

COMPUTATIONAL INTELLIGENCE

(INTRODUCTION TO MACHINE LEARNING)

SS 17

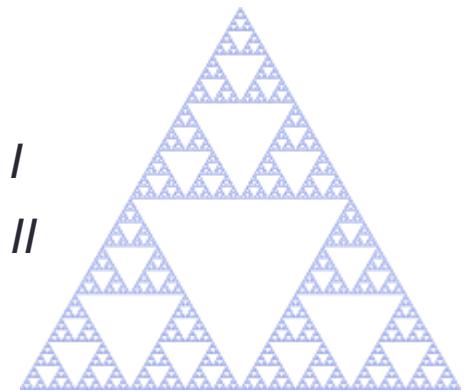
2 VO 442.070 + 1 UE 708.070

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)
 - Assoc. Prof. Dr. Franz Pernkopf (SPSC)
- Practical / UE:
 - First practical on **Friday 10th of March**, HS i11
 - 12:30-13:30 – if your last name starts with **A-L**
 - 14:00-15:00 – if your last name starts with **M-Z**
 - 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

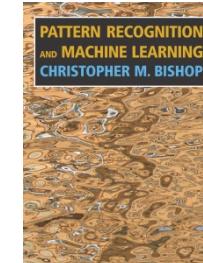


Organization

- Office hours:
 - Both Anand and Guillaume:
 - Time: Every Tuesday 14:00 – 15:00
 - Place: Our offices at Inffeldgasse 16b/1
- Exam:
 - Written exam for this year's course:
 - July onwards
 - 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?

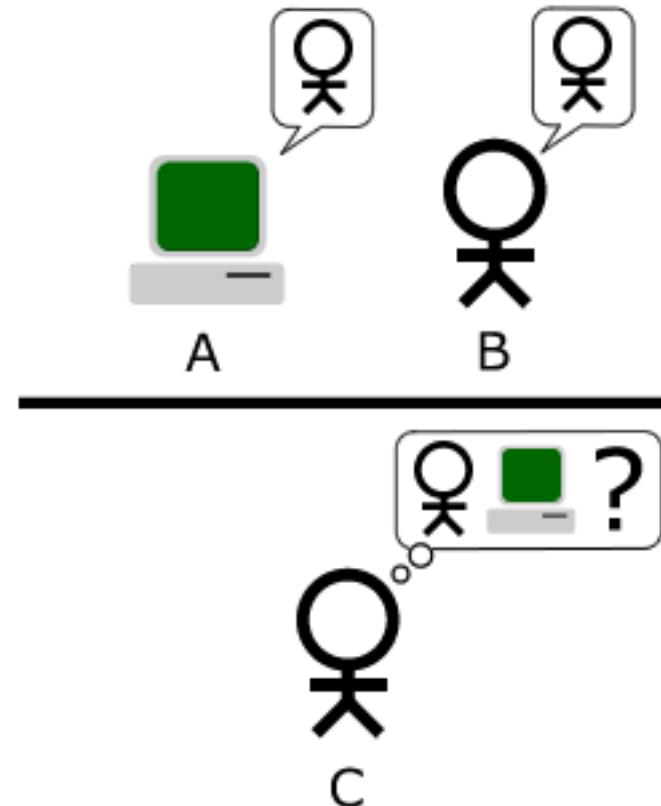
<p>"The exciting new effort to make computers think . . . <i>machines with minds</i>, in the full and literal sense" (Haugeland, 1985)</p> <p>"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)</p>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

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

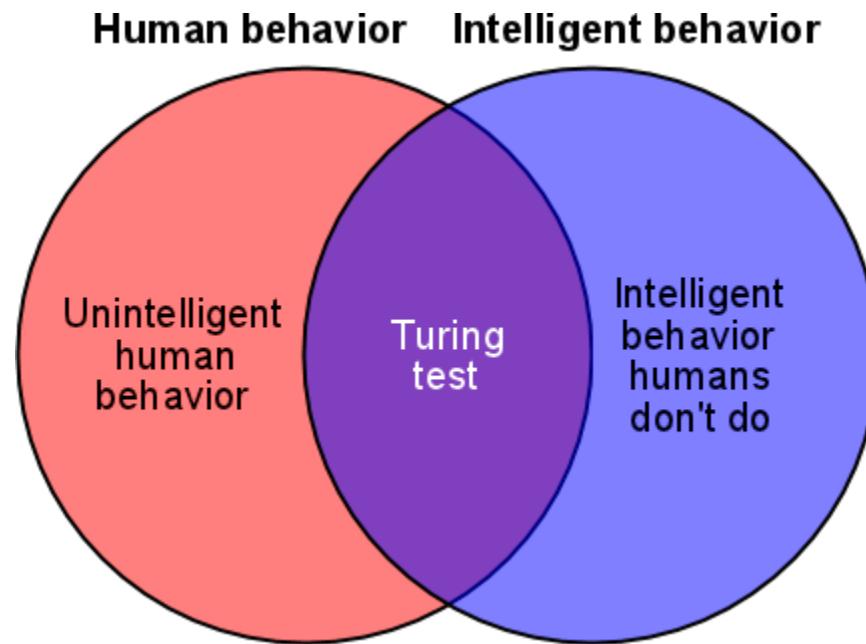
Systems that think like humans.	Systems that think rationally.
Systems that act like humans.	Systems that act rationally.

But what really is 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?

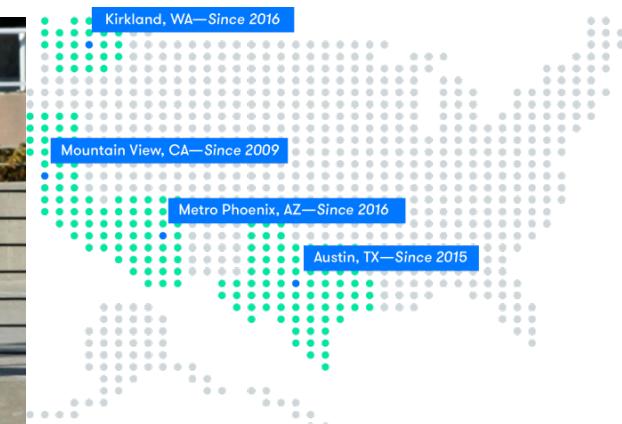
- When human expert knowledge is missing
 - For example, predicting whether some new substance could be an effective treatment for a disease
- When humans can only do it “intuitively”
 - Flying a helicopter
 - Recognize visual objects
 - Natural language processing
- When we need to learn about something 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
 -

Autonomous car

- Waymo/Alphabet
 - <https://www.youtube.com/watch?v=TsaES--OTzM>



+ UK, France,
Switzerland,
Singapore

Bipedal robot

- ATLAS (Boston Dynamics/Alphabet)
 - <https://www.youtube.com/watch?v=rVlhMGQgDkY>



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	prediction
<p>Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only \$5.95 for shipping mfw rlk</p> <p