COMPUTATIONAL INTELLIGENCE

(INTRODUCTION TO MACHINE LEARNING)

SS 16

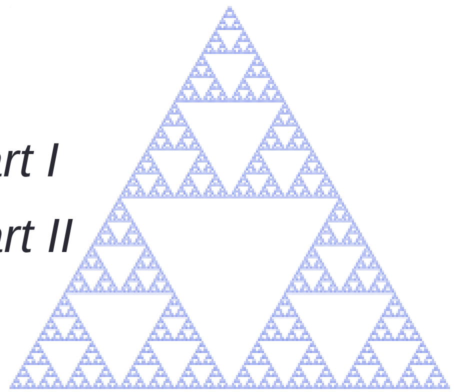
2 VO 442.070 + 1 UE 708.070

Institute for Theoretical Computer Science (IGI)
TU Graz, Inffeldgasse 16b / first floor
www.igi.tugraz.at

Institute for Signal Processing and Speech Communication (SPSC)
TU Graz, Inffeldgasse 16c / ground floor
www.spsc.tugraz.at

Organization

- Lecture / VO:
  - Tuesday, 11:00, HS i13
    - Anand Subramoney and Guillaume Bellec (IGI) *Part I*
    - Assoc. Prof. Dr. Franz Pernkopf (SPSC) *Part II*



- Practical / UE:
  - First practical on **Friday 4th of March**, 12:30, HS i11
    - Anand Subramoney and Guillaume Bellec (IGI) *Part I*
    - Dipl.-Ing. Christian Knoll (SPSC) *Part II*
  - Homework in **teams of up to 3** (use newsgroup to form teams)

- Website: http://www.spsc.tugraz.at/courses/computational-intelligence
- Newsgroup: tu-graz.lv.ci

Organization

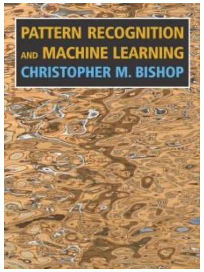
- Office hours:
  - Both Anand and Guillaume:
    - Time: Every Wednesday 16:00 – 17:00
    - Place: Our offices at Inffeldgasse 16b/1
- Exam:
  - Written exam for this year's course:
    - 08.07.2016
  - Exam has two parts: IGI (first half of semester) + SPSC (second half)
  - Language: English
  - Positive grade if positive on **both parts**!

Materials (for IGI part)

- No textbook required
- Lecture slides and further reading on **course website**

- Materials for further study:
  - Online Machine Learning course
    by Andrew Ng (Stanford):
    https://www.coursera.org/course/ml

  - Book by C. Bishop, Pattern Recognition and Machine Learning, Springer 2007.



- For SPSC part (second half):
  - Announced by Franz Pernkopf

Acknowledgments

- IGI Slides based on material from Stefan Häusler (IGI), Zeno Jonke (IGI), David Sontag (NYU), Andrew Ng (Stanford), Xiaoli Fern (Oregon State)

INTRODUCTION + MOTIVATION

Machine Learning

- Grew out of **Artificial Intelligence**

What is Artificial Intelligence?

"The exciting new effort to make computers think . . . *machines with minds,* in the full and literal sense" (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	"The study of mental faculties through the use of computational models" (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act" (Winston, 1992)
"The art of creating machines that perform functions that require intelligence when performed by people" (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better" (Rich and Knight, 1991)	"A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes" (Schalkoff, 1990) "The branch of computer science that is concerned with the automation of intelligent behavior" (Luger and Stubblefield, 1993)

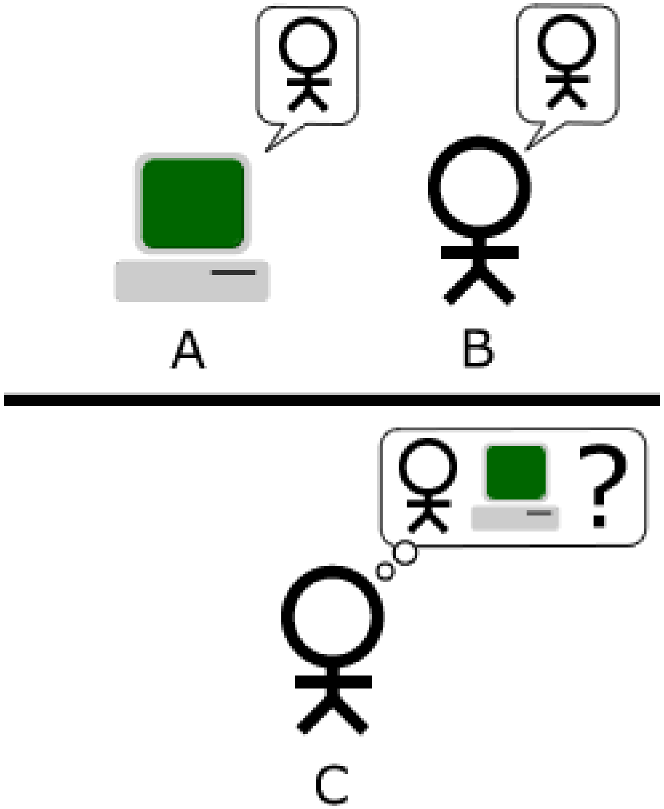
Figure 1.1 Some definitions of AI. They are organized into four categories:

Systems that think like humans.	Systems that think rationally.
Systems that act like humans.	Systems that act rationally.

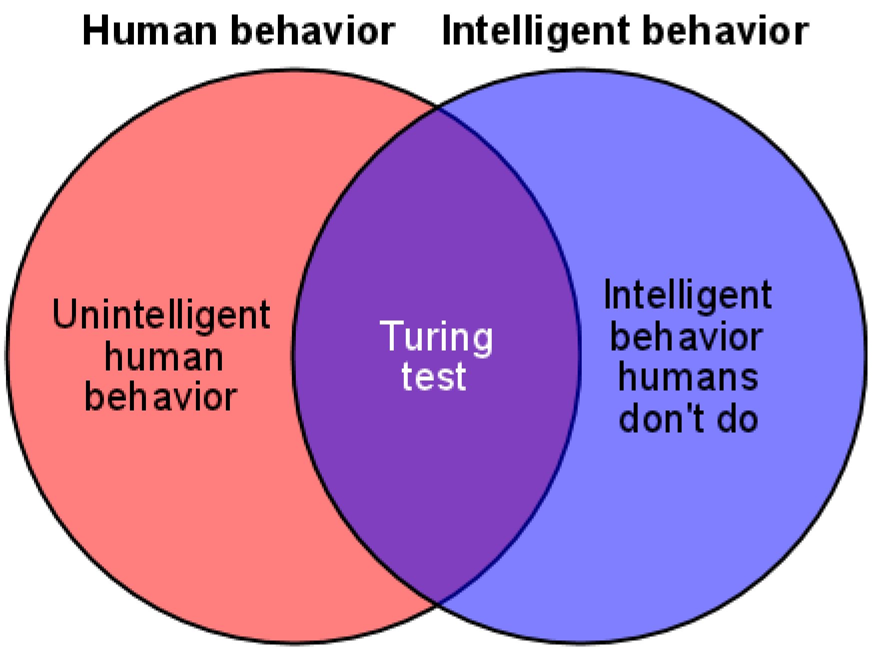
Source -- Artificial Intelligence: A Modern Approach by Stuart Russell and Peter Norvig

But what is really AI?

Turing test



Turing test



AI – “You’ll know it when you see it”

Components of AI

- *Natural language processing*
- Knowledge representation
- Automated reasoning
- **Machine learning**
- *Computer vision*
- *Robotics*

-- Russel and Norvig

Machine Learning

- Grew out of **Artificial Intelligence**
- The ability to “adapt to new circumstances and to detect and extrapolate patterns” – Russel and Norvig
- Arthur Samuel (1959). “Field of study that gives computers the ability to learn without being explicitly programmed.”

When do we need computers to learn?

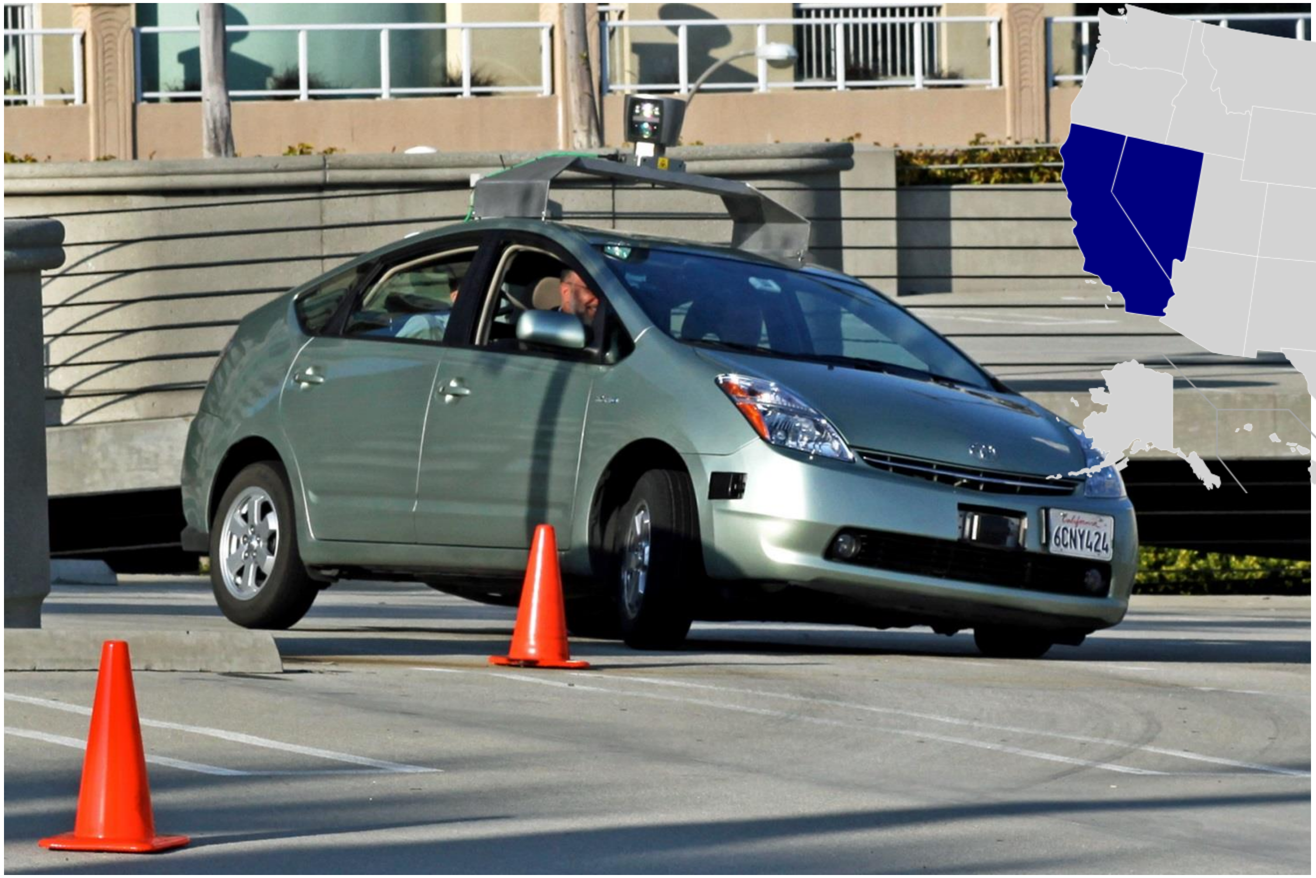
- When human expert knowledge is missing
  - For example, predicting whether some new substance could be an effective treatment for a disease
- When humans can only do it “intuitively”
  - Flying a helicopter
  - Recognize visual objects
  - Natural language processing
- When we need to learn about something that changes frequently
  - Stock market analysis
  - Weather forecasting
  - Computer network routing
- Customized learning
  - Spam filters, movie/product recommendations

