Applied Machine Learning
Overview

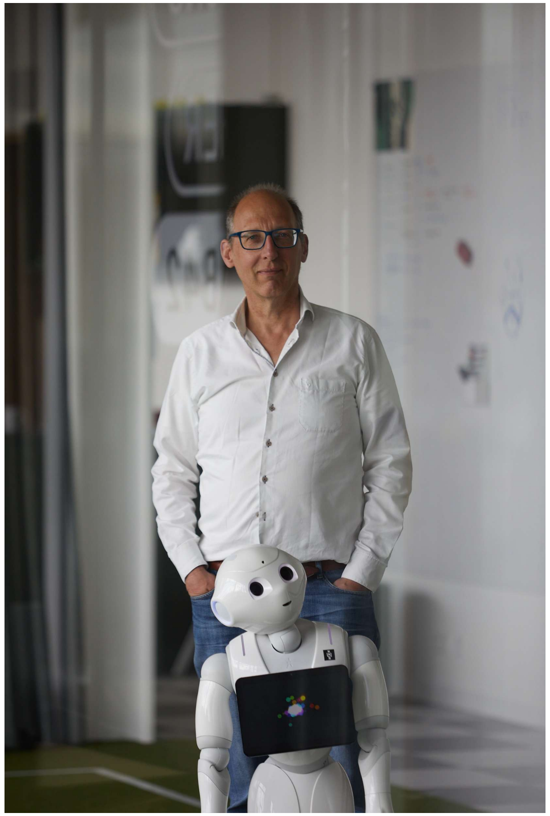
BSc course Informatiekunde 2026
https://staff.fnwi.uva.nl/a.visser/education/AML

Arnoud Visser
Intelligent Robotics Lab & Computer Vision Lab
Informatics Institute
Universiteit van Amsterdam
A.Visser@uva.nl

Illustrations courtesy of Maarten Marx, Sarah Guido, Benjamin Bengfort, and many others.

A little bit about Arnoud Visser

Universiteit van Amsterdam



Universiteit Leiden

Arnoud Visser

PhD in Computer Science,

Master in Physics, Minor in BioChemistry

A little bit about Rein Lukkes

TU Delft

Technische Universiteit Delft

Universiteit Utrecht

Rein Lukkes

Medical Imaging master,

Bsc in Computer Science & Game Technology

A little bit about Pim Dijkhuizen

Universiteit van Amsterdam

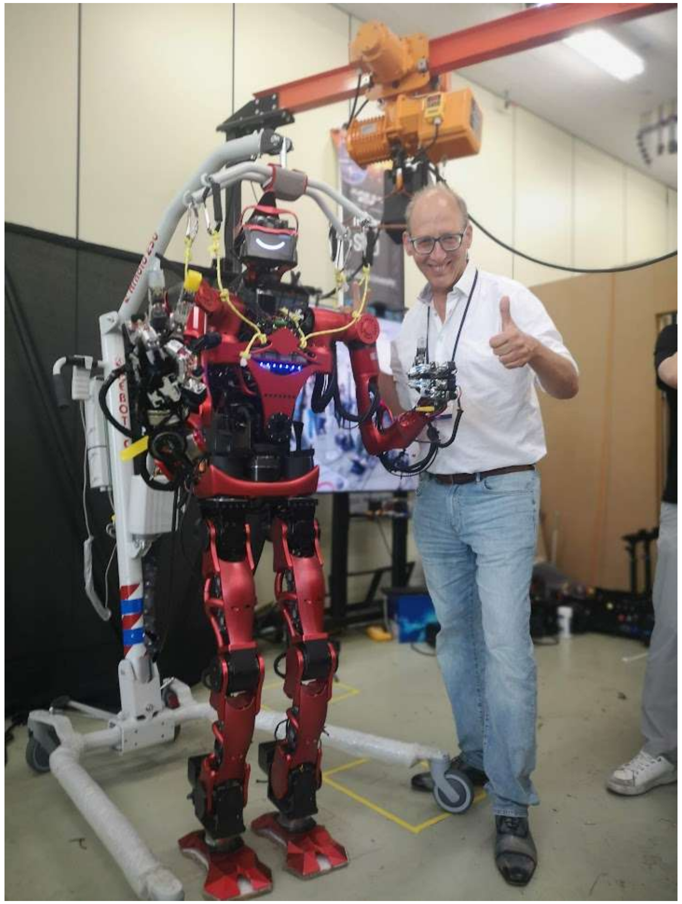
UNIVERSITIES AUSTRALIA

Pim Dijkhuizen

Bsc Informatiekunde

A little bit about Arnoud Visser

Universiteit van Amsterdam



Universiteit Leiden

Arnoud Visser

Interaction with humanoid robots

Artificial Intelligence

Predicting Damage of Dutch Road Markings

37th Benelux Conference on Artificial Intelligence

Bringing the RT-1-X Foundation Model to a SCARA robot

36th Benelux Conference on Artificial Intelligence

Learning to walk with a soft actor-critic approach

35th Benelux Conference on Artificial Intelligence

Combining Structure from Motion with visual SLAM for the Katwijk Beach

32nd Benelux Conference on Artificial Intelligence

A residual neural-network model to predict visual cortex measurements

31st Benelux Conference on Artificial Intelligence

Intelligent News Conversation with the Pepper Robot

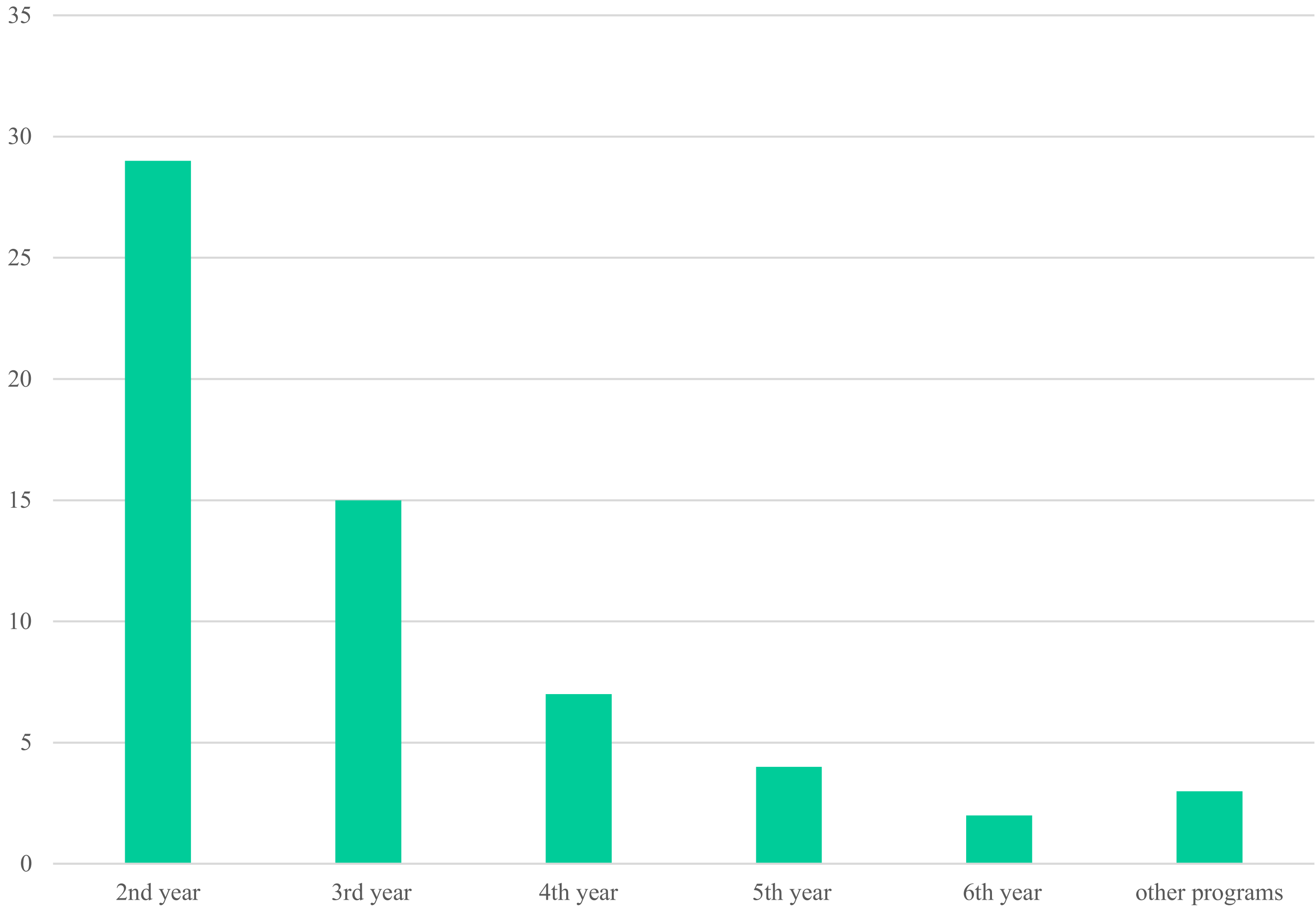
29th Benelux Conference on Artificial Intelligence

BNAIC BeNeLearn

BNVKI BAIAI

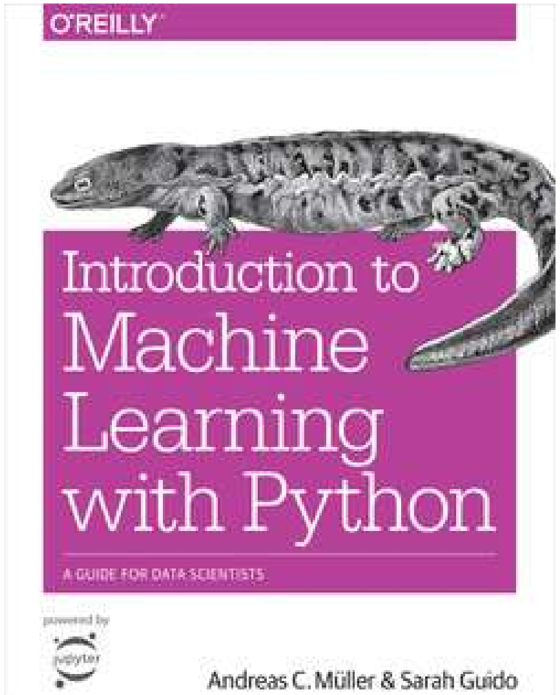
Collaborations

How about you?



