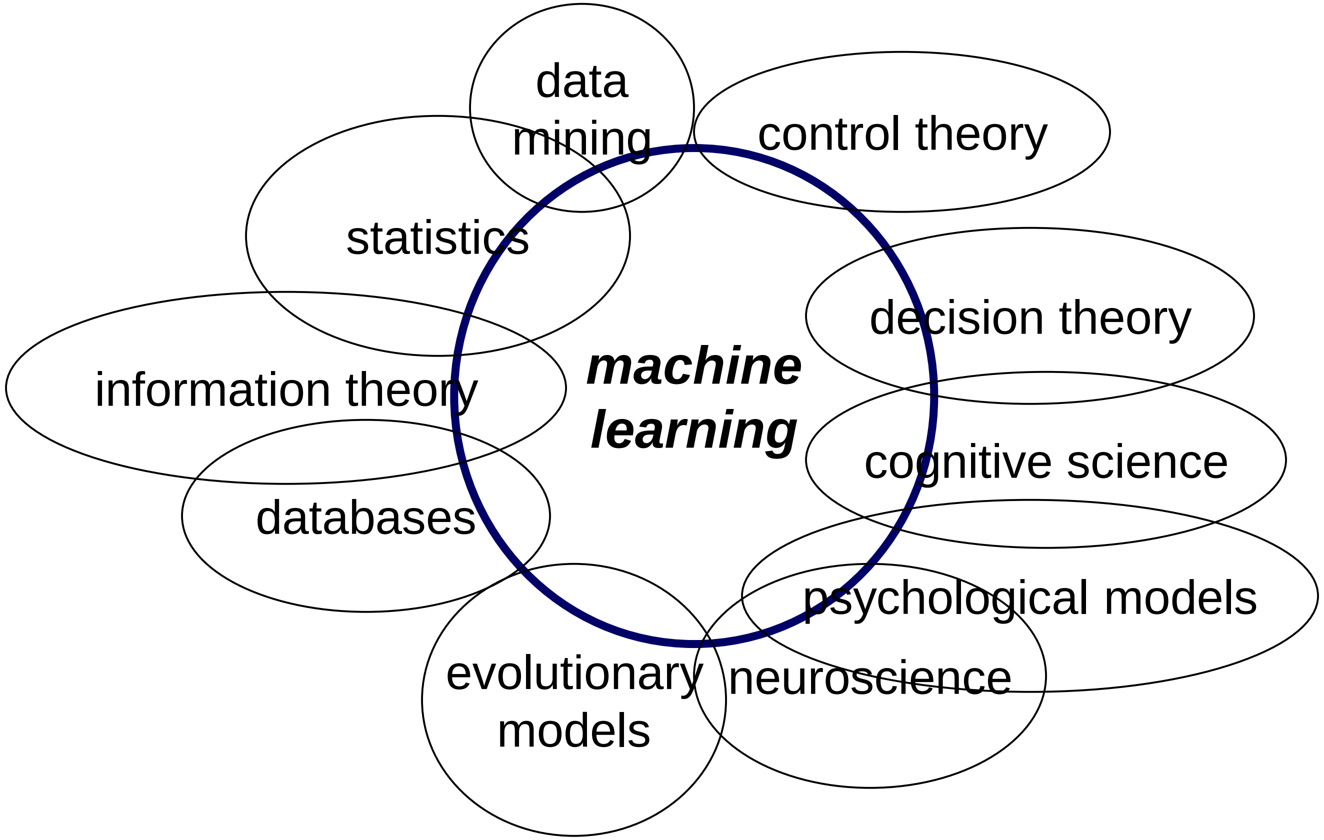


Why do we see this “explosion” now?

- More training data
- More computation power
- New algorithms (Deep Learning)…
  - But: same principles as from 40 years ago are still in use

- "when you go from 10,000 training examples to 10 billion training examples, it all starts to work. Data trumps everything.” Google translate engineer

Machine Learning is interdisciplinary



ML: uses these disciplines to create more accurate and efficient computer systems.

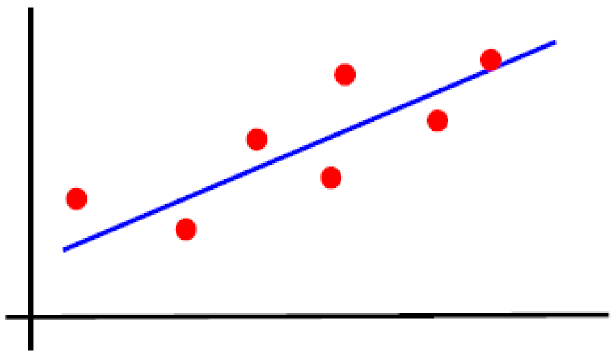
Different Types of Learning

1. **Supervised Learning**
   – Training data includes target values
2. **Unsupervised Learning**
   – Training data does not include target values
3. **Reinforcement Learning**
   – No target values, but evaluation (reward) of the output