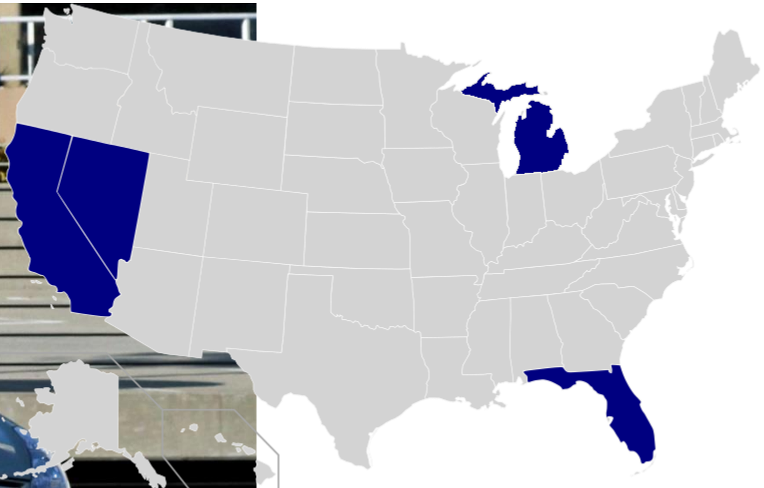
Applications of Machine learning

- Machine learning is used in a wide range of fields including:
  - Bio-informatics
  - Brain-Machine interfaces
  - Computational finance
  - Game playing
  - Information Retrieval
  - Internet fraud detection
  - Medical diagnosis
  - Natural language processing
  - Online advertising
  - Recommender systems
  - Robot locomotion
  - Search engines
  - Sentiment analysis
  - Software engineering
  - Speech and handwriting recognition
  - Stock market analysis
  - Economics and Finance
  - Credit card fraud detection
  - …….

Google car

- https://www.youtube.com/watch?v=TsaES--OTzM





+ UK, France, Switzerland

ATLAS (Boston Dynamics / Alphabet)

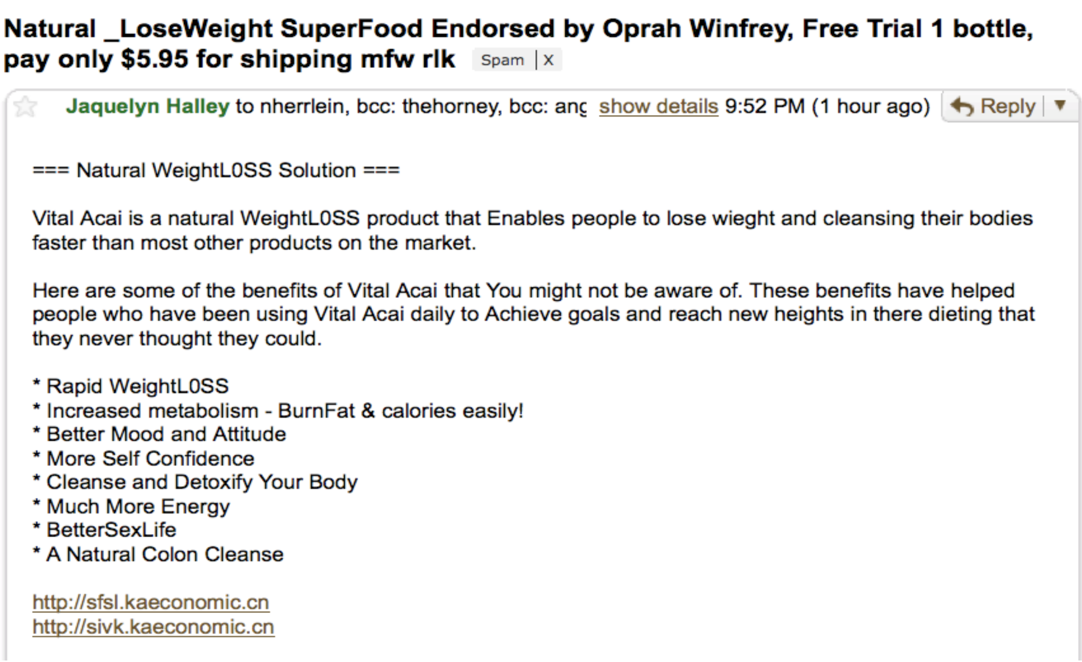
- https://www.youtube.com/watch?v=rVlhMGQgDkY

http://spectrum.ieee.org/automaton/robotics/humanoids/boston-dynamics-marc-raibert-on-nextgen-atlas

Spam filtering

- "Spam in email started to become a problem when the Internet was opened up to the general public in the mid-1990s. It grew exponentially over the following years, and today composes some 80 to 85% of all the email in the world, by a "conservative estimate".
- Source: http://en.wikipedia.org/wiki/Spamming

data


Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only $5.95 for shipping mfw rlk Spam | X

Jaquelyn Halley to nherrlein, bcc: thehorney, bcc: ang show details 9:52 PM (1 hour ago) Reply

=== Natural WeightL0SS Solution ===

Vital Acai is a natural WeightL0SS product that Enables people to lose wieght and cleansing their bodies faster than most other products on the market.

Here are some of the benefits of Vital Acai that You might not be aware of. These benefits have helped people who have been using Vital Acai daily to Achieve goals and reach new heights in there dieting that they never thought they could.

* Rapid WeightL0SS
* Increased metabolism - BurnFat & calories easily!
* Better Mood and Attitude
* More Self Confidence
* Cleanse and Detoxify Your Body
* Much More Energy
* BetterSexLife
* A Natural Colon Cleanse

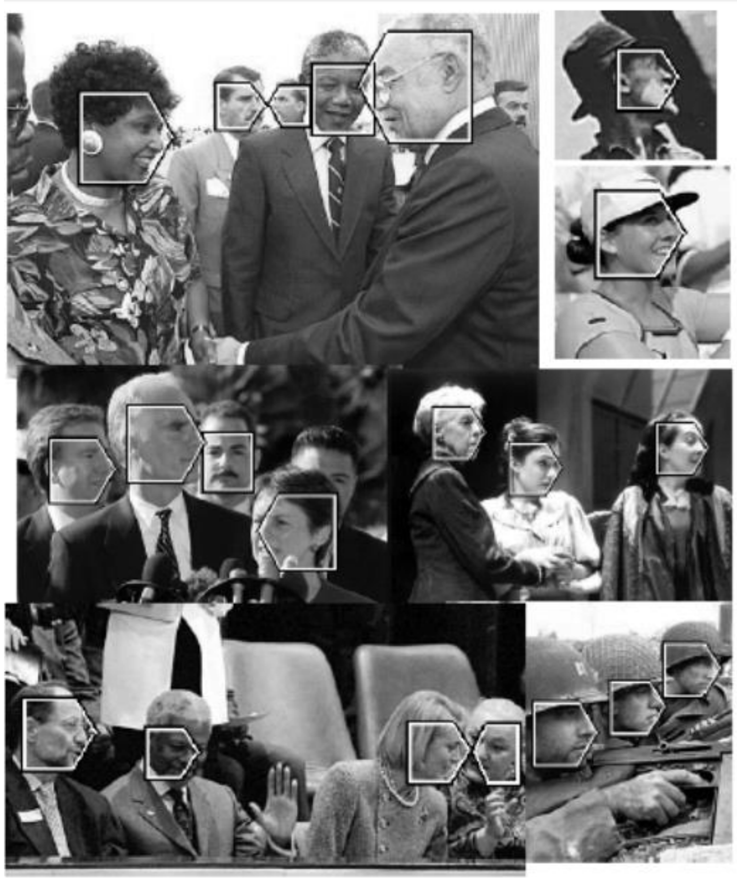
http://sfsl.kaeconomic.cn
http://sivk.kaeconomic.cn

prediction

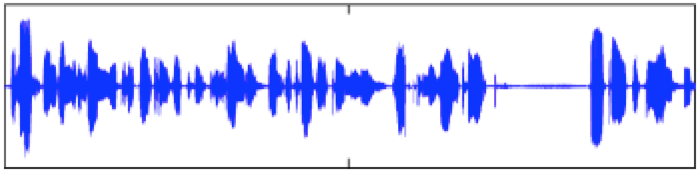
Spam
vs.
Not Spam

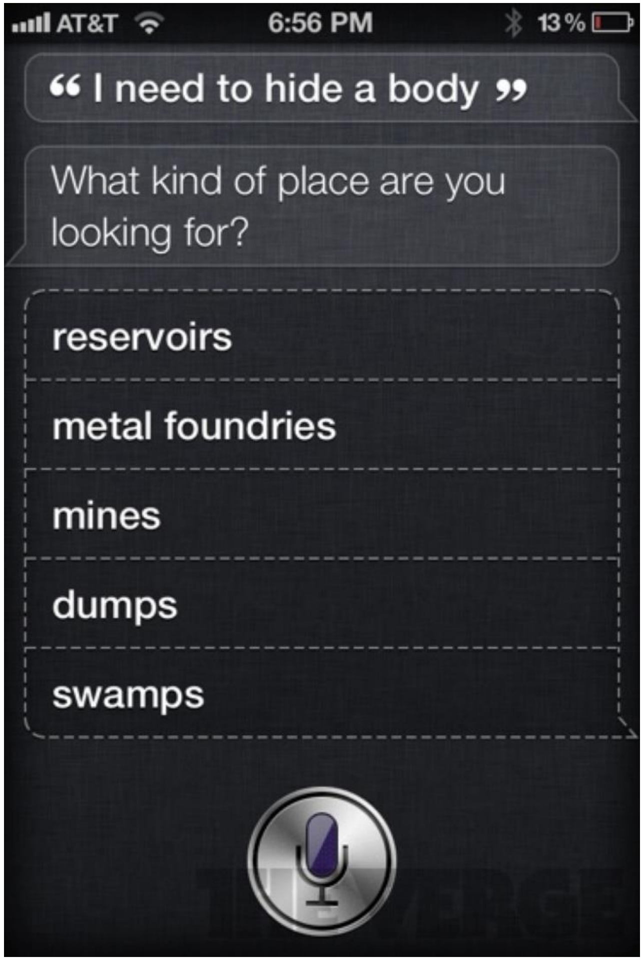
Face recognition

- Facebook http://www.youtube.com/watch?v=l4Rn38_vrLQ
- iPhoto
- Cameras, etc.