Introduction to Machine Learning

“The tools introduced in this book have been applied to diverse scientific problems such as understanding stars, finding distant stars, discovering new particles, analyzing DNA sequences, and providing personalized cancer treatments”



Andreas C. Müller, Sarah Guido, Introduction to Machine Learning with Python, O’Reilly Media, October 2016

Impact of Machine Learning

https://ai100.stanford.edu/

Impact of Artificial Intelligence

100 years of AI

AI100 timeline

1950 In his famous paper Computing Machinery and Intelligence, Alan Turing posits that computer programs could think like humans and proposes a test to ascertain whether a computer's behavior is "intelligent."

1956 Stanford computer scientist John McCarthy, above, convenes the Dartmouth conference on "artificial intelligence," a term he defined. At this conference Herbert Simon and Allen Newell demonstrate a program that uses artificial intelligence to prove theorems in Principia Mathematica, by Bertram Russell and Alfred North Whitehead about logical foundations of mathematics. Simon and Newell also start work on computerized chess.

1962 Arthur Samuel, an IBM computer scientist who later became a Stanford professor, creates a self-learning program that proves capable of defeating one of America's top-ranked checkers champions.

1965-1970 Stanford researchers Ed Feigenbaum, seated above, Joshua Lederberg, Bruce Buchanan and Carl Djerassi create DENDRAL, the first "expert system." It creates scientific hypotheses about molecular structure using measured data.

1970-1980 Researchers develop more expert systems with applications to biology, medicine, engineering and the military.

1973 SRI's Artificial Intelligence Group creates Shakey the Robot, which crosses an obstacle-filled room autonomously using vision and locomotion systems. Shakey is the Computer History Museum's iconic exhibit for AI and Robotics.

1997 IBM's Deep Blue beats world chess champion Garry Kasparov in a six-game match, capping what Simon and Newell started four decades earlier.

2000 Statistical machine learning research that began in the 1980s achieves widespread practical use in major software services and mobile devices.

2005 Computer scientist Sebastian Thrun, above, and a team from the Stanford Artificial Intelligence Laboratory build a driverless car named Stanley. It becomes the first autonomous vehicle to complete a 132-mile course in the Mojave Desert, winning the DARPA Grand Challenge. Stanley is now on exhibit in the Smithsonian.

2009 Computer scientist Eric Horvitz assembles an AAAI study group on long-term AI futures, which holds its final meeting at Asilomar in California.

2011 IBM's Watson supercomputing system beats the two best human players of the TV game show Jeopardy, demonstrating an ability to understand and answer the types of nuanced questions that had previously bedeviled computer programs.

2014 Stanford accepts proposal to host One-Hundred-Year Study on Artificial Intelligence.

Stanford | One Hundred Year Study on Artificial Intelligence (AI100)

ARTIFICIAL INTELLIGENCE AND LIFE IN 2030

A forecast on the impact of AI
on daily life in western city in the year 2030

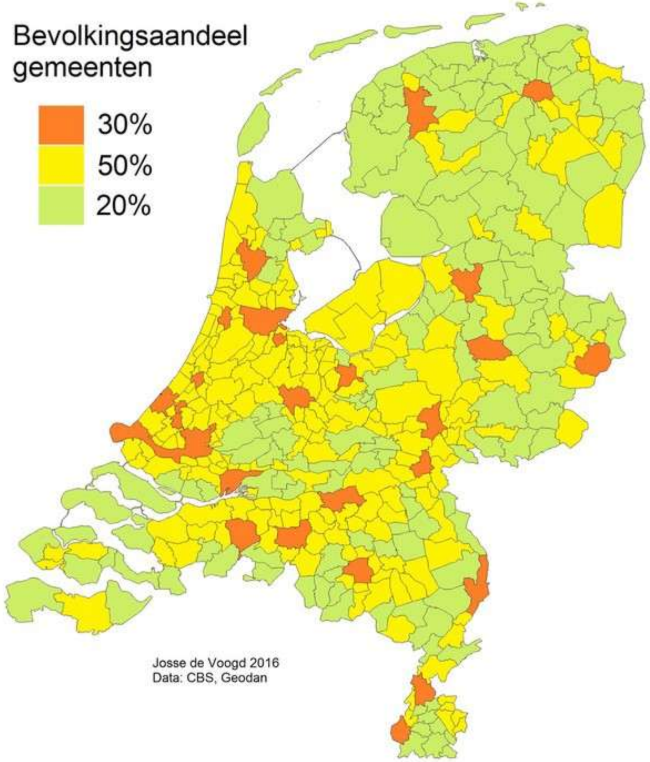
Stanford | One Hundred Year Study on Artificial Intelligence (AI100)

Move to the city

The percentage of China's population living in cities rose from 13% to 40% between 1950 and 2005. It is predicted to rise to 60% by 2030. Source: The Guardian.

Middle Town

- In the Netherlands, most people don't live in a city



One big city

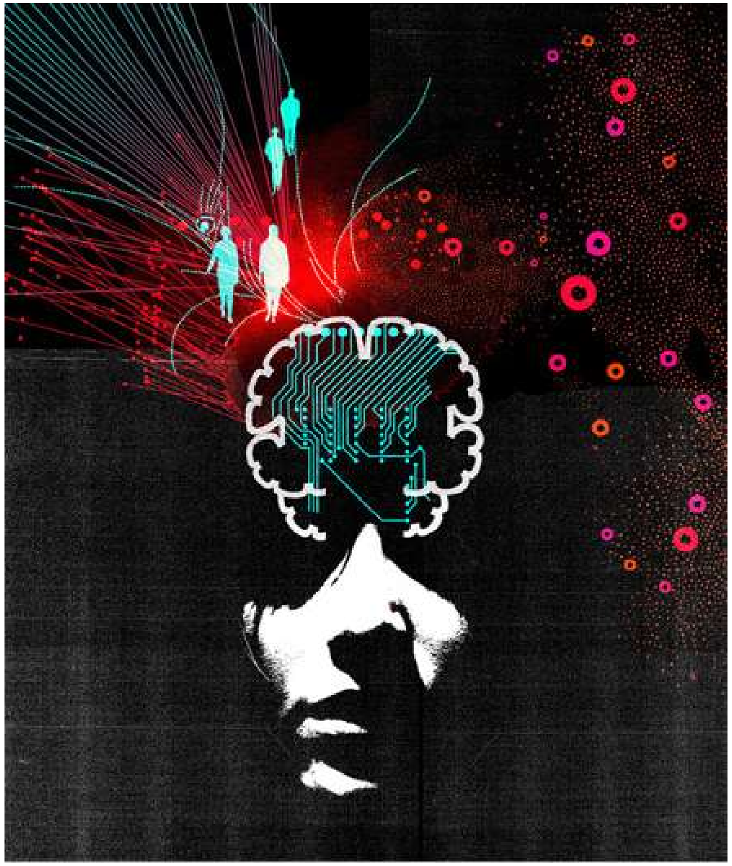
Stanford | One Hundred Year Study on Artificial Intelligence (AI100)

Eight Domains

- ☐ transportation;
- ☐ service robots;
- ☐ healthcare;
- ☐ education;
- ☐ low-resource communities;
- ☐ public safety and security;
- ☐ employment and workplace;
- ☐ entertainment

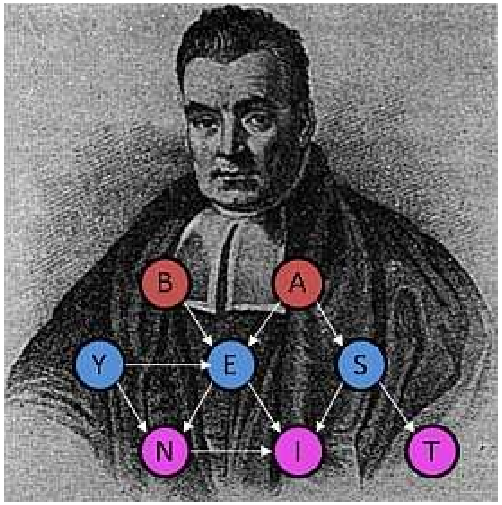
AI research trends

- Large-scale machine learning
- Deep Learning
- Reinforcement learning
- Robotics
- Computer vision
- Natural Language Processing
- Collaborative systems
- Crowdsourcing and human computation
- Algorithmic game theory and computational social choice
- Internet of Things (IoT)
- Neuromorphic Computing



The traditional paradigms of AI

- Logic-based knowledge representation and reasoning
- Planning based on modeling assumptions
- Bayesian reasoning and graphical models