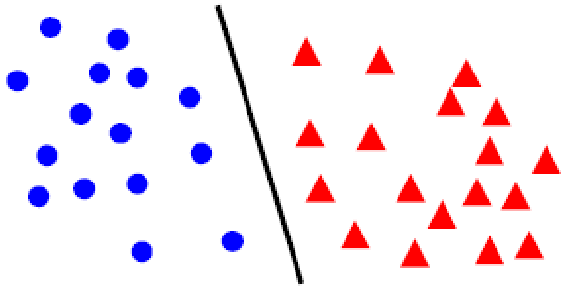
Supervised Learning

Training data includes targets

- **Regression:**
  - Learn continuous function
  - Example: line



- **Classification:**
  - Learn class labels
  - Example: Digit recognition

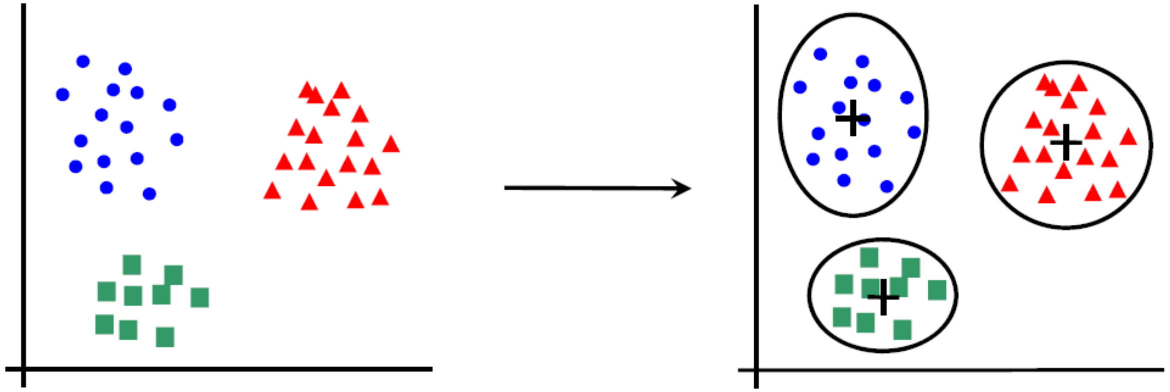


Unsupervised Learning

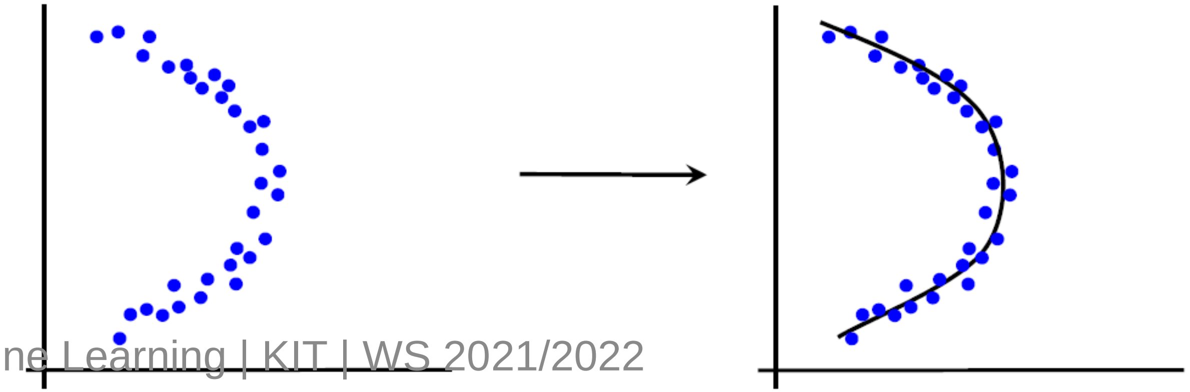
Training data does not include target values