Speech recognition



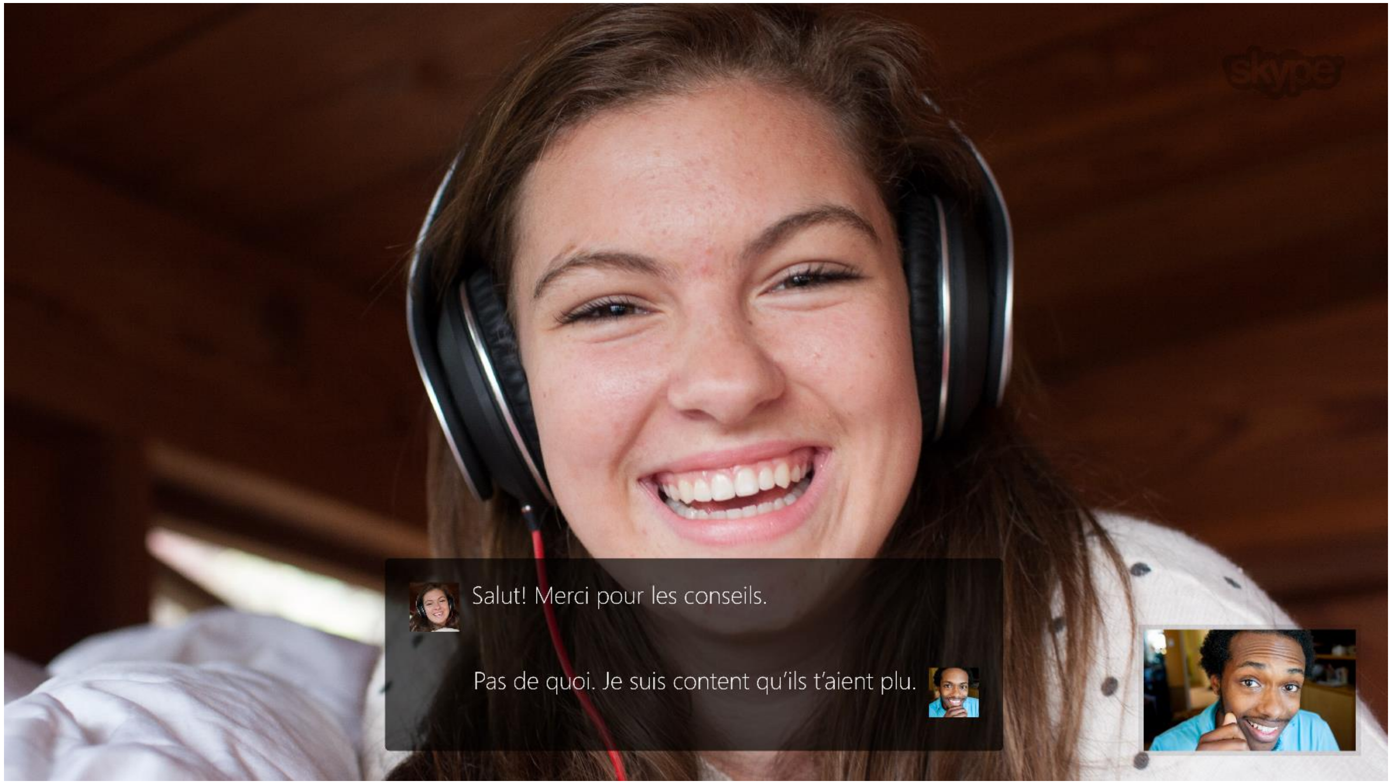


Siri:
http://www.youtube.com/watch?v=8ciagGASro0

Machine Translation

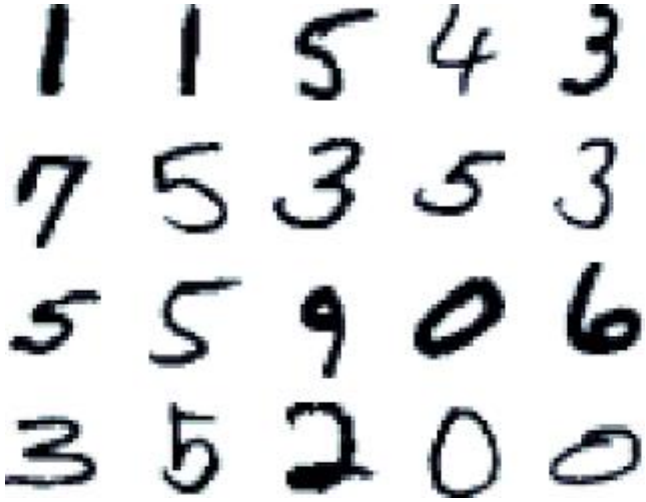
- Skype real-time translation

https://www.youtube.com/watch?v=rek3jjbYRLo



Optical character recognition

- Car plate recognition
- Zip code recognition (mail delivery)
- Electronic versions of printed books (searchable PDFs)



http://detexify.kirelabs.org/classify.html

Web search

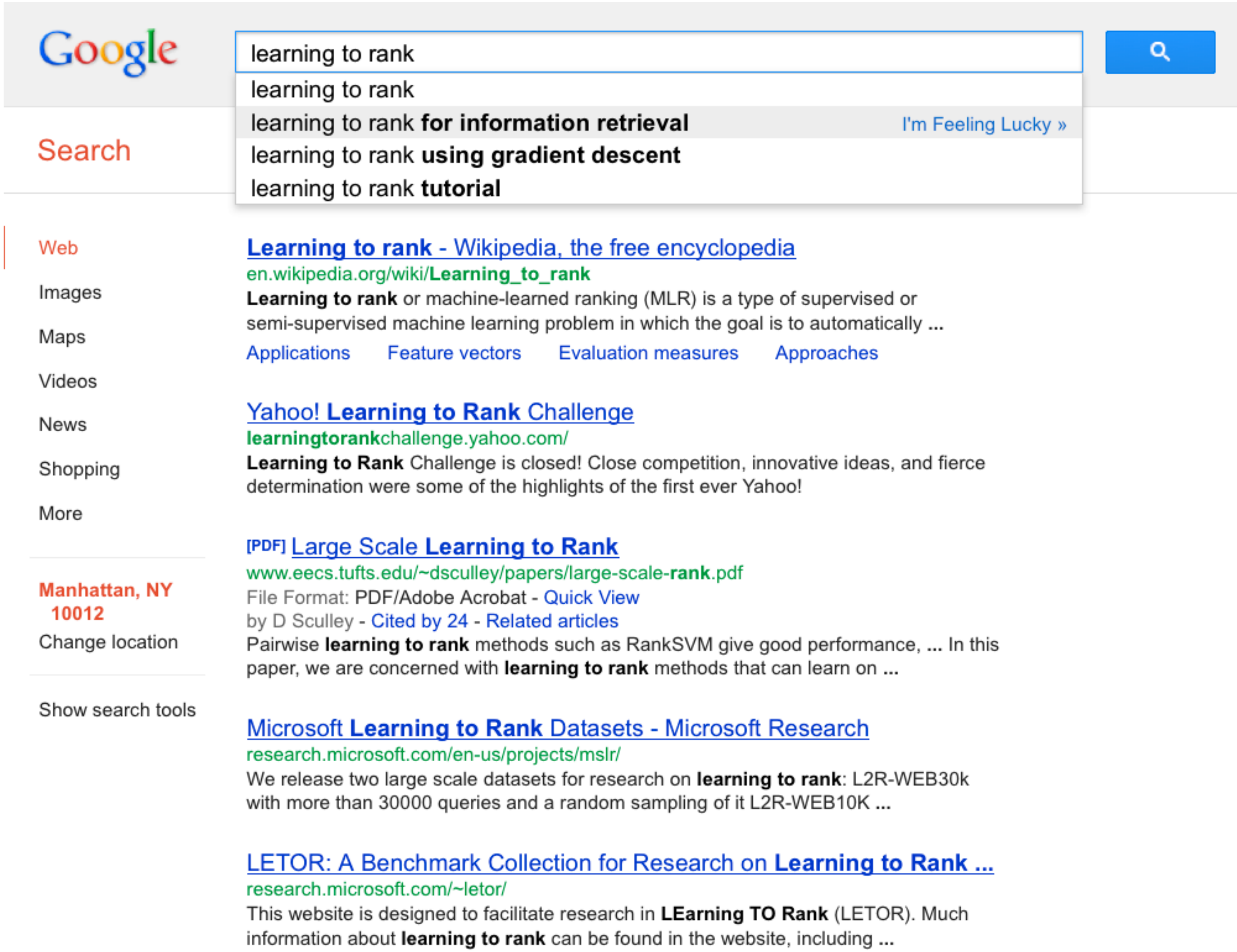
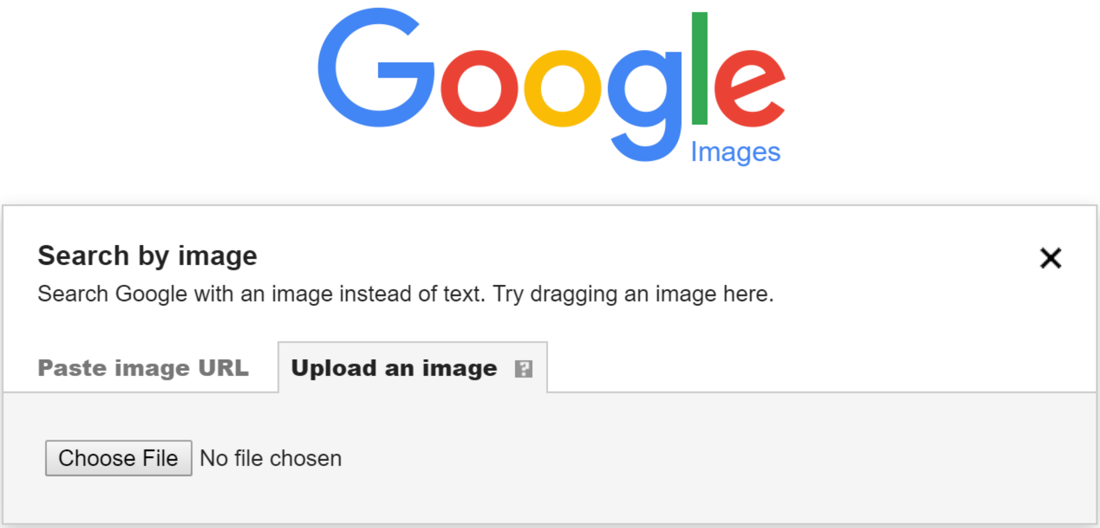