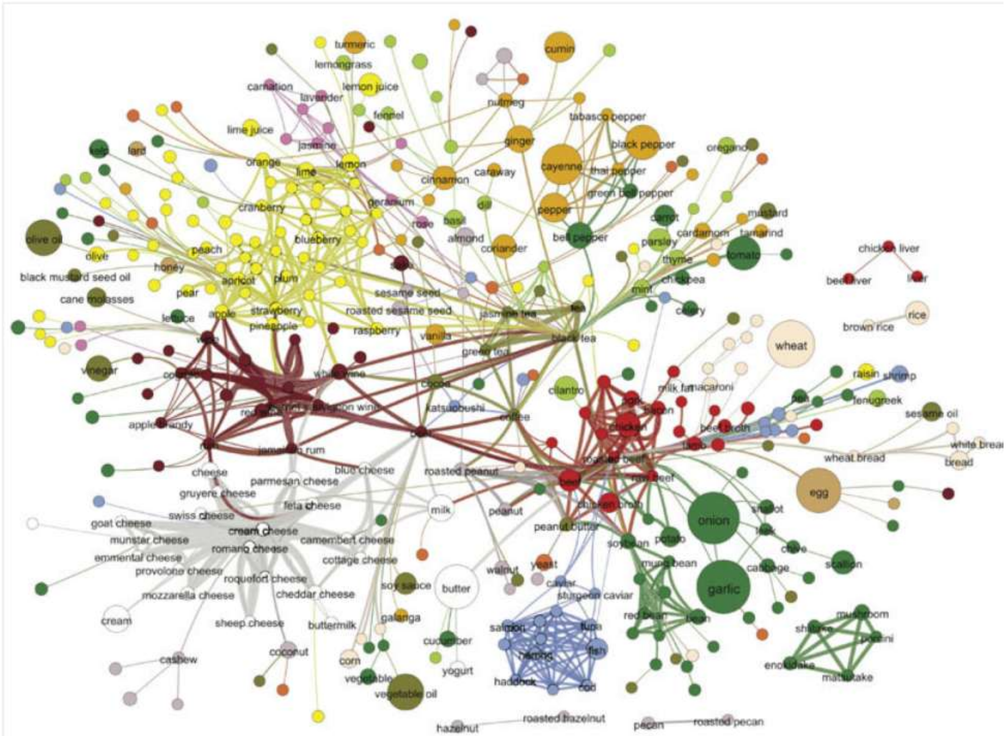


$$p(x \mid y) = \frac{p(y \mid x)\, p(x)}{p(y)} = \frac{p(y \mid x)\, p(x)}{\sum_{x'} p(y \mid x')\, p(x')} \quad \text{(discrete)}$$

$$p(x \mid y) = \frac{p(y \mid x)\, p(x)}{p(y)} = \frac{p(y \mid x)\, p(x)}{\int p(y \mid x')\, p(x')\, dx'} \quad \text{(continuous)}$$

Large-scale machine learning

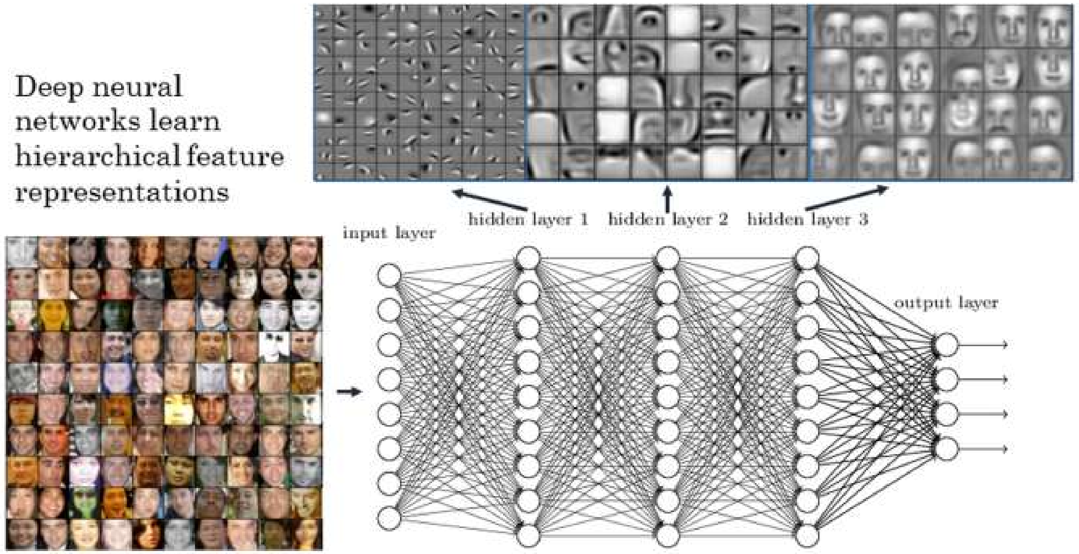
Many of the basic problems in machine learning (such as supervised and unsupervised learning) are well-understood. A major focus of current efforts is to scale existing algorithms to work with extremely large data sets. For example, whereas traditional methods could afford to make several passes over the data set, modern ones are designed to make only a single pass; in some cases, only sublinear methods (those that only look at a fraction of the data) can be admitted.



[Study of Taste]

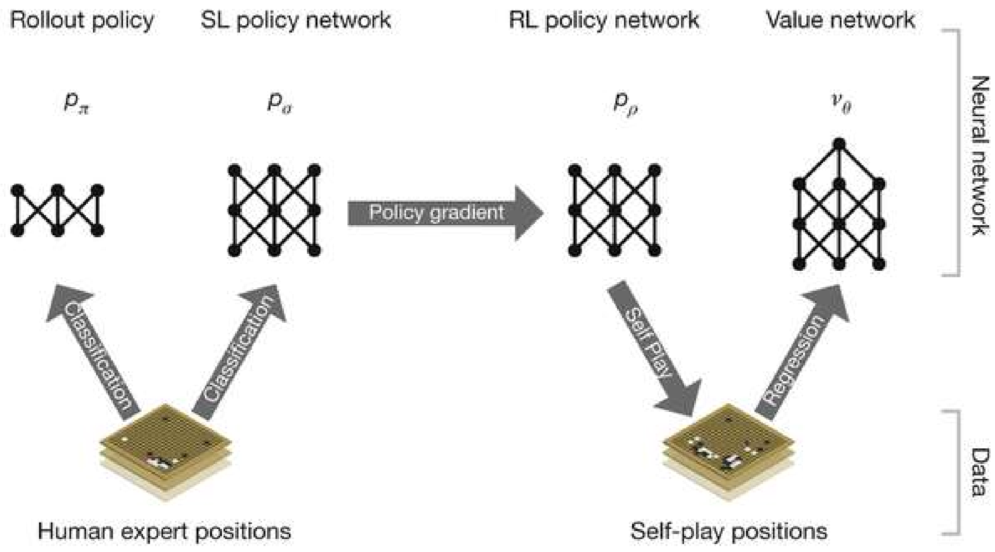
Deep Learning

The ability to successfully train convolutional neural networks has most benefited the field of computer vision, with applications such as object recognition, video labeling, activity recognition, and several variants thereof. Deep learning is also making significant inroads into other areas of perception, such as audio, speech, and natural language processing.



Reinforcement learning

Whereas traditional machine learning has mostly focused on pattern mining, reinforcement learning shifts the focus to decision making, and is a technology that will help AI to advance more deeply into the realm of learning about and executing actions in the real world. It has existed for several decades as a framework for experience-driven sequential decision-making, but the methods have not found great success in practice, mainly owing to issues of representation and scaling. However, the advent of deep learning has provided reinforcement learning with a "shot in the arm." The recent success of AlphaGo, a computer program developed by Google Deepmind that beat the human Go champion in a five-game match, was due in large part to reinforcement learning. AlphaGo was trained by initializing an automated agent with a human expert database, but was subsequently refined by playing a large number of games against itself and applying reinforcement learning.

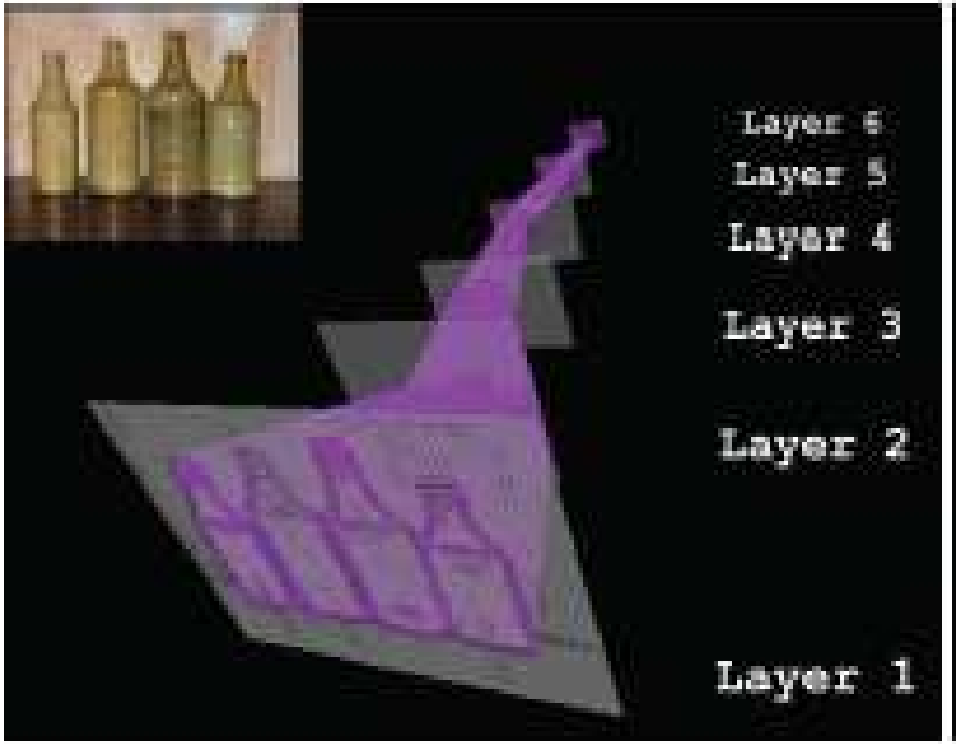


Robotics

Robotic navigation, at least in static environments, is largely solved. Current efforts consider how to train a robot to interact with the world around it in generalizable and predictable ways. A natural requirement that arises in interactive environments is manipulation, another topic of current interest. The deep learning revolution is only beginning to influence robotics, in large part because it is far more difficult to acquire the large labeled data sets that have driven other learning-based areas of AI. Reinforcement learning, which obviates the requirement of labeled data, may help bridge this gap but requires systems to be able to safely explore a policy space without committing errors that harm the system itself or others. Advances in reliable machine perception, including computer vision, force, and tactile perception, much of which will be driven by machine learning, will continue to be key enablers to advancing the capabilities of robotics.