- Model the data

- **Clustering:**



- **Dimensionality reduction:**

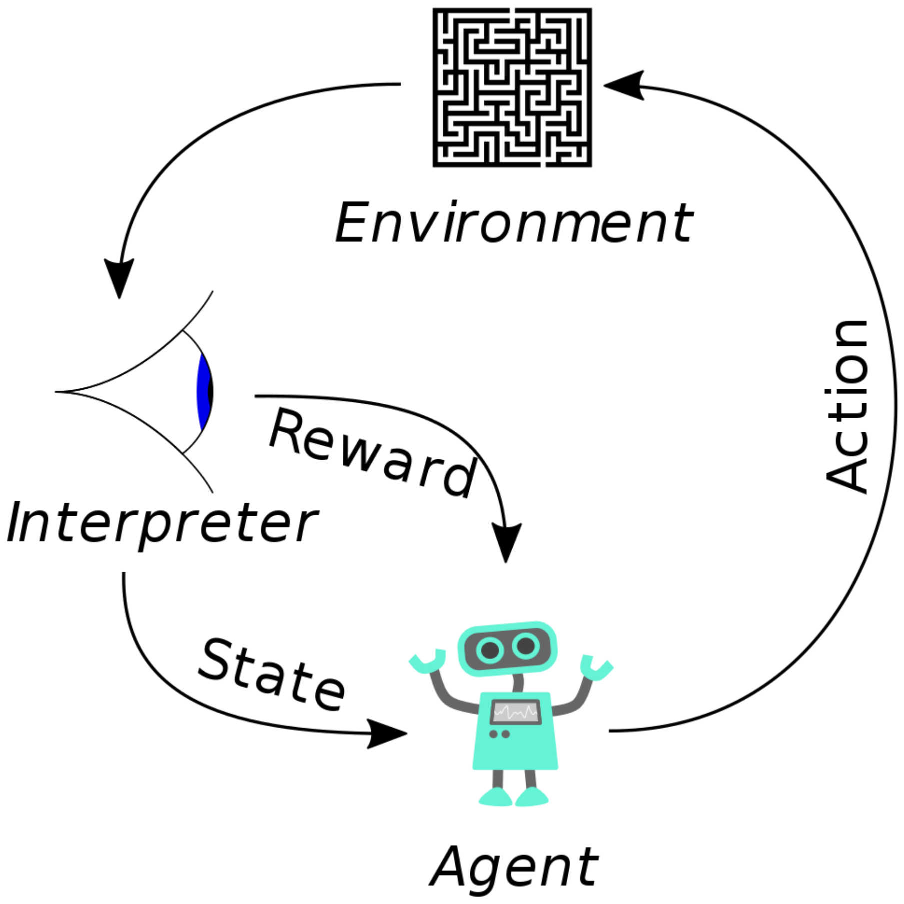


Reinforcement Learning

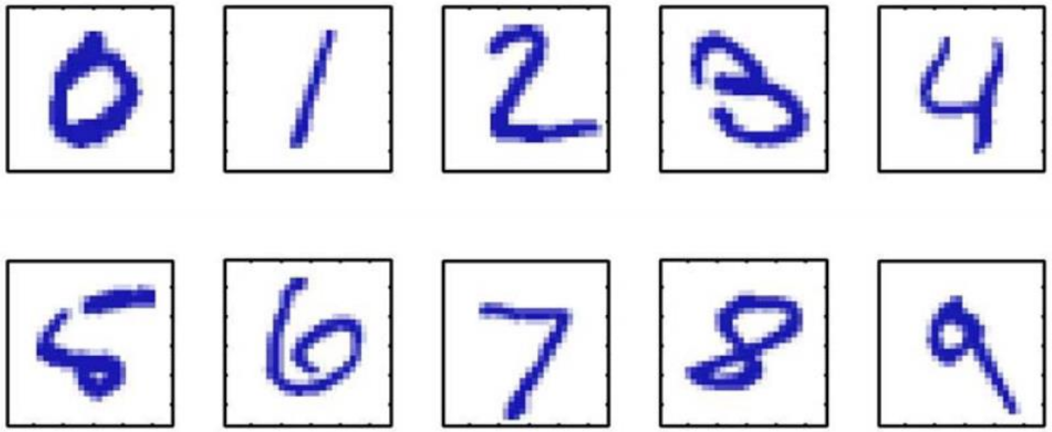
- No supervisor, but reward signal
- Selected actions also influence future states

Not part of this lecture!

… but: New lecture in this semester!



Example 1: hand-written digit recognition



Images are 28 x 28 pixels

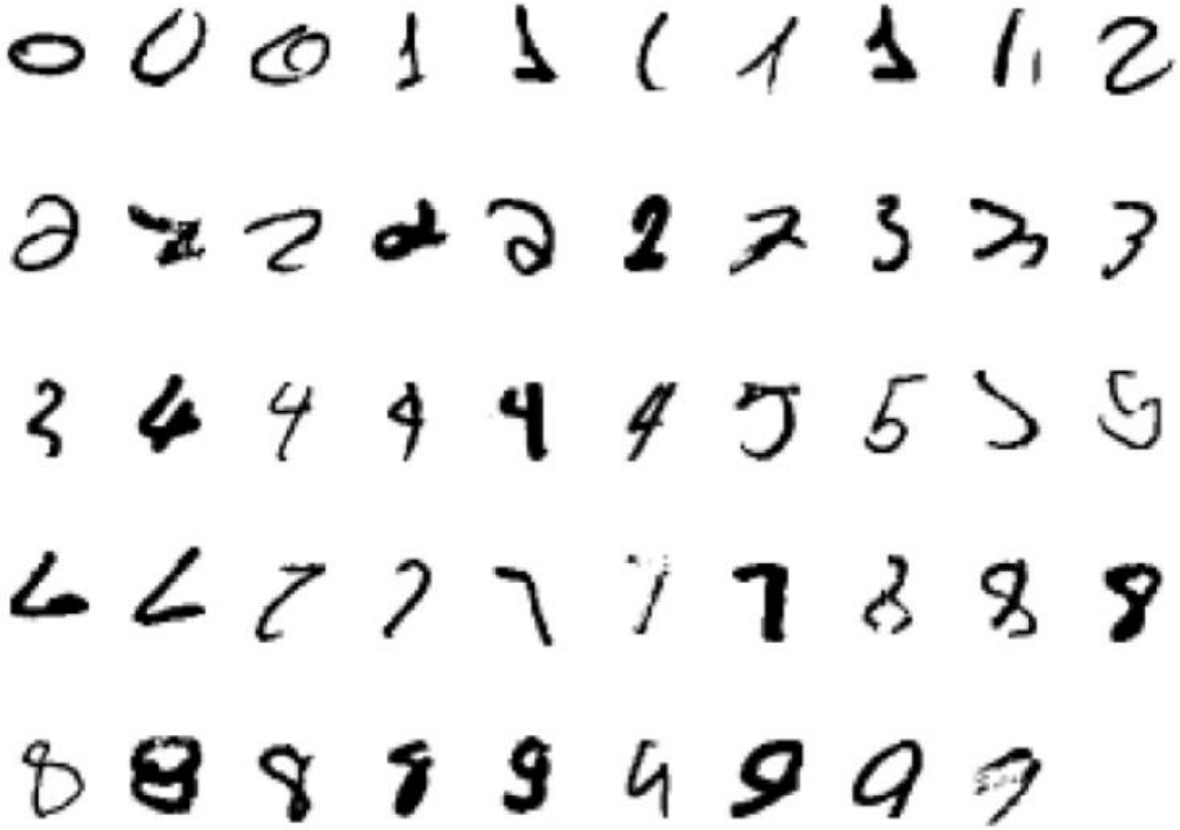
- Represent image as vector
- Learn Classifier