Google
learning to rank
learning to rank
learning to rank for information retrieval
I'm Feeling Lucky »
learning to rank using gradient descent
learning to rank tutorial
Search
Web
Images
Maps
Videos
News
Shopping
More
Manhattan, NY 10012
Change location
Show search tools
Learning to rank - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Learning_to_rank
Learning to rank or machine-learned ranking (MLR) is a type of supervised or semi-supervised machine learning problem in which the goal is to automatically ...
Applications
Feature vectors
Evaluation measures
Approaches
Yahoo! Learning to Rank Challenge
learningtorankchallenge.yahoo.com/
Learning to Rank Challenge is closed! Close competition, innovative ideas, and fierce determination were some of the highlights of the first ever Yahoo!
[PDF] Large Scale Learning to Rank
www.eecs.tufts.edu/~dsculley/papers/large-scale-rank.pdf
File Format: PDF/Adobe Acrobat - Quick View
by D Sculley - Cited by 24 - Related articles
Pairwise learning to rank methods such as RankSVM give good performance, ... In this paper, we are concerned with learning to rank methods that can learn on ...
Microsoft Learning to Rank Datasets - Microsoft Research
research.microsoft.com/en-us/projects/mslr/
We release two large scale datasets for research on learning to rank: L2R-WEB30k with more than 30000 queries and a random sampling of it L2R-WEB10K ...
LETOR: A Benchmark Collection for Research on Learning to Rank ...
research.microsoft.com/~letor/
This website is designed to facilitate research in LEarning TO Rank (LETOR). Much information about learning to rank can be found in the website, including ...