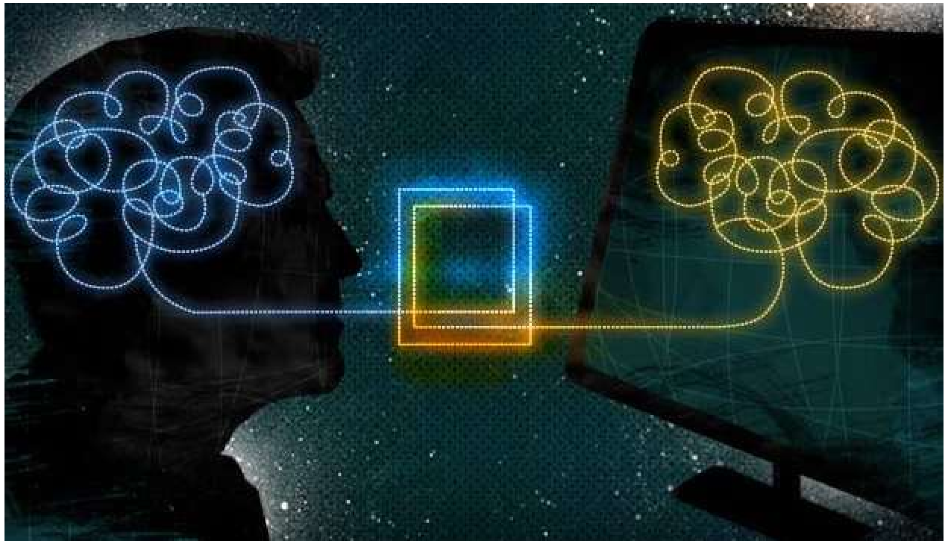
Computer vision

Computer vision is currently the most prominent form of machine perception. It has been the sub-area of AI most transformed by the rise of deep learning. Until just a few years ago, support vector machines were the method of choice for most visual classification tasks. But the confluence of large-scale computing, especially on GPUs, the availability of large datasets, especially via the internet, and refinements of neural network algorithms has led to dramatic improvements in performance on benchmark tasks (e.g., classification on ImageNet). For the first time, computers are able to perform some (narrowly defined) visual classification tasks better than people. Much current research is focused on automatic image and video captioning.



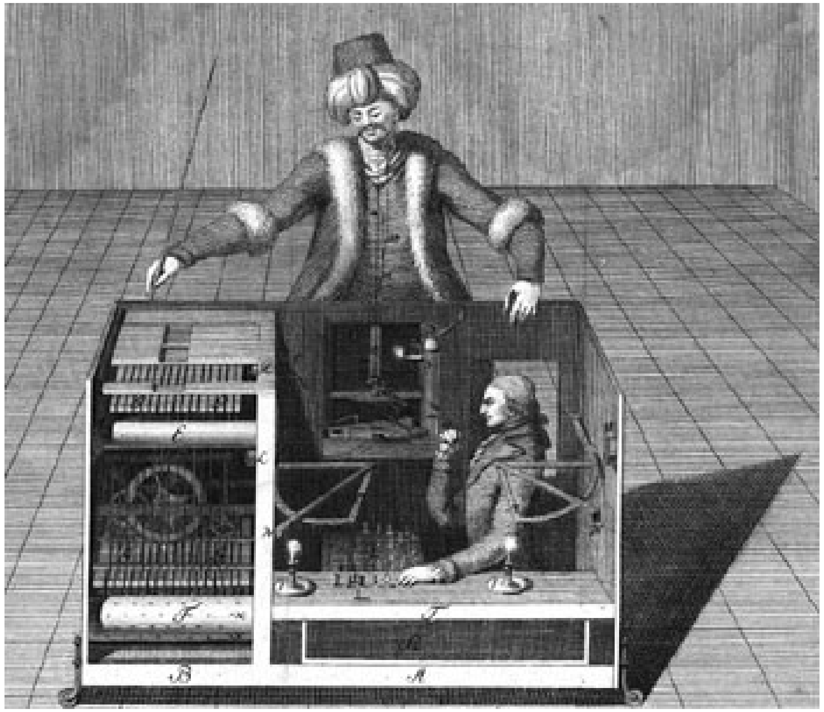
Natural Language Processing

Often coupled with automatic speech recognition, Natural Language Processing is another very active area of machine perception. It is quickly becoming a commodity for mainstream languages with large data sets. Google announced that 20% of current mobile queries are done by voice,18 and recent demonstrations have proven the possibility of real-time translation. Research is now shifting towards developing refined and capable systems that are able to interact with people through dialog, not just react to stylized requests.



Crowdsourcing and human computation

- Since human abilities are superior to automated methods for accomplishing many tasks, research on crowdsourcing and human computation investigates methods to augment computer systems by utilizing human intelligence to solve problems that computers alone cannot solve well. Introduced only about fifteen years ago, this research now has an established presence in AI. The best-known example of crowdsourcing is Wikipedia, a knowledge repository that is maintained and updated by netizens and that far exceeds traditionally-compiled information sources, such as encyclopedias and dictionaries, in scale and depth. Crowdsourcing focuses on devising innovative ways to harness human intelligence. Citizen science platforms energize volunteers to solve scientific problems, while paid crowdsourcing platforms such as Amazon Mechanical Turk provide automated access to human intelligence on demand. Work in this area has facilitated advances in other subfields of AI, including computer vision and NLP, by enabling large amounts of labeled training data and/or human interaction data to be collected in a short amount of time. Current research efforts explore ideal divisions of tasks between humans and machines based on their differing capabilities and costs.



Algorithmic game theory and computational social choice

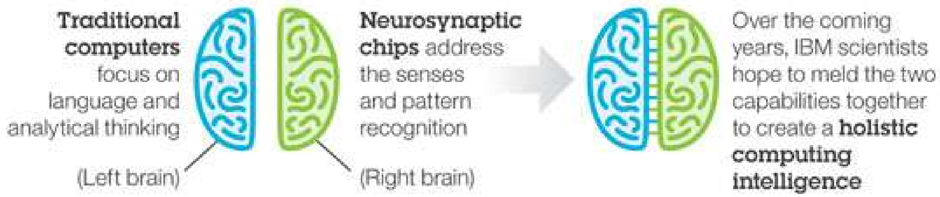
New attention is being drawn to the economic and social computing dimensions of AI, including incentive structures. Distributed AI and multi-agent systems have been studied since the early 1980s, gained prominence starting in the late 1990s, and were accelerated by the internet. A natural requirement is that systems handle potentially misaligned incentives, including self-interested human participants or firms, as well as automated AI-based agents representing them. Topics receiving attention include computational mechanism design (an economic theory of incentive design, seeking incentive-compatible systems where inputs are truthfully reported), computational social choice (a theory for how to aggregate rank orders on alternatives), incentive aligned information elicitation (prediction markets, scoring rules, peer prediction) and algorithmic game theory (the equilibria of markets, network games, and parlor games such as Poker—a game where significant advances have been made in recent years through abstraction techniques and no-regret learning).

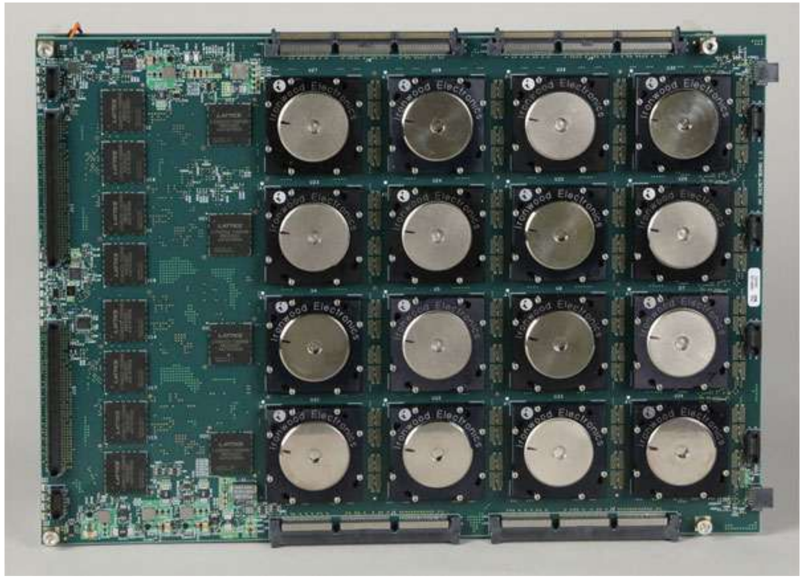
Internet of Things (IoT)

A growing body of research is devoted to the idea that a wide array of devices can be interconnected to collect and share their sensory information. Such devices can include appliances, vehicles, buildings, cameras, and other things. While it's a matter of technology and wireless networking to connect the devices, AI can process and use the resulting huge amounts of data for intelligent and useful purposes. Currently, these devices use a bewildering array of incompatible communication protocols. AI could help tame this Tower of Babel.