$$\mathbf{x} \in \mathbb{R}^{784}$$

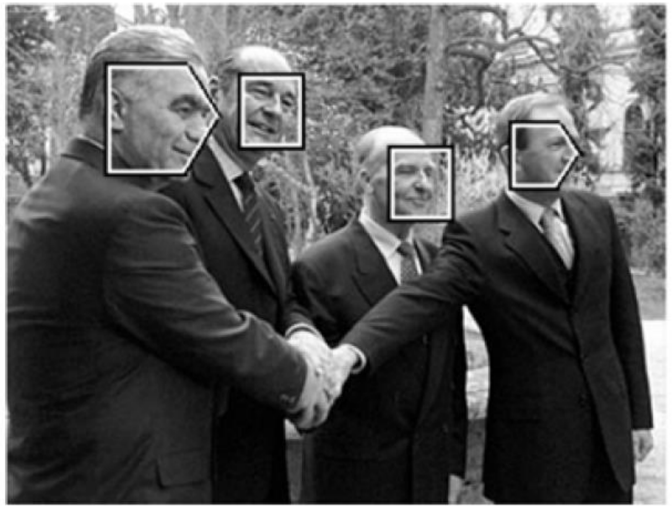
$$f : \mathbf{x} \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Example 1: hand-written digit recognition

- Supervised Classification Problem
- Training-Set: 6000 examples per class
- Error on a test-set: 0.4%
- One of the first applications used in a commercial product (ZIP-Codes, cheques,…)



Example 2: Face Detection



- **Classification Problem**
- **Classify image windows in 3 classes**
  - Non-face
  - Frontal-face
  - Profile-face