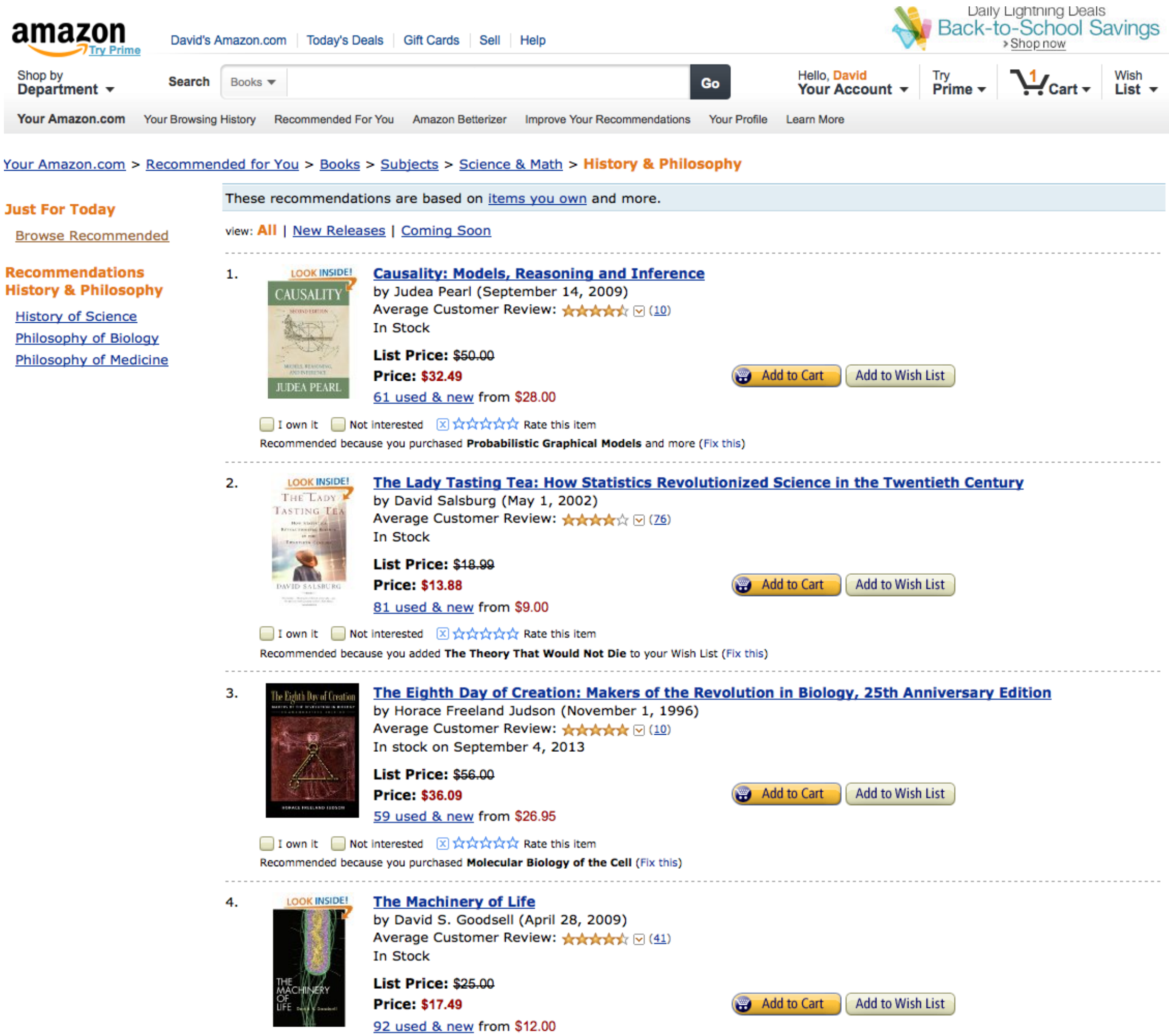
Image search

- Google image search

  https://images.google.com

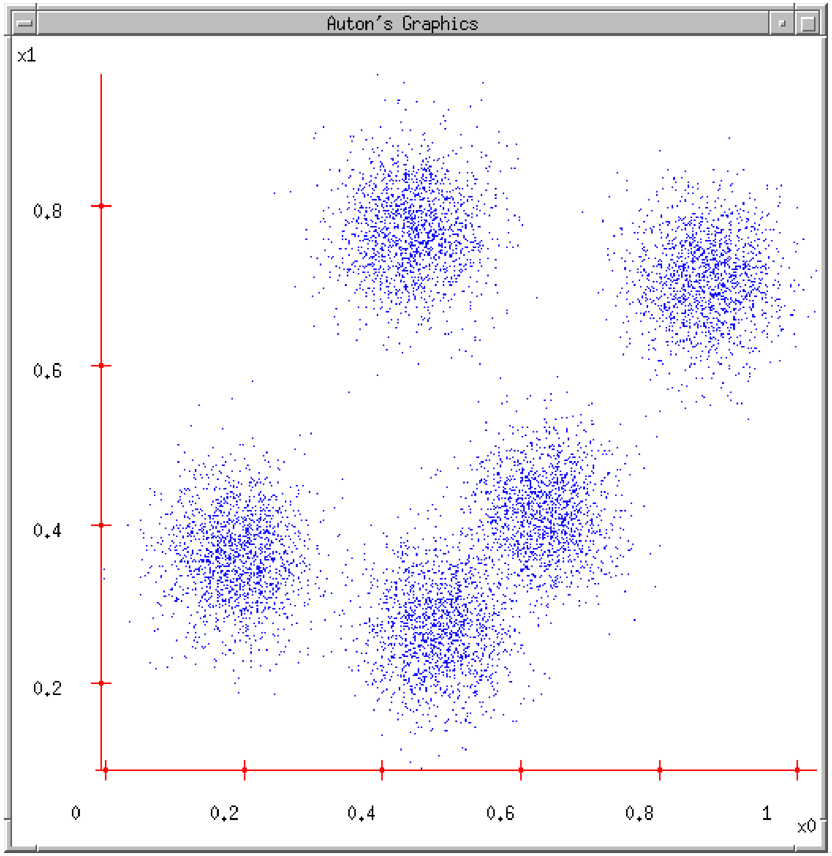


Recommendation systems

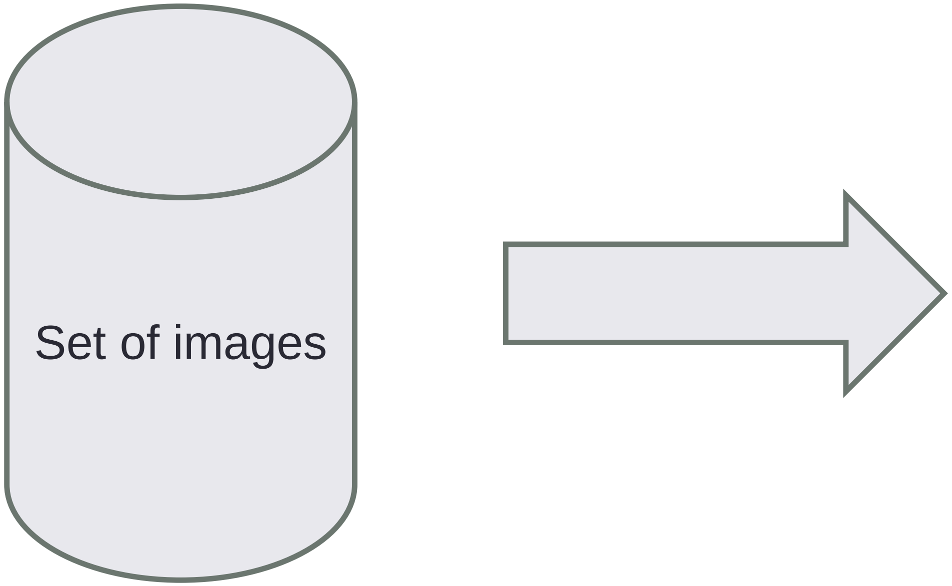


Clustering

- Clustering data into similar groups

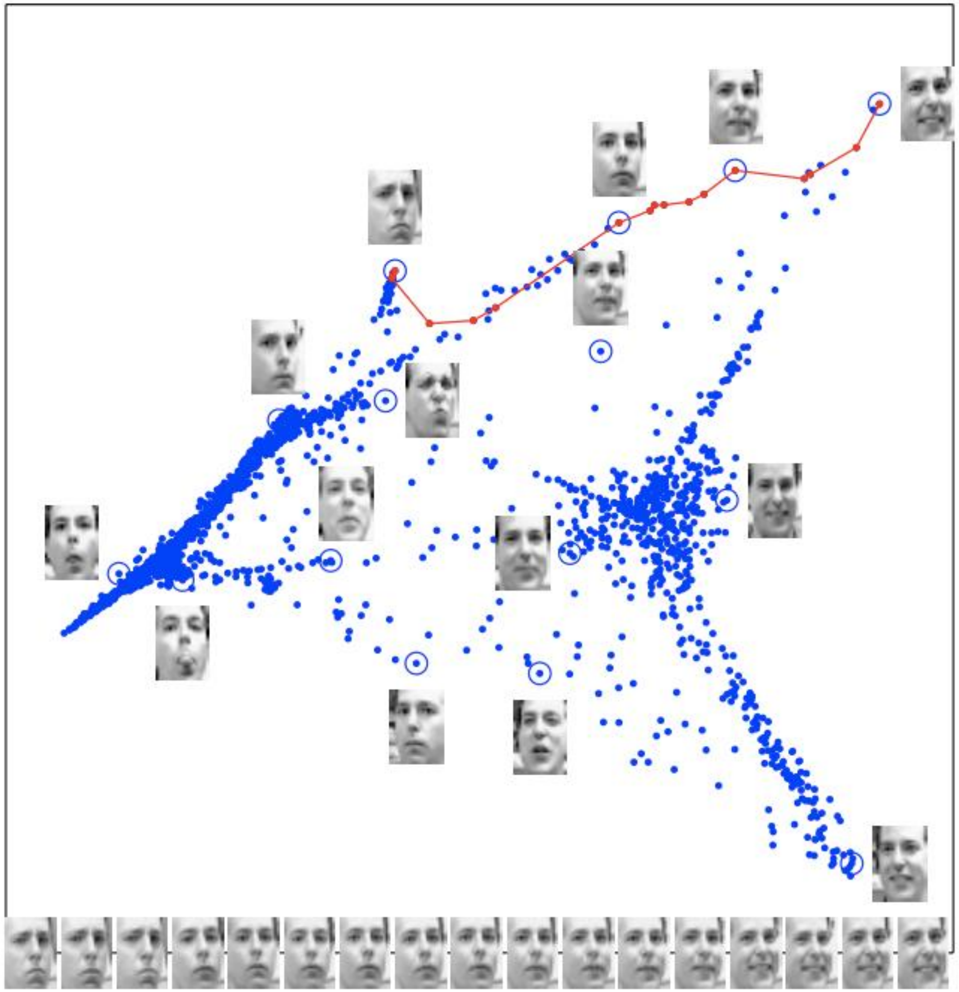


Clustering images



Data visualization (Embedding images)

- Images have thousands or millions of pixels.
- Can we give each image a coordinate, such that similar images are near each other?



[Saul & Roweis ‘03]

Growth of Machine Learning

- Preferred approach to
  - Speech recognition, Natural language processing
  - Computer vision
  - Robot control
  - Computational biology
  - ….
- Accelerating trend
  - Big data (data mining)
  - Improved algorithms
  - Faster computers
  - Availability of good open-source software

COURSE CONTENT

What we will cover

- IGI part:
  - Introduction
  - Linear regression
  - Non-linear basis functions
  - Logistic regression
  - Under- and over-fitting
  - Model selection
  - k-NN
  - Cross-validation
  - Regularization
  - Neural networks
  - SVM
  - Kernel methods
  - Multiclass classification
- SPSC part:
  - Parametric & non-parametric density estimation
  - Bayes classifier
  - Gaussian mixture model
  - K-means
  - Markov model & Hidden Markov model
  - Graphical models
  - PCA
  - LDA

INTRODUCTION:
TYPES OF ML ALGORITHMS

Types of Machine Learning algorithms

- Supervised learning
  - Given: Training examples **with target values**
  - Goal: Predict target values for new examples
  - Examples: optical character recognition, speech recognition, etc.
- Unsupervised learning
  - Given: Training examples **without** target values
  - Goal: Detect and extract structure from data
  - Examples: clustering, segmentation, embedding (visualization), compression, automatic speaker separation

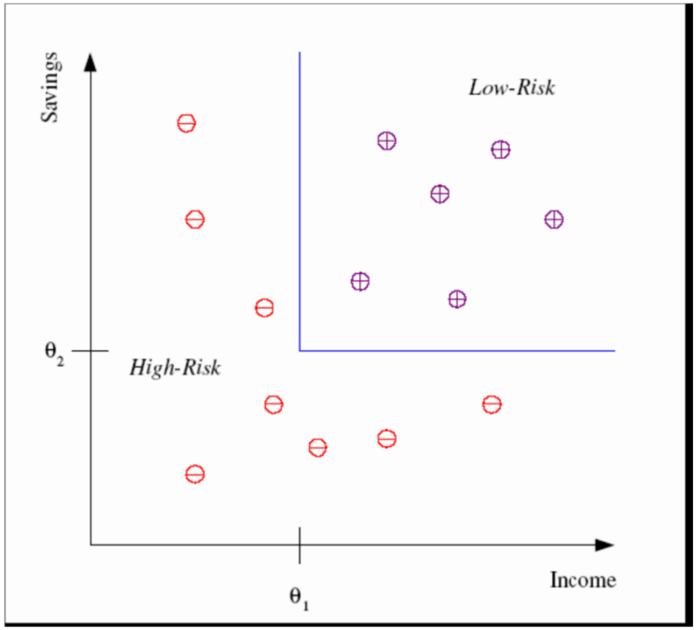
$\Big\}$ *Learning from examples (data)*

- Reinforcement learning (not in this course)
  - Given: Feedback (reward/cost) during trial-and-error episodes
  - Goal: Maximize Reward/minimize cost
  - Examples: learning to control a robot/car/helicopter etc.
  - see Master‘s course „Machine Learning B“

$\Big\}$ *Learning by doing (trial and error)*

Supervised Learning: Example

- Learn to predict output from input *(learning from examples)*
  - Target values (output) can be continuous (regression) or discrete (classification)
  - E.g. predict the risk level (high vs. low) of a loan applicant based on income and savings

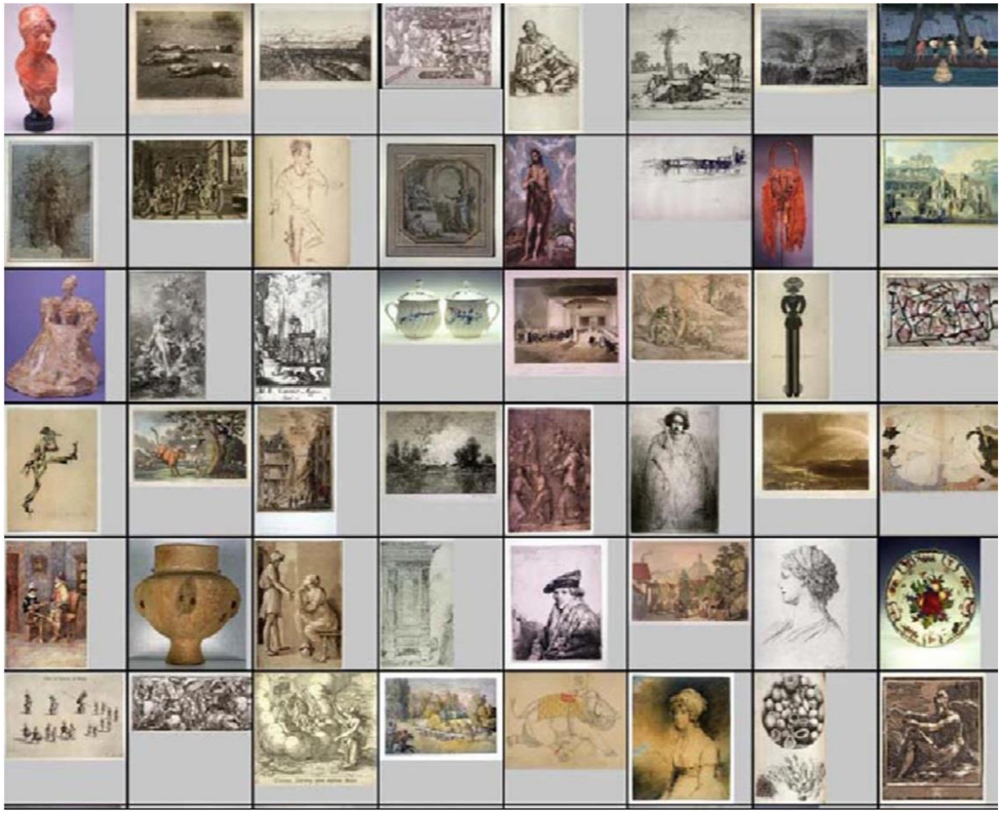


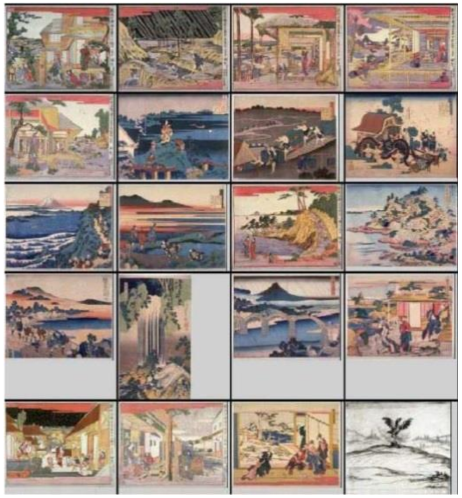
Applications:

- *Spam filters*
- *Character recognition*
- *Speech recognition*
- *Collaborative filtering (predicting if a customer will be interested in an advertisement …)*
- *Medical diagnosis*
- …