Neuromorphic Computing

Traditional computers implement the von Neumann model of computing, which separates the modules for input/output, instruction-processing, and memory. With the success of deep neural networks on a wide array of tasks, manufacturers are actively pursuing alternative models of computing—especially those that are inspired by what is known about biological neural networks—with the aim of improving the hardware efficiency and robustness of computing systems. At the moment, such "neuromorphic" computers have not yet clearly demonstrated big wins, and are just beginning to become commercially viable. But it is possible that they will become commonplace (even if only as additions to their von Neumann cousins) in the near future. Deep neural networks have already created a splash in the application landscape. A larger wave may hit when these networks can be trained and executed on dedicated neuromorphic hardware, as opposed to simulated on standard von Neumann architectures, as they are today.





Conclusion

AI effects nearly every aspect in city life!

Stanford | One Hundred Year Study on Artificial Intelligence (AI100)

AI for Good

How does it affect low-resource communities?

Stanford | One Hundred Year Study on Artificial Intelligence (AI100)

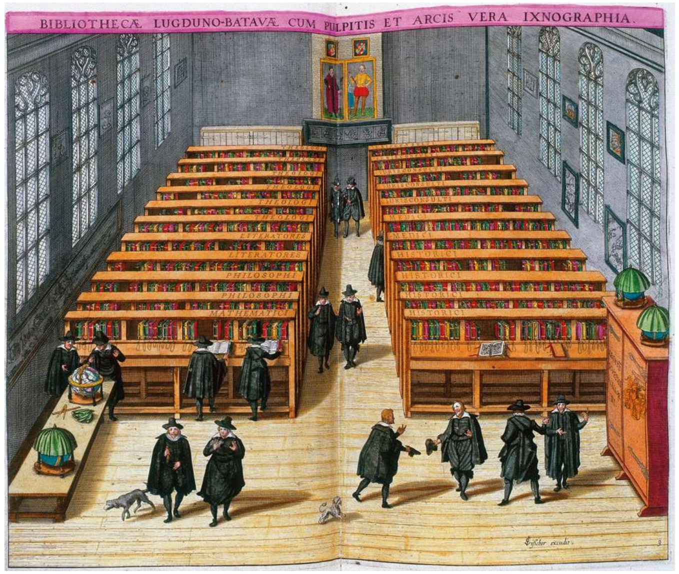
Goals for the Course

- Insight in the mathematical foundation of the techniques and algorithms applied in the field
- Experience with selecting the right model for the right problem
- Practical experience with applying the techniques to a "real data"

Machine Leaning vs Data Science

Literature

Andreas C. Müller, Sarah Guido, [Introduction to Machine Learning with Python](), O'Reilly Media, October 2016



Universiteit Leiden

Universiteitsbibliotheek

Universiteit van Amsterdam

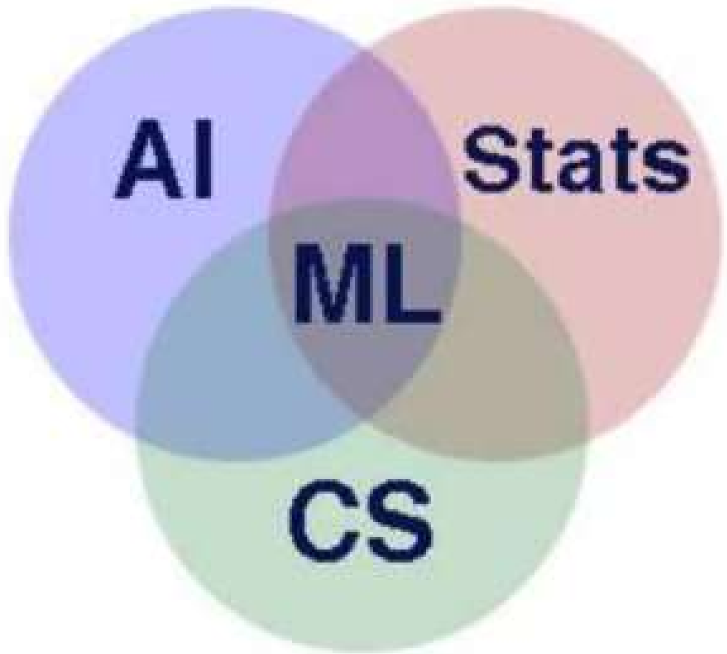
Artis bibliotheek

Topics covered in the course

- Fundamental concepts and applications of machine learning
- Advantages and shortcomings of widely used algorithms
- Data aspects of machine learning
- Evaluation and training methods
- Pipelines to streamline your work

What is machine learning?

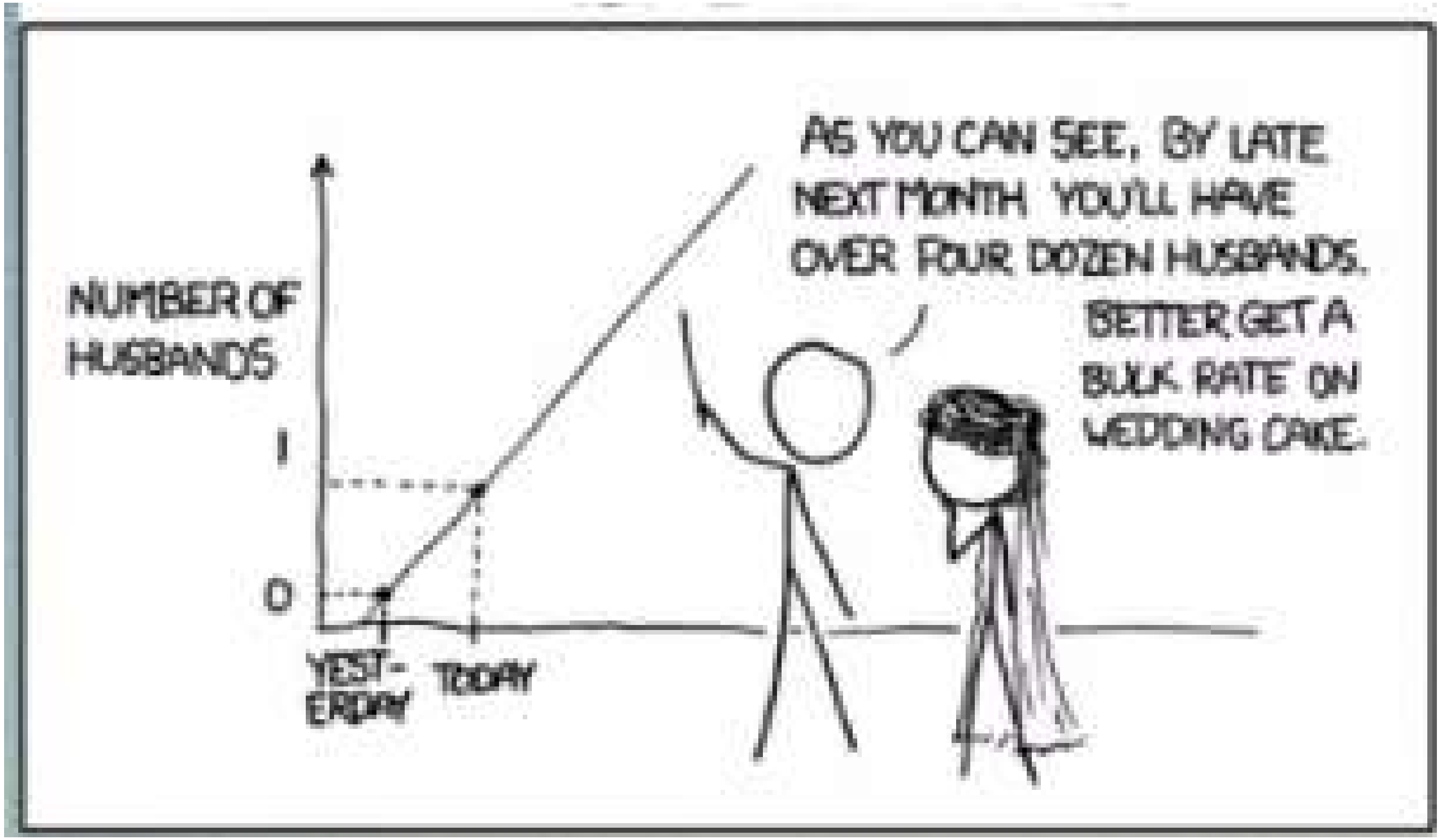
- Learn to find correlations in data



- Estimates new data points in between observed data points and predict beyond examples seen in training data