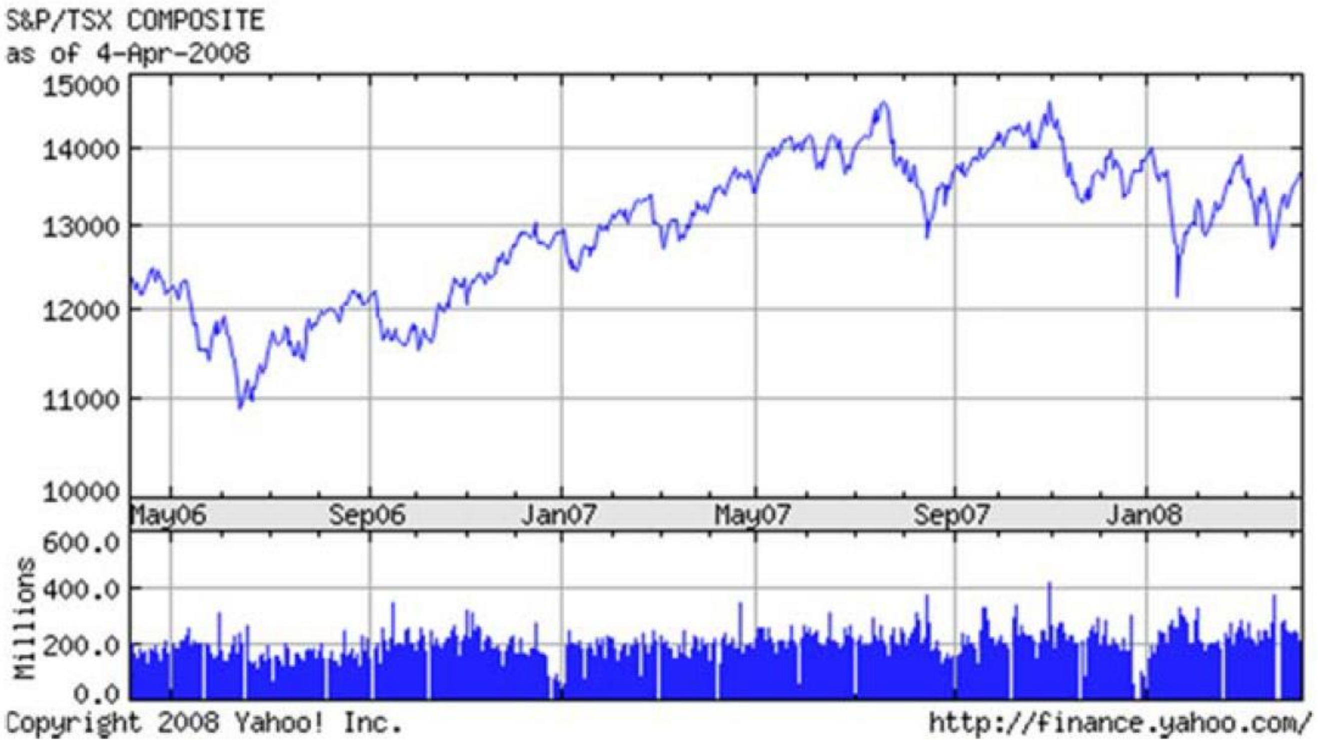
Example 2: Face Detection

Training data for frontal phases

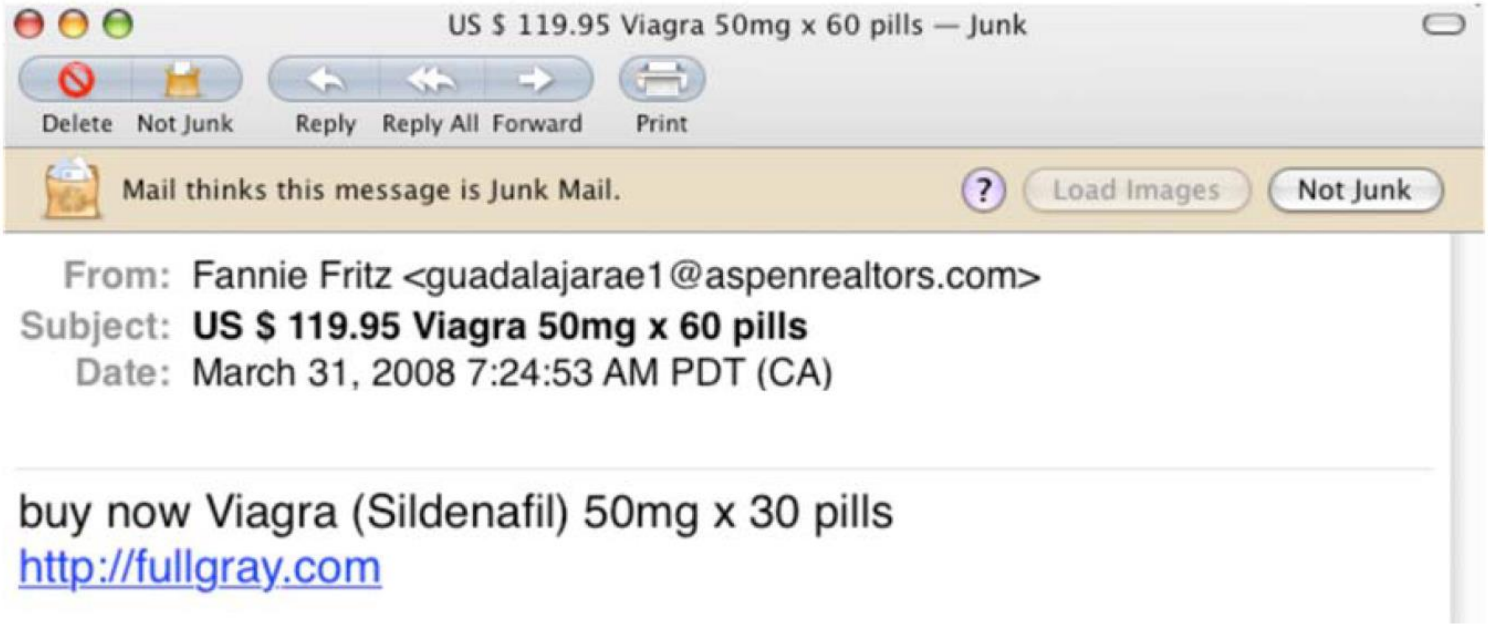
- 5000 faces
  - All near frontal
  - All ages, races, gender, lighting
- 10^8 non-faces
- Normalization (scale, translation)

Example 3: Stock-price prediction

- Prediction of the stock prices
- Regression problem
  continuous outputs)



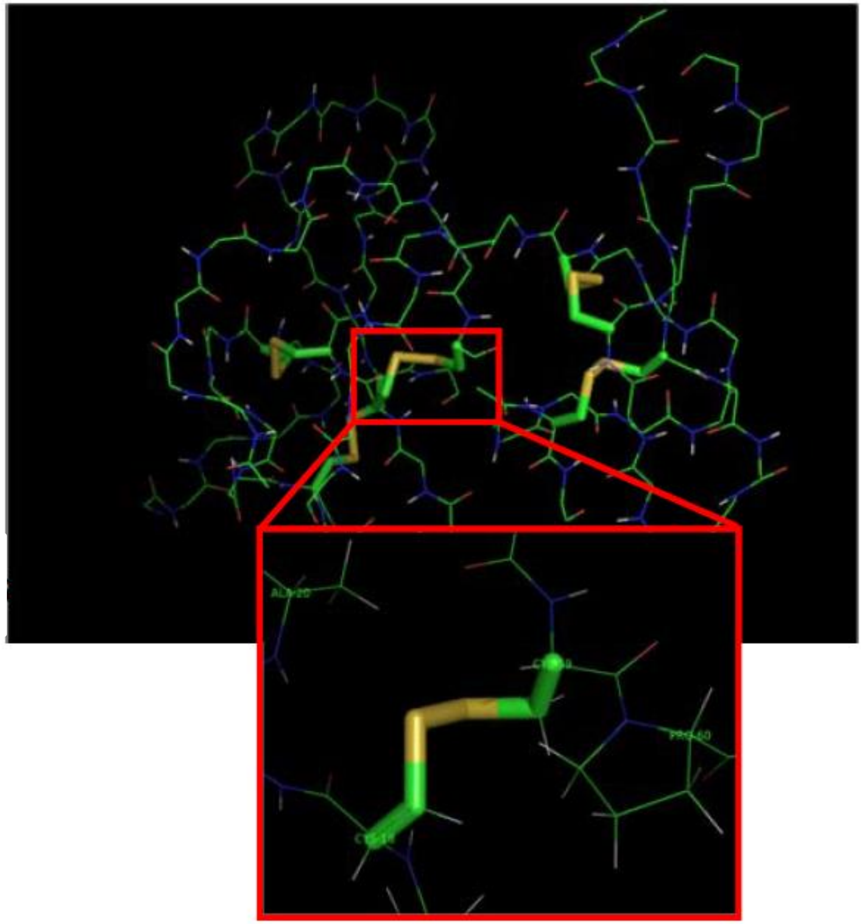
Example 4: Spam Detection



- Classify emails in spam / non-spam
- Data x represented by word count, z.B.: “Viagra”, “outperform”, “you may be surprised to be contacted”…
- Spam strategies change -> we need learning