Unsupervised Learning: Example

- Find patterns and structure in data





Clustering art

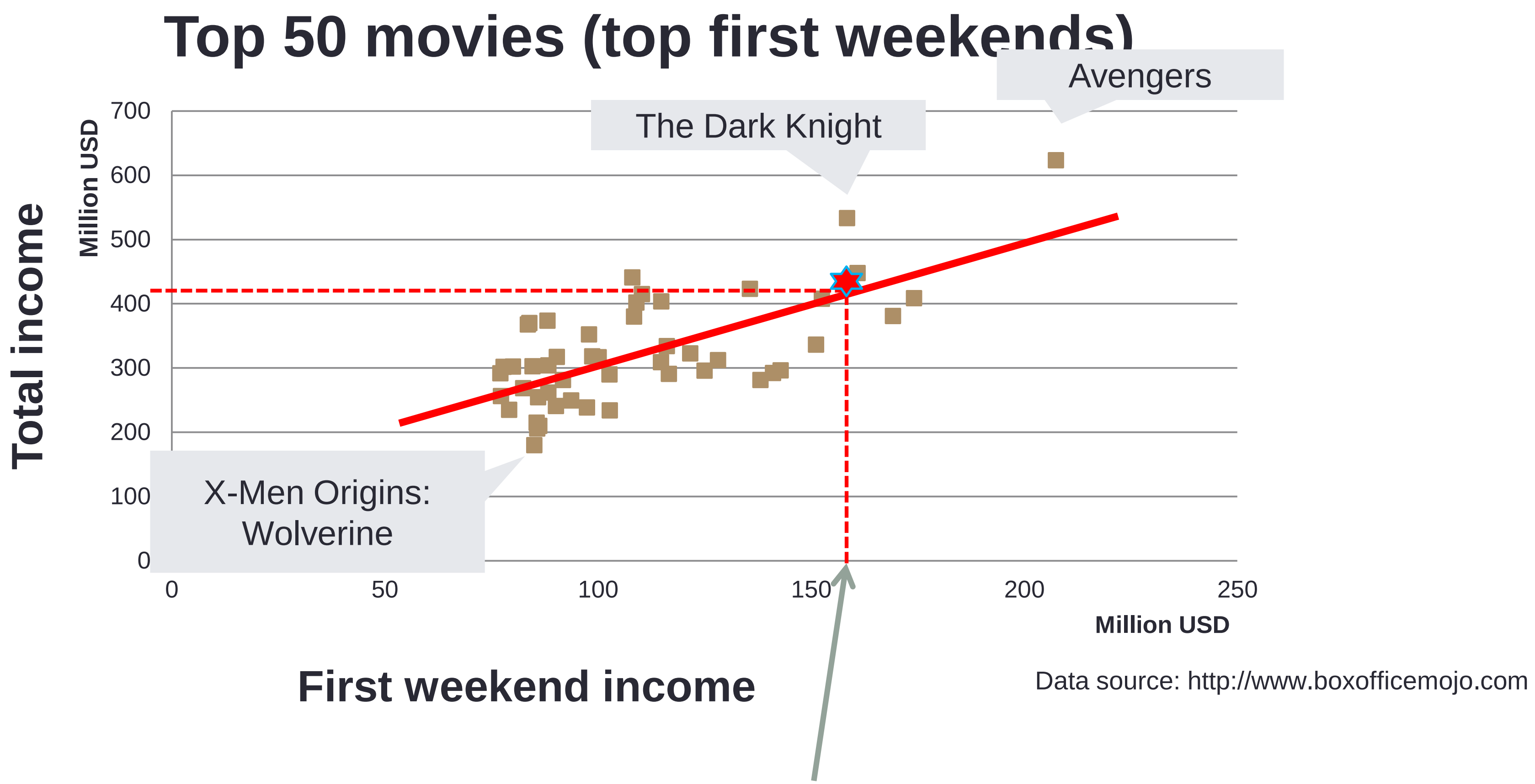
Unsupervised Learning: Applications

- Market partition: divide a market into distinct subsets of customers
  - Find clusters of similar customers, where each cluster may conceivably be selected as a market target to be reached with a distinct marketing strategy
- Image, document, web clustering
  - Automatic organization of pictures
  - Generate a categorized view of a collection of documents
  - For organizing search results etc.
- Bioinformatics
  - Clustering the genes based on their expression profile
  - Find clusters of similarly regulated genes – functional groups

INTRODUCTION: SUPERVISED LEARNING

Regression and classification

Simple regression example

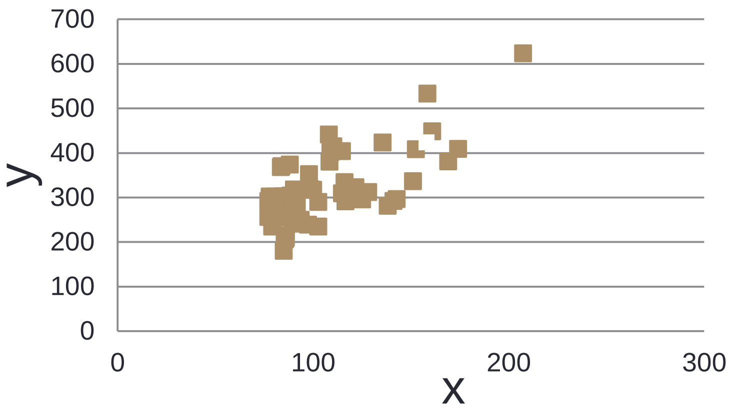


„The Hunger Games: Catching Fire“: **158 Mio. USD** on opening weekend.
How much in total?

Predicted: ~418 Mio., Actual: 424 Mio.

Simple regression example (cont'd)

- **Data set**: Input $x^{(i)}$, Output $y^{(i)}$

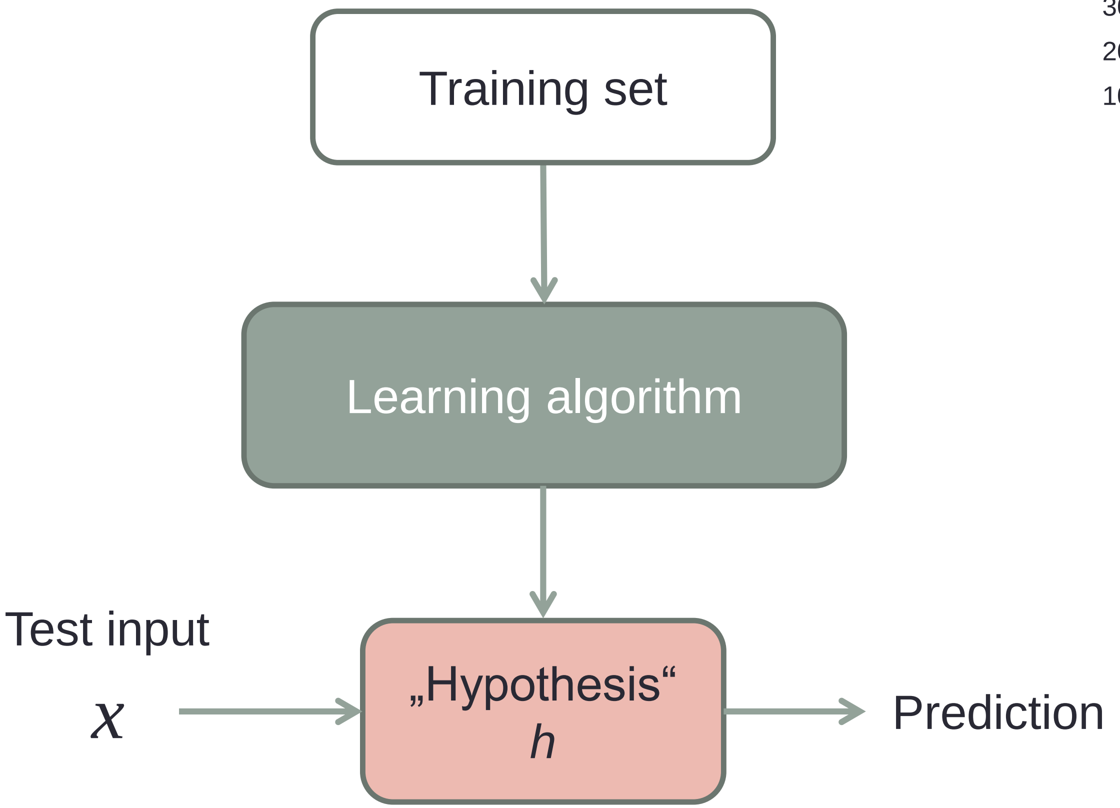


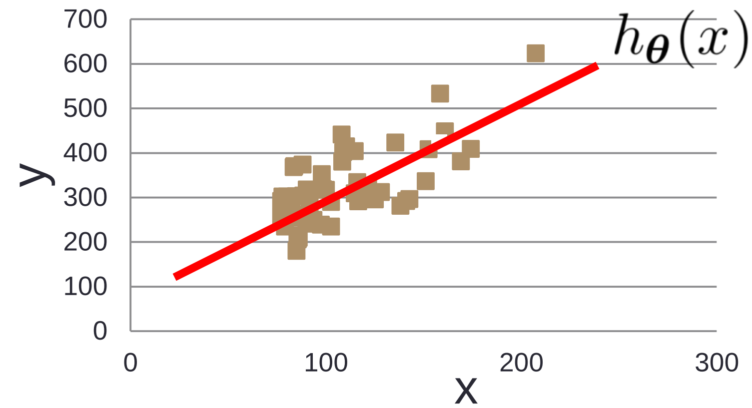
	i	First weekend $x^{(i)}$	Total $y^{(i)}$
Avengers	1	207	623
Iron Man 3	2	174	409
Harry Potter and the Deathly…	3	169	381
The Dark Knight Rises	4	161	449
The Dark Knight	5	158	533
		…	…

m data points (data samples)

Simple regression example (cont'd)

- **Data set**: $\langle x^{(1)}, y^{(1)} \rangle$ … $\langle x^{(m)}, y^{(m)} \rangle$





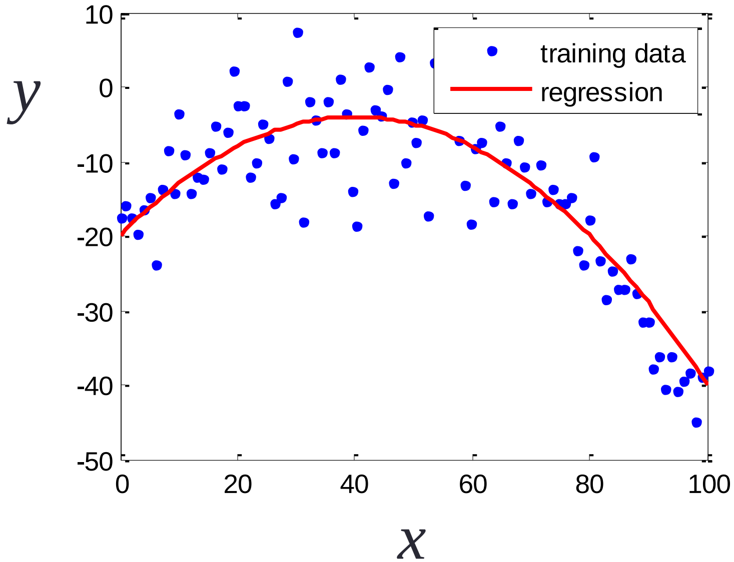
Hypothesis

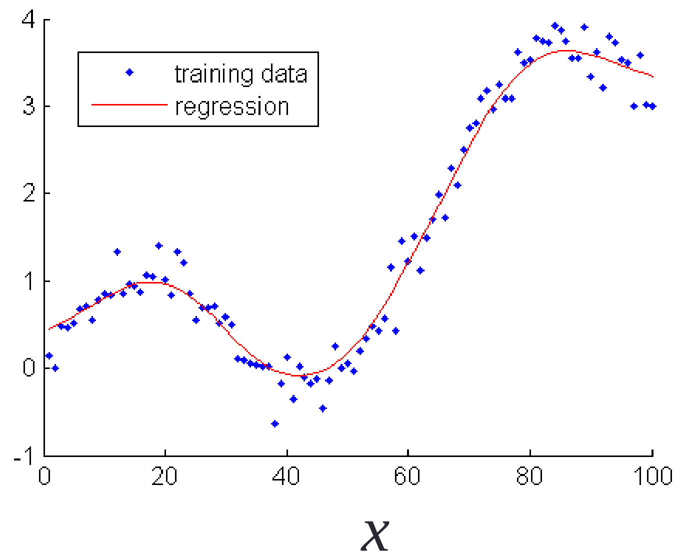
$$h_{\boldsymbol{\theta}}(x) = \theta_0 + \theta_1 \cdot x$$