Slide idea from Sarah Guido

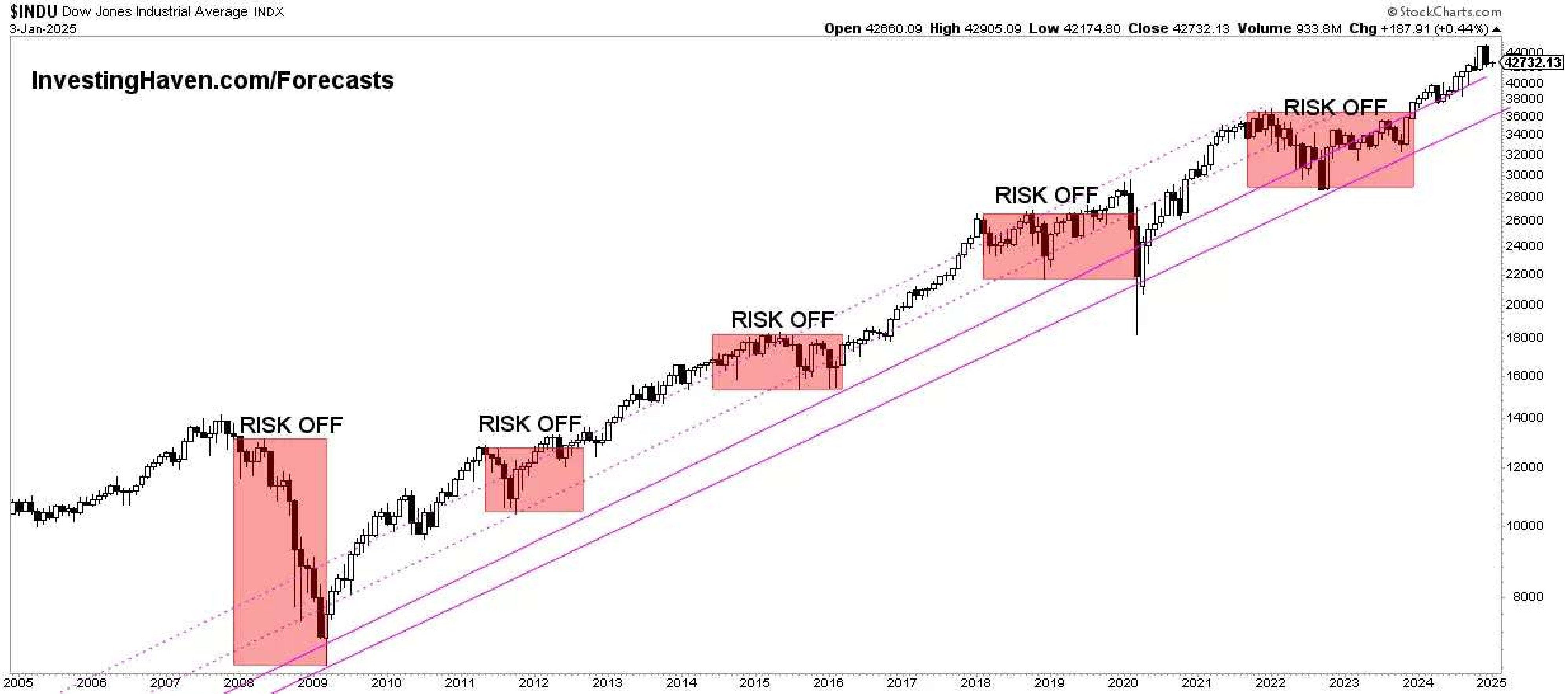
Extrapolation



- Difficult to predict beyond examples seen in training data

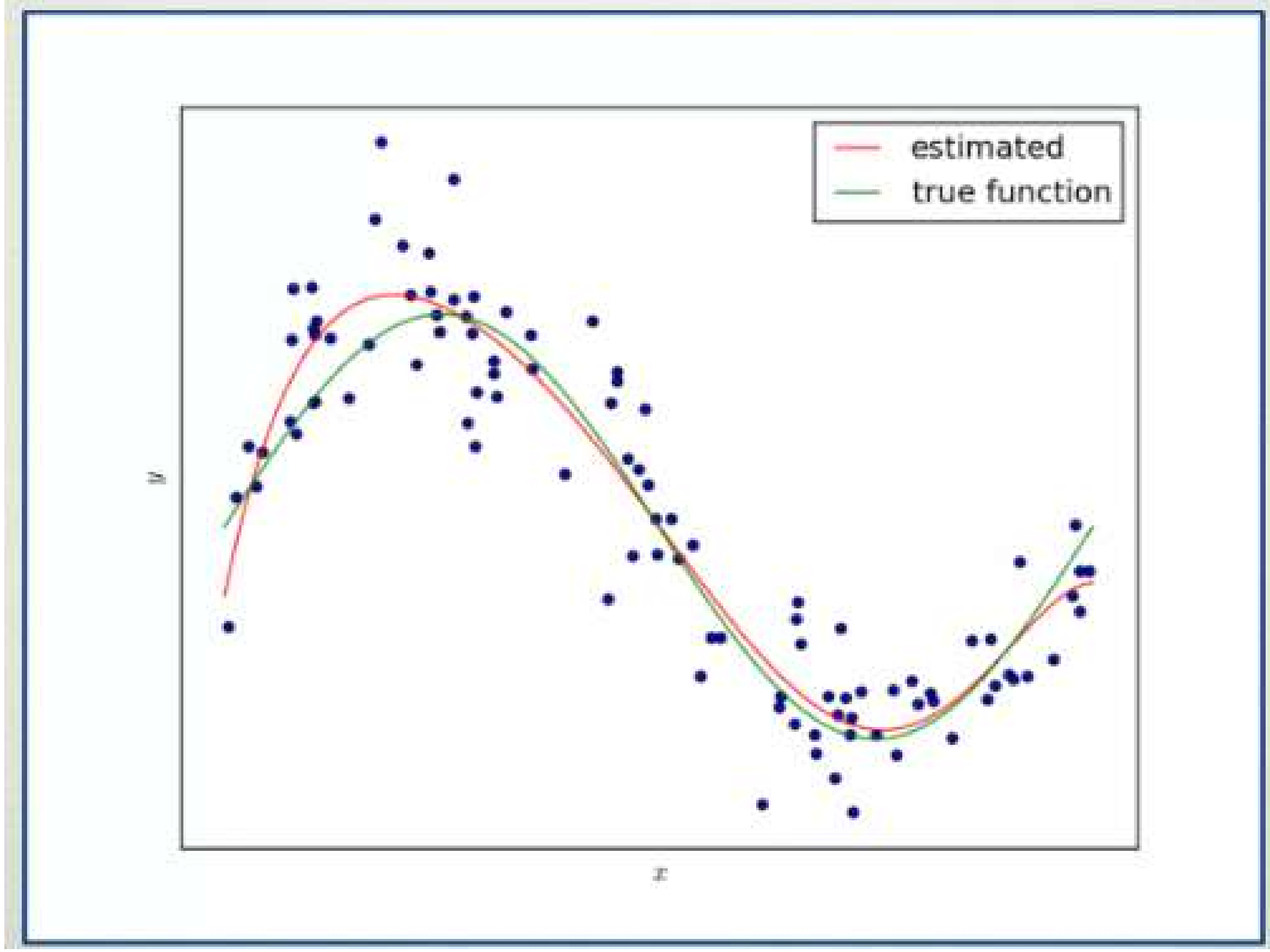
Slide idea from Benjamin Bengfort

Extrapolation



- Difficult to predict beyond examples seen in training data

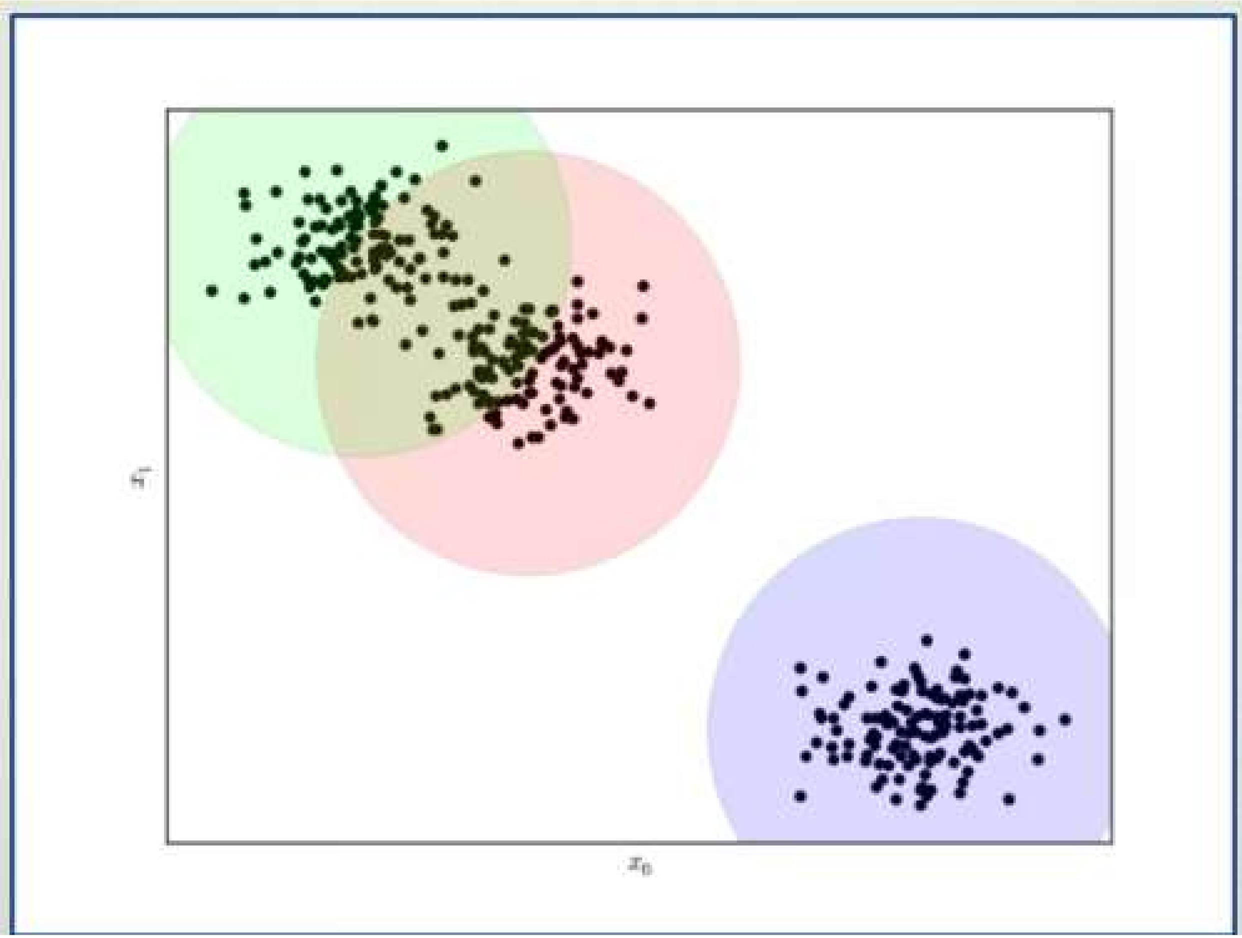
Regression



- Given training data predict the continues values in between – the function between input and output