Beispiel 5: Computational Biology

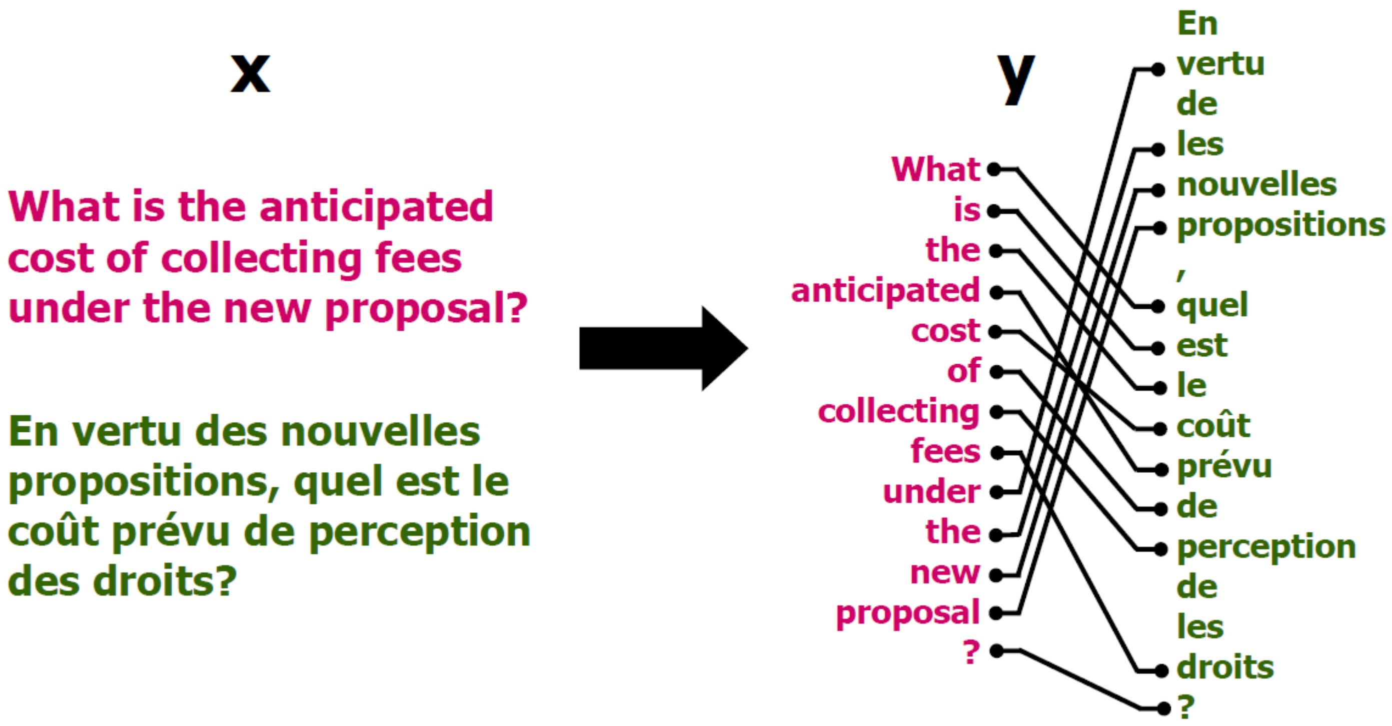
AVITGACERDLQCG
KGTCCAVSLWIKSV
RVCTPVGTSGEDCH
PASHKIPFSGQRMH
HTCPCAPNLACVQT
SPKKFKCLSK



- Given the protein sequence, predict 3D structure

Example 6: Language Translation

- Learn from aligned text translations



e.g. Google translate

Example 6: Language Translation

Example 7: Recommender Systems

Example 8: Robot Manipulation

- Reinforcement Learning: Learning to grasp unknown objects (Google)

Example 9: Tossing objects

- Reinforcement Learning: Learning to throw objects into bins

Example 10: Semantic Segmentation

Example 11: Creating Image Captions

Description:

A falcon is eating during sunset.
The falcon is standing on earth.

Poem:

Like a falcon by the night
Hunting as the black knight
Waiting to take over the fight
With all of it's mind and might

Example 12: Image generation

Figure: synthetic images generated by [pg-GAN]() from Nvidia. None of these images are real!

More success stories in the last 3 years

- AI beats human in lip reading
- Improved translation with Google Neural Machine Translation System
- OpenAI Bot dominates DOTA 2 Champions
- CMU AI beats top poker players
- Google Deepmind beats world champion in “Go”

More success stories

- Deepmind reduces Google’s Data Center cooling costs by 40%
- Google Duplex: AI System performing every-days phone calls
- Tacotron 2: Generating human speech from text
- State-of-the-art speech recognition via Sequence to Sequence models

Conceptual view on ML

- Thousands of learning algorithms
- Hundreds of new algorithms every year
- We will only look at the fundamental algorithms and discuss the principles that connect them

Every ML algorithm consists of 3 parts:

- **Representation**
- **Evaluation**
- **Optimization**

Representation

What is the underlying representation of our model?