Parameters

$$\boldsymbol{\theta} = (\theta_0, \theta_1)$$

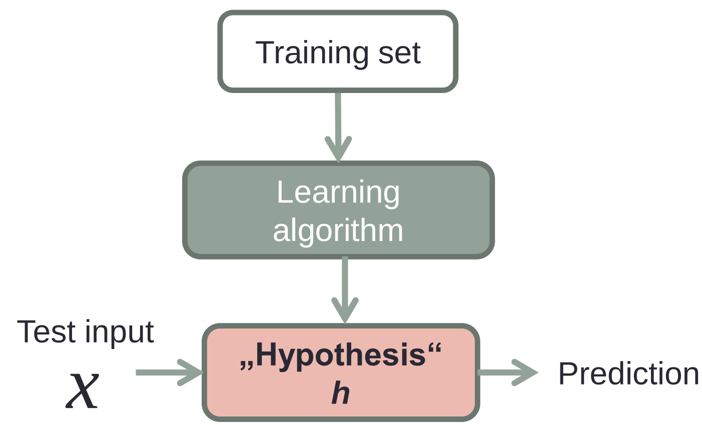
Non-linear regression



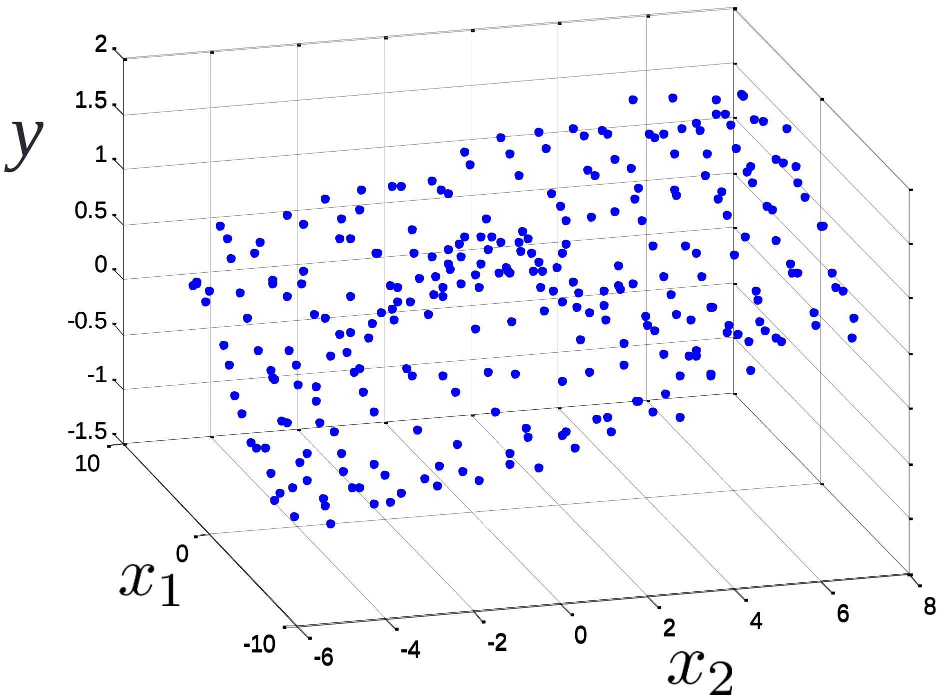


Non-linear hypothesis, for example

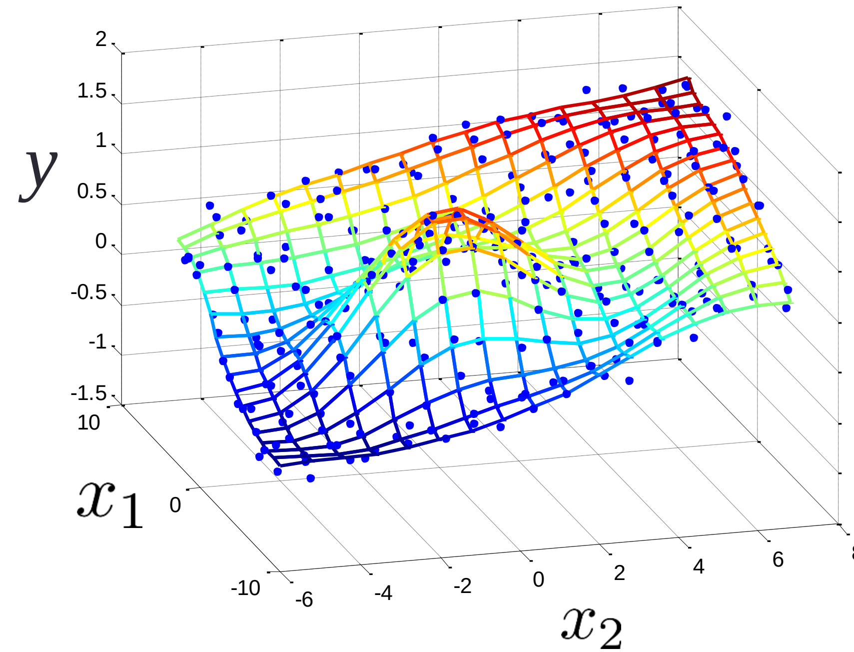
$$h_{\boldsymbol{\theta}}(x) = \theta_0 + \theta_1 \cdot x + \theta_2 \cdot x^2$$



Regression with multiple inputs



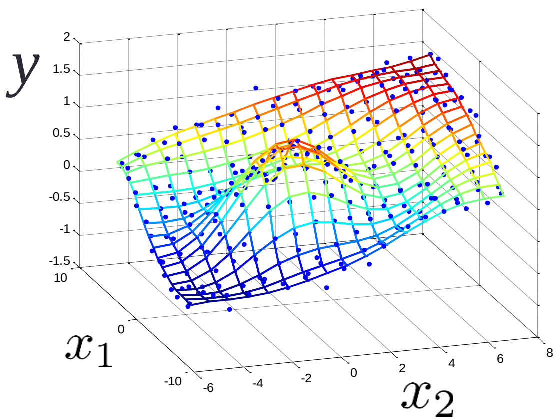
linear hypothesis



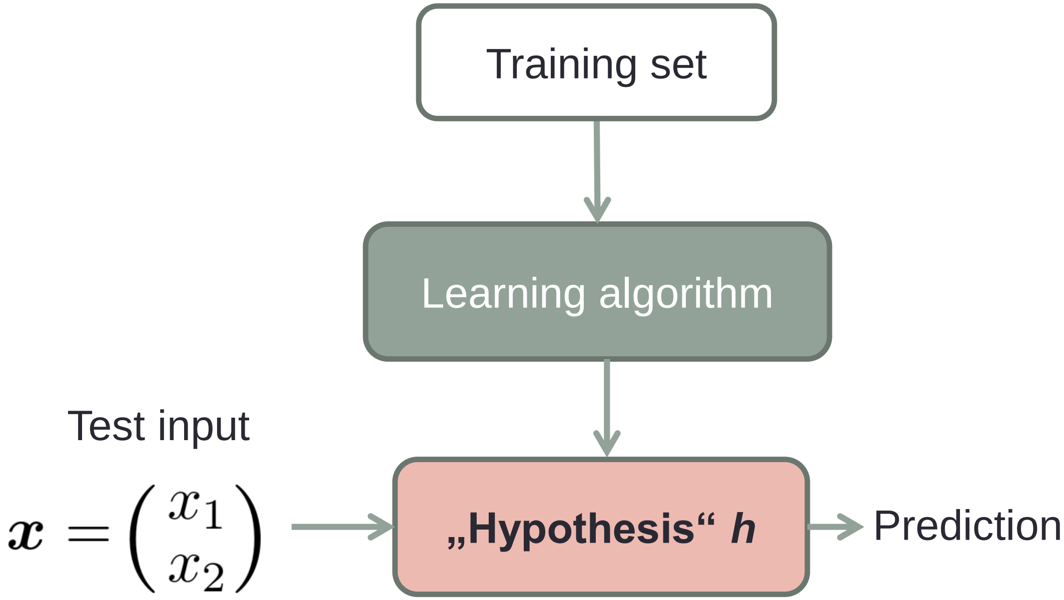
non-linear hypothesis

$$h_{\boldsymbol{\theta}}(x_1, x_2) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2$$

Multiple inputs continued



i	$x_1^{(i)}$	$x_2^{(i)}$	$y^{(i)}$
1	5.3	-2.1	2.31
2	0.4	3.5	-1.3
3	1.2	0.9	1.9
4	-0.3	0.1	-0.7
5	...	...	...