Slide idea from Benjamin Bengfort

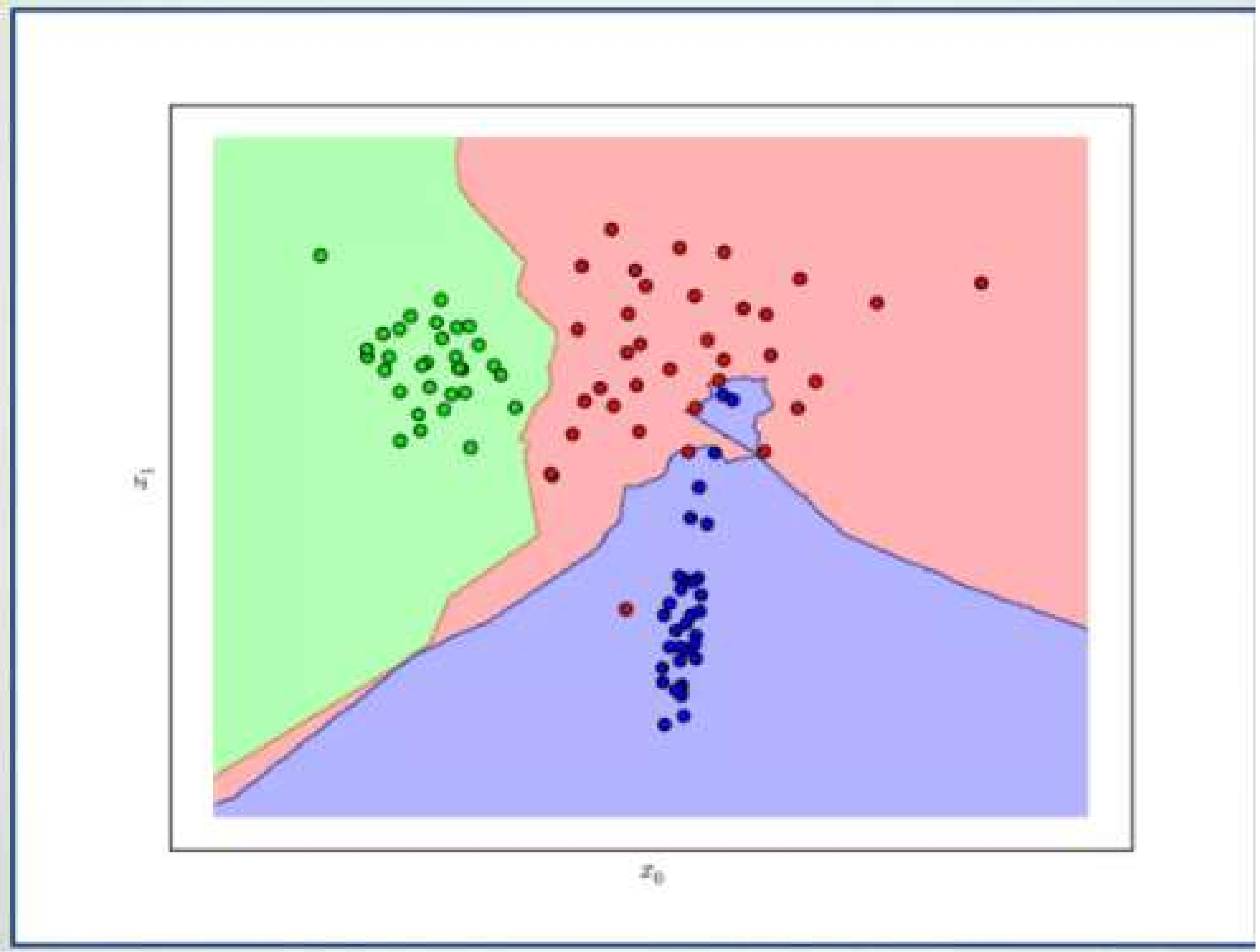
Clustering



- Given training data, determine a pattern of associated data points via their simularity of distance from each other.

Slide idea from Benjamin Bengfort

Classification



- Given training data, fit a function that can determine for any input, what the label is.

Slide idea from Benjamin Bengfort

Features

Feature

	buying	maint	doors	persons	lug_boot	safety
0	vhigh	vhigh	2	2	small	low
1	vhigh	vhigh	2	2	small	med
2	vhigh	vhigh	2	2	small	high
3	vhigh	vhigh	2	2	med	low
4	vhigh	vhigh	2	2	med	med

Class

	class
0	unacc
1	unacc
2	unacc
3	unacc
4	unacc

Instance

- the class label of an instance is based on its features.

Slide idea from Sharon Guido

Scikit-Learn

- Model = EstimatorObject()
- Model.fit(dataset.data, dataset.target)
  - dataset.data = dataset $\leftarrow X$
  - dataset.target = labels $\leftarrow y$
- Model.predict(dataset.data)

Class

	class
0	unacc
1	unacc
2	unacc
3	unacc
4	unacc

vector ↓

Matrix ↓

$$y = f(X)$$

↓ output

↑ input

Feature

	buying	maint	doors	persons	lug_boot	safety
0	vhigh	vhigh	2	2	small	low
1	vhigh	vhigh	2	2	small	med
2	vhigh	vhigh	2	2	small	high
3	vhigh	vhigh	2	2	med	low
4	vhigh	vhigh	2	2	med	med

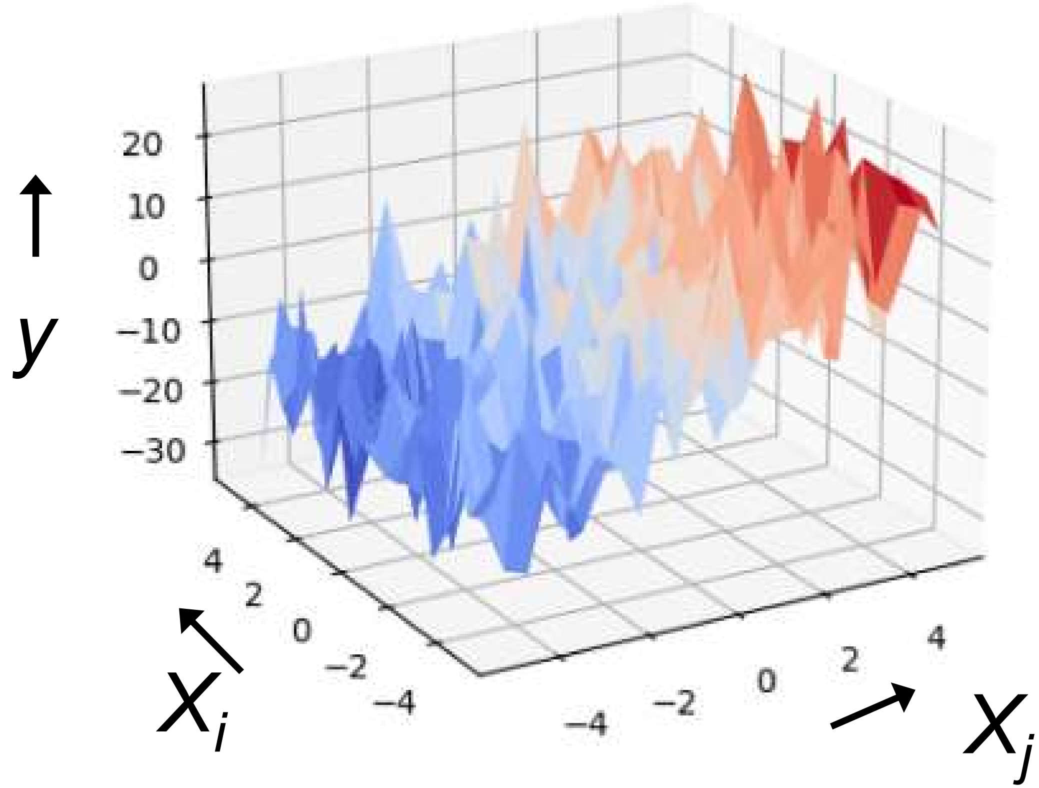
Instance

Slide idea from Sharon Guido

Scikit-Learn

$y \longleftarrow$ Model.predict(X)

$$y = f(X)$$



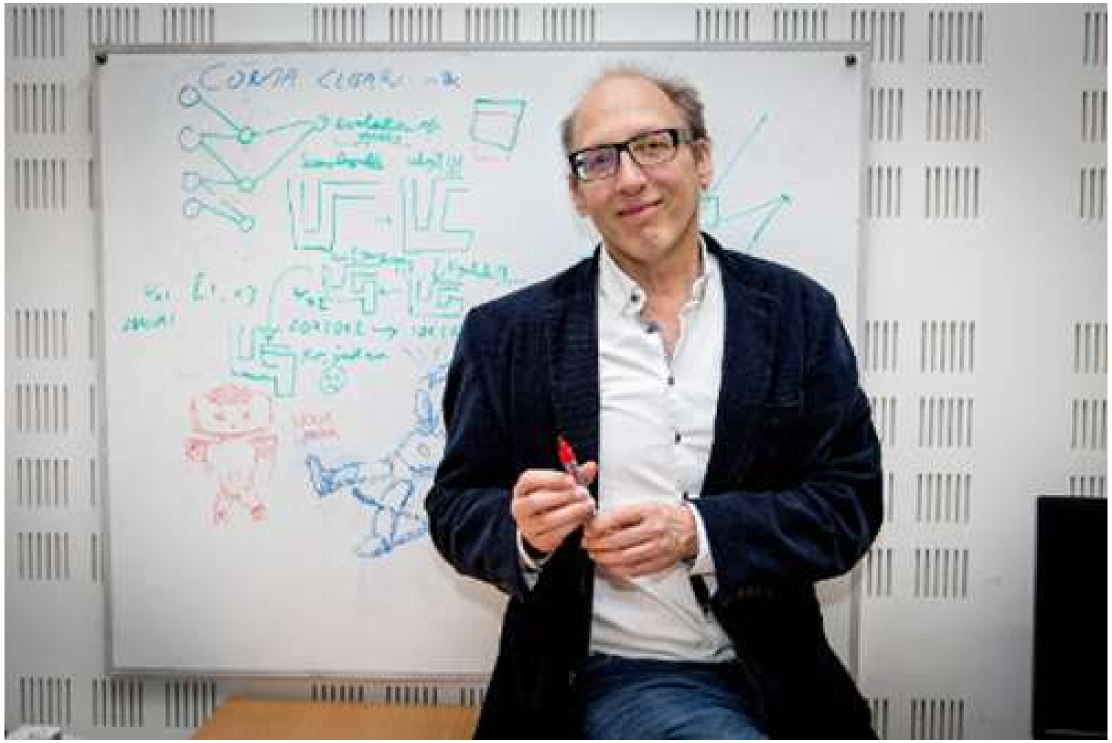
Schedule

- Week 1: Python/Pandas/Numpy prerequisites + Data preprocessing
- Week 2: Regression I: KNN + Linear Regression + Metrics
- Week 3: Regression II: Linear regression with polynomial features + under/overfitting + regularization

- Week 5: Classification I: KNN + Logistic Regression + Metrics
- Week 6: Classification II: Trees/RF + SVMs + Grid search
- Week 7: Clustering: K-means + DBSCAN

Plus much more subjects!

What’s new since 2025

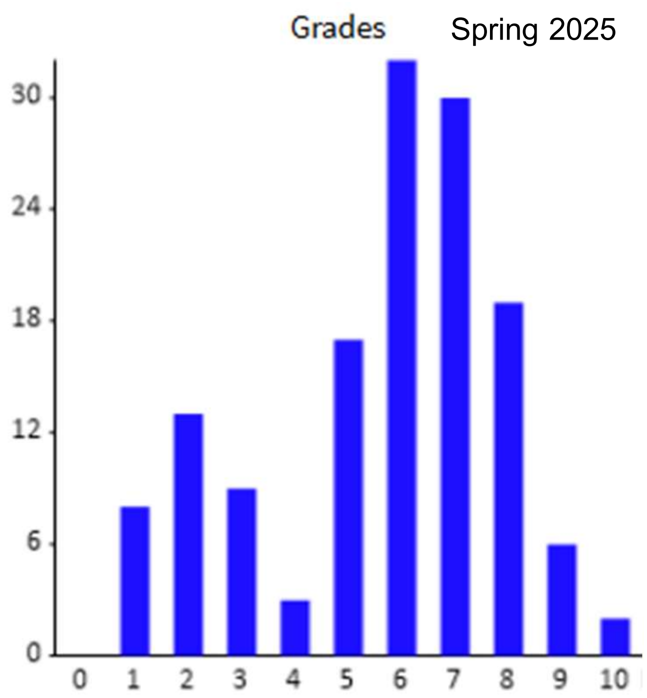


- New lecturer
- Tutorials to head-start your homework
- Direct feedback via CodeGrade

- Quizzes during the workgroups

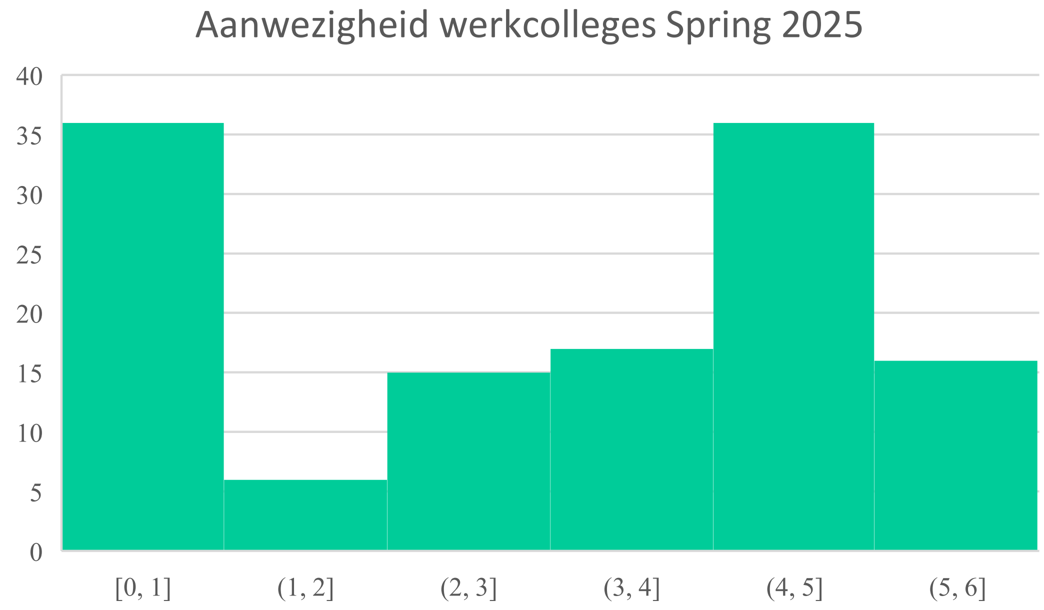
Grading

- 5% Quizzes (bonus)
- 10% Homework
- 40% Digital exam I
- 50% Digital exam II ← Will also cover the material of the first half



Attendance Workgroups

- 5% Quizzes
- 10% Homework



4,97 ± 2,86

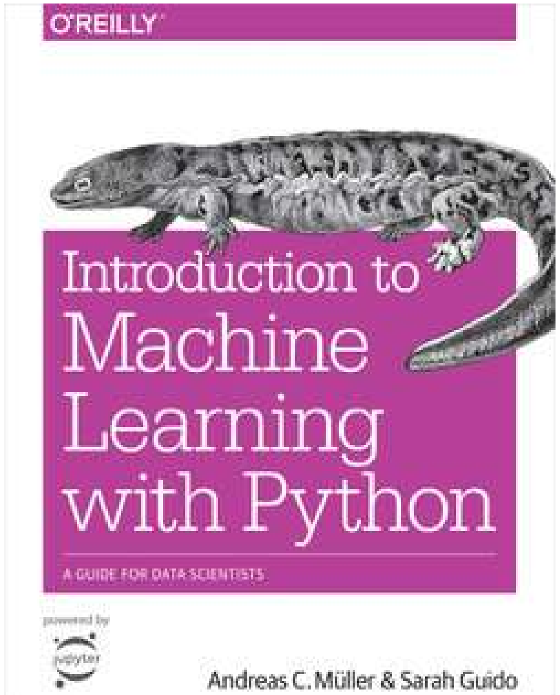
5,85 ± 2,27

Who to ask?

- Lectures & Exam: Arnoud (a.visser@uva.nl)
- Practicals: Rein (r.lukkes@uva.nl)

Introduction to Machine Learning

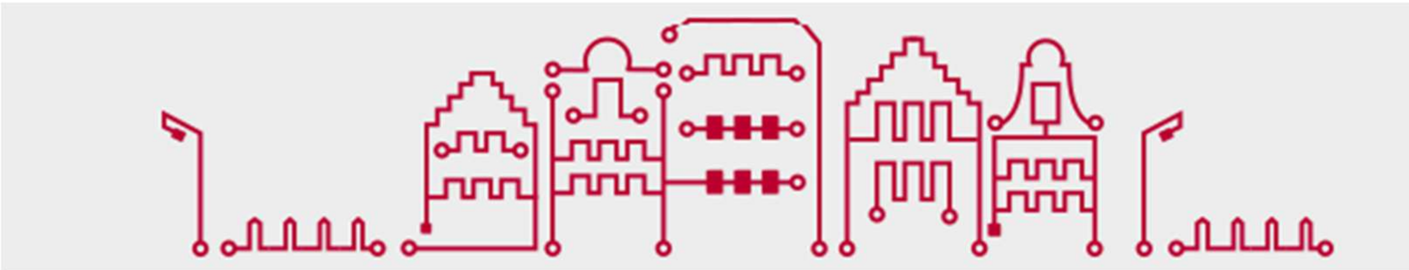
- 1.1 Why Machine Learning
- 1.2 Why Python
- 1.3 Why Scikit-Learn



Andreas C. Müller, Sarah Guido, [Introduction to Machine Learning with Python](), O'Reilly Media, October 2016

Conclusion

- AI effects nearly every aspect in life



- It is about patterns in data, Data, DATA

- The functions & boundaries that you define
  can have real impact!

  (growth / decline) (rich / poor) (life / dead)