- Decision trees
- Instances
- Mixture Models
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

Evaluation

What are we optimizing for ?

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- KL divergence
- Etc.

Optimization

How do we optimize?

- Least squares solution
- Gradient descent
- Gradient descent with adaptive learning rate
- 2^{nd} order methods
- Constraint optimization
- Random search

What will we cover?

We will cover the fundamentals for

- **Representation**
- **Evaluation**
- **Optimization**

With a strong mathematical focus on understanding and deriving the algorithms

Lecture Content

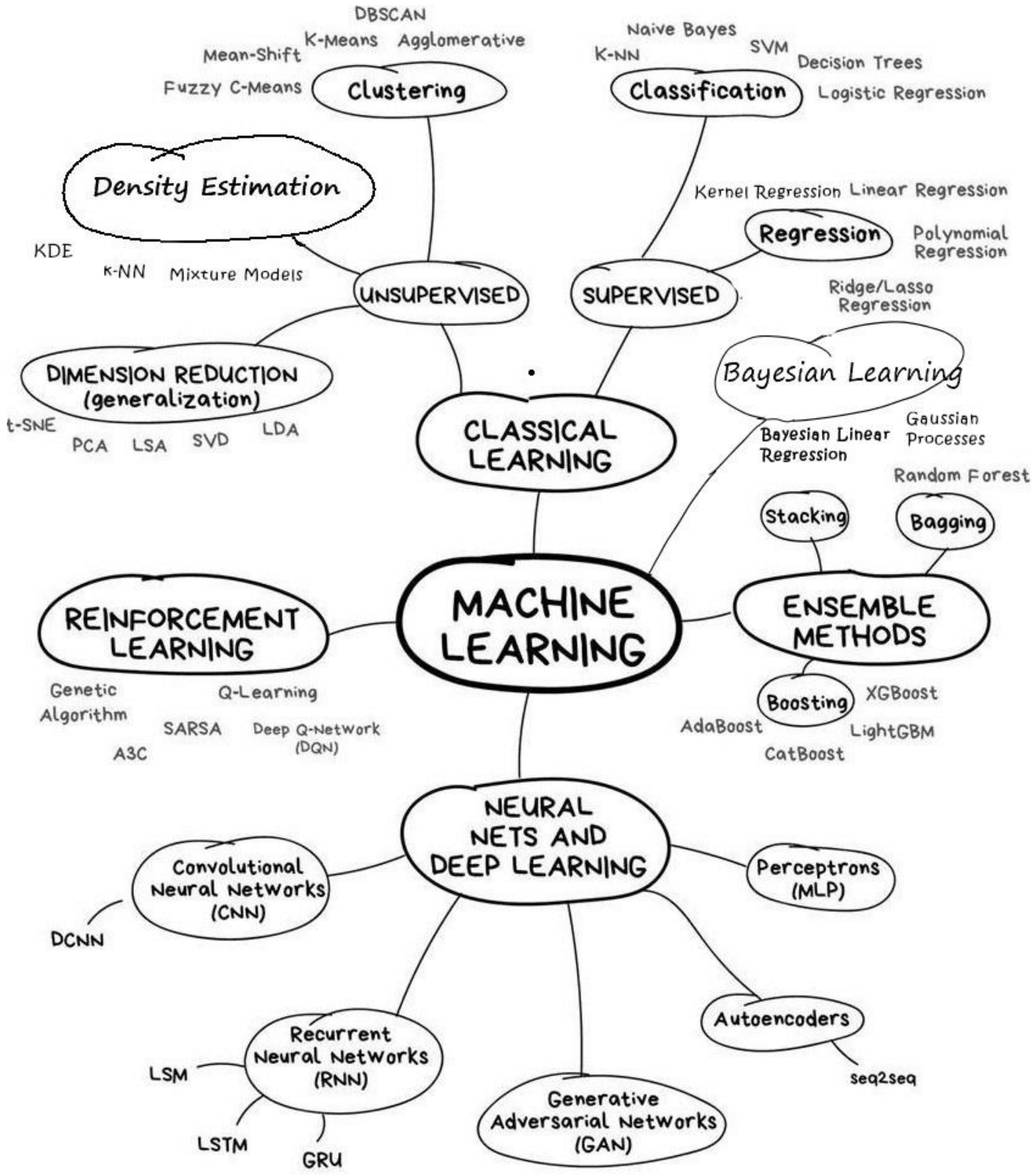
Chapter 1: Classical Supervised Learning
- Lecture 1: Linear Regression, Ridge Regression
- Lecture 2: Linear Classification
- Lecture 3: Model Selection
- Lecture 4: k-Nearest Neighbors, Trees and Forests

Chapter 2: Classical Unsupervised Learning
- Lecture 5: Dimensionality Reduction and Clustering
- Lecture 6: Density Estimation and Mixture Models

Chapter 3: Kernel Methods
- Lecture 7: Kernel-Regression
- Lecture 8: Support Vector Machines

Chapter 4: Bayesian Learning
- Lecture 9: Bayesian Linear Regression and Gaussian Processes

Chapter 5: Neural Networks
- Lecture 10: Neural Networks and Backpropagation
- Lecture 11: CNNs and LSTMs
- Lecture 12: Variational Auto-Encoders (?)



Lecture Content

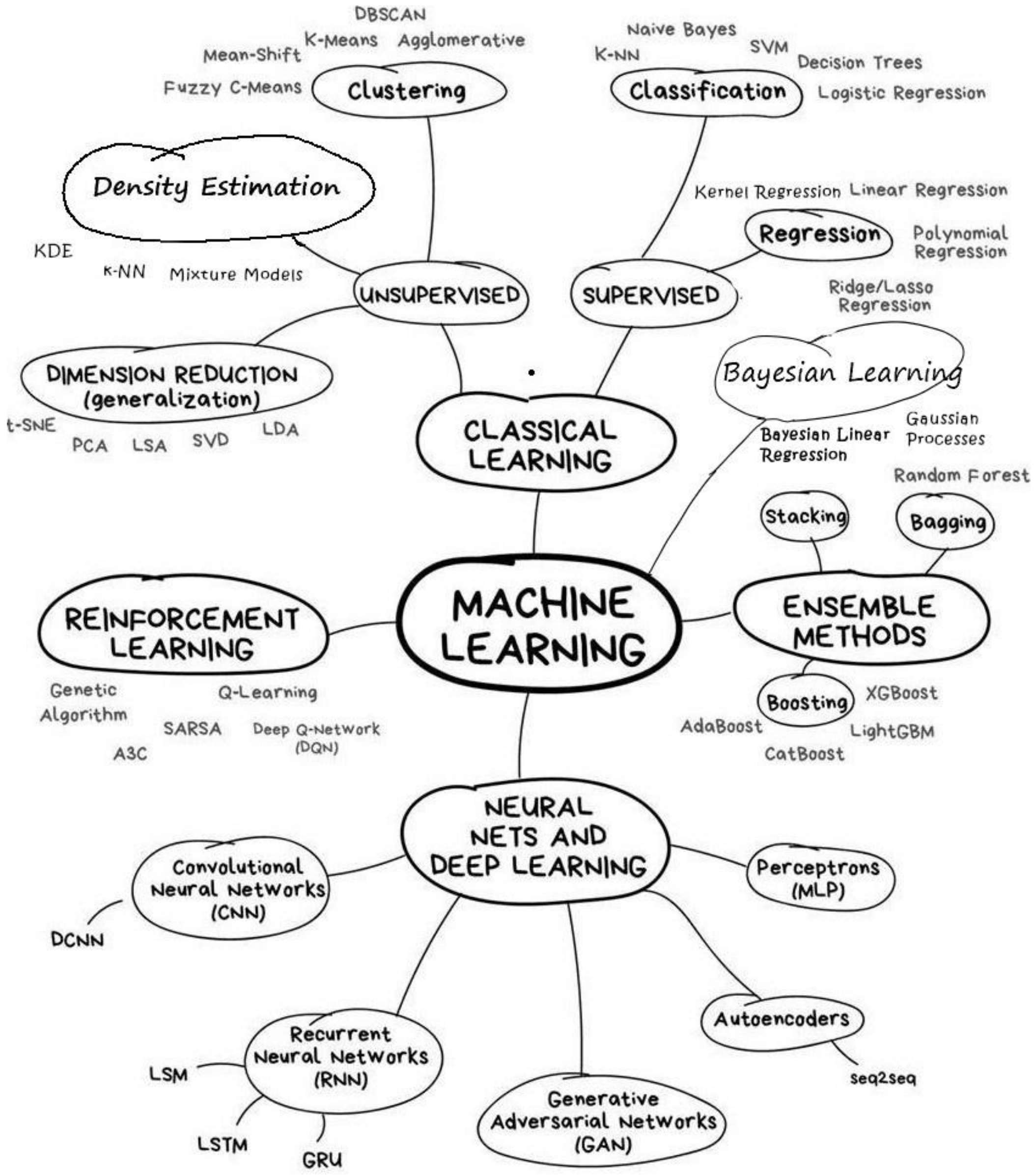
Chapter 1: Classical Supervised Learning
- Lecture 1: Linear Regression, Ridge Regression
- Lecture 2: Linear Classification
- Lecture 3: Model Selection
- Lecture 4: k-Nearest Neighbors, Trees and Forests

Chapter 2: Classical Unsupervised Learning
- Lecture 5: Dimensionality Reduction and Clustering
- Lecture 6: Density Estimation and Mixture Models

Chapter 3: Kernel Methods
- Lecture 7: Kernel-Regression
- Lecture 8: Support Vector Machines

Chapter 4: Bayesian Learning
- Lecture 9: Bayesian Linear Regression and Gaussian Processes

Chapter 5: Neural Networks
- Lecture 10: Neural Networks and Backpropagation
- Lecture 11: CNNs and LSTMs
- Lecture 12: Variational Auto-Encoders (?)



Why so many algorithms?

- **Different use cases**
  - regression, classification, density estimation, etc...
- **Different properties of the data:**
  - Vectors, images, text, graphs, time-series, low-D vs high-D,...
- **Different trade-offs**
  - Complexity vs. number of needed data points
  - Computation speed (training and testing)
  - Interpretability
  - Sensitivity to overfitting
- **Nowadays everyone is talking about Deep Learning...**
  - In most cases we would also choose a deep learning method, but...
  - We also need to understand the classical algorithms to understand fundamental problems in machine learning. They also provide a good introduction
  - In same cases, simple methods still work very well (e.g. small data regimes)

What will we not cover?

- Reinforcement Learning
- Genetic Algorithms
- Natural Language Processing (NLP)
- Generative Adversarial Networks
- Graphical Models
- Sampling Methods
- Variational Inference
- Recommender Systems
- Topic Models
- …

Please visit the advanced courses

The end…

- Any further questions?