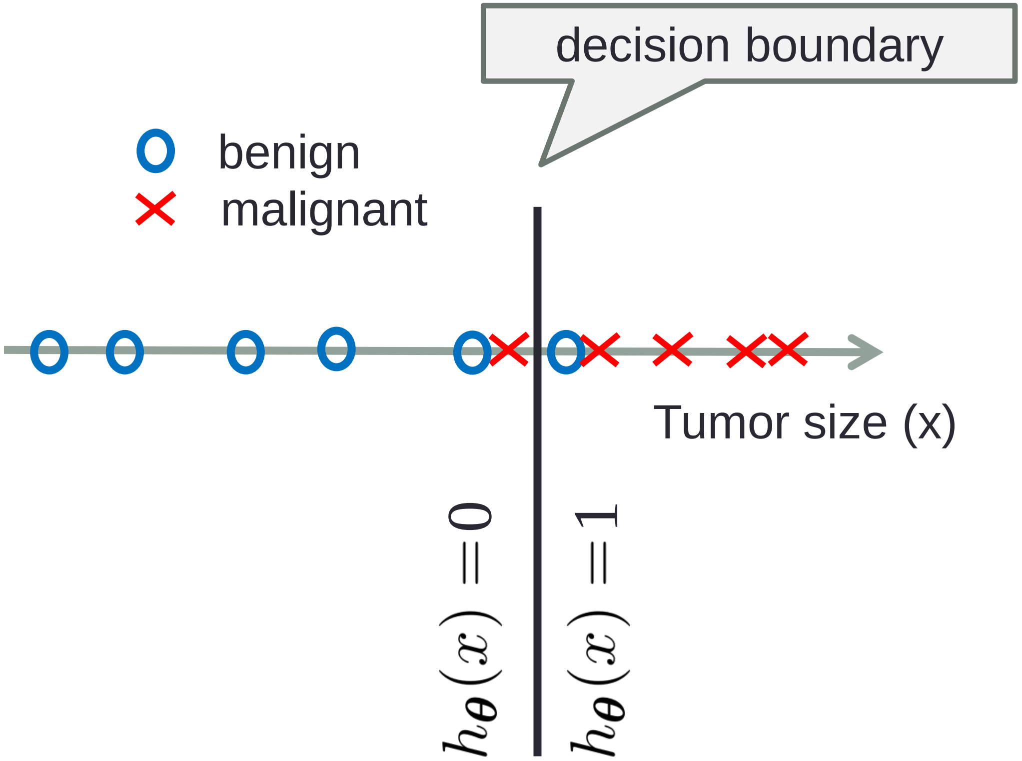


Simple classification example

„labeled data“

i	Tumor size (mm) x	Malignant? y
1	2.3	**0 (N)**
2	5.1	**1 (Y)**
3	1.4	**0 (N)**
4	6.3	**1 (Y)**
5	5.3	**1 (Y)**
	…	…

labels

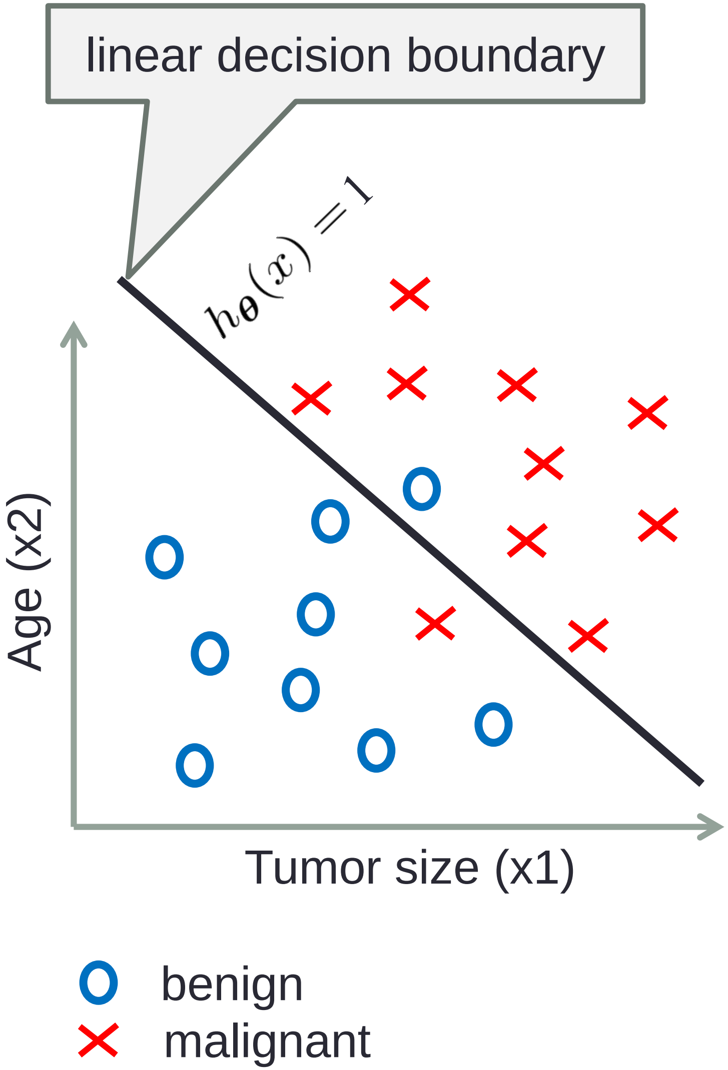


Example hypothesis:

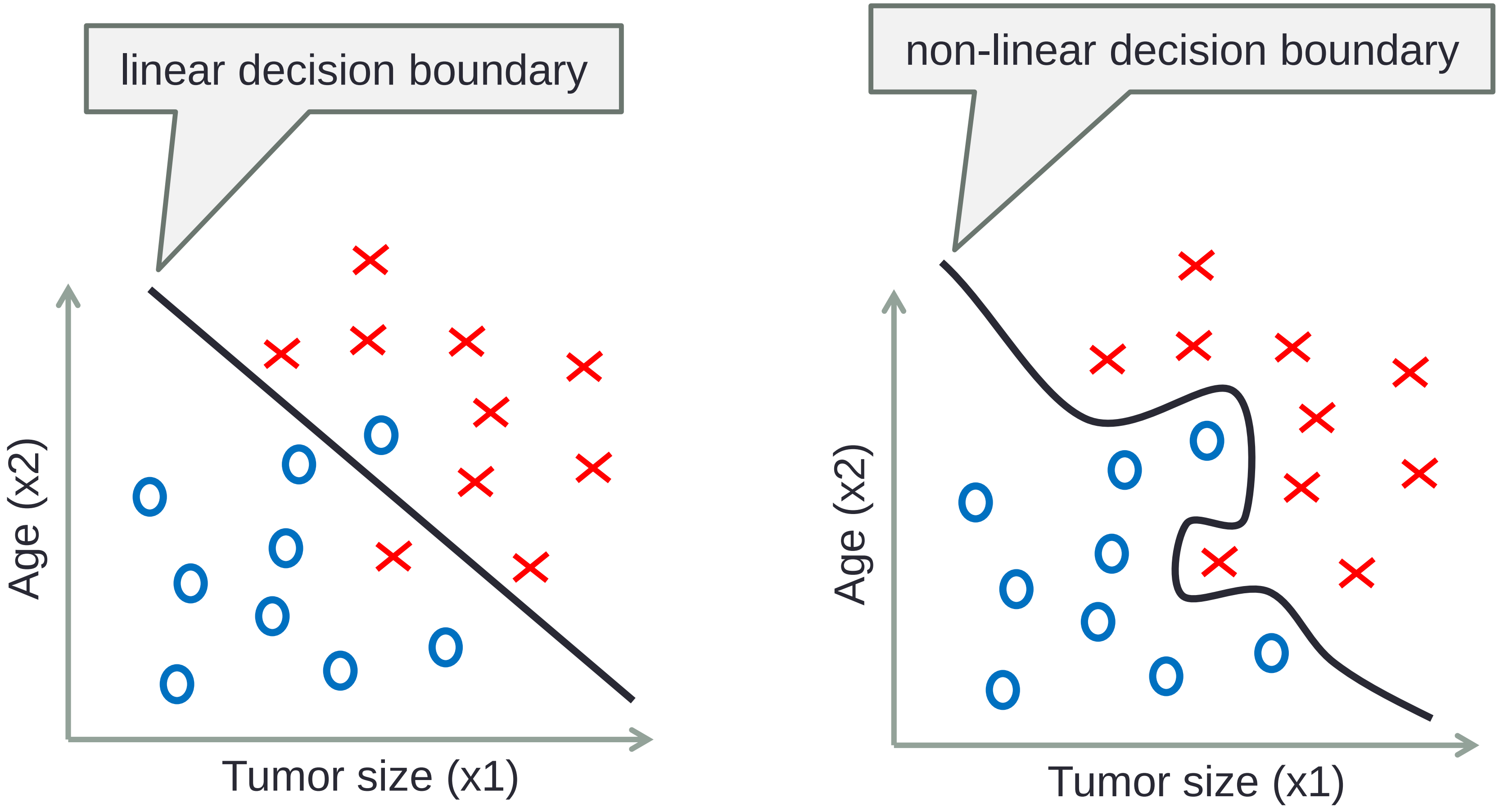
$$h_{\boldsymbol{\theta}}(x) = 1 \quad \text{if x} > \theta_0$$

Classification with multiple inputs

i	Tumor size (mm) x1	Age x2	Malignant? y
1	2.3	25	0 (N)
2	5.1	62	1 (Y)
3	1.4	47	0 (N)
4	6.3	39	1 (Y)
5	5.3	72	1 (Y)
	…		…



Non-linear classification



Both hypotheses fit the data quite well.
Which one fits the **training data** better?
Which one would you trust more for **prediction**?

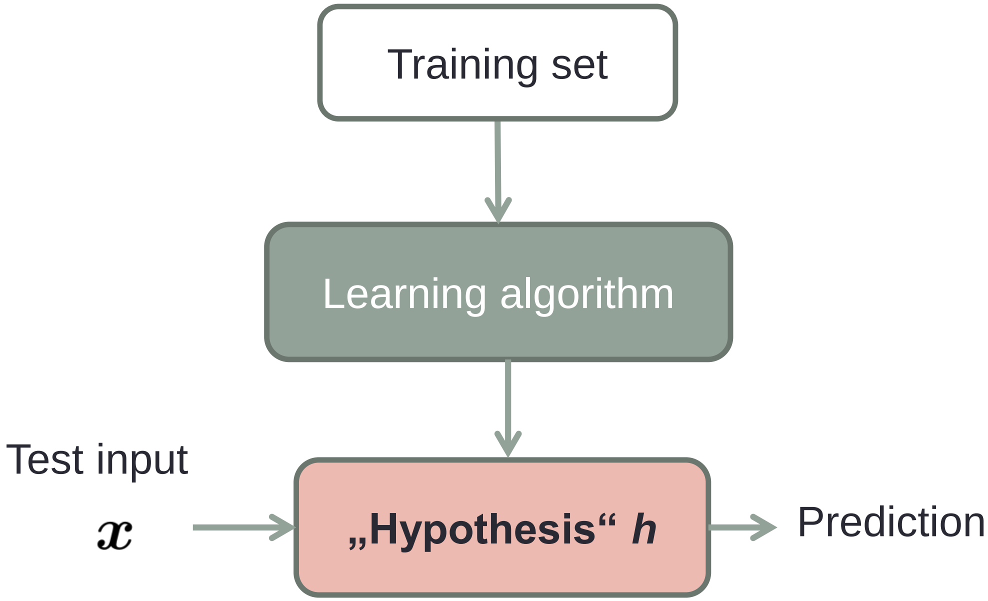
Supervised learning (Regr., Class.)

- Discrete vs. continuous outputs (classification vs. regression)

In the next few classes we'll cover:

- **Learning algorithms** for regression and classification
  (linear regression, neural nets, SVMs, etc.)

- Supervised learning **in practice** (overfitting, etc.)



What is next?

- Lecture by Guillaume Bellec on:
    - Linear regression
    - Gradient descent
    - Non-linear basis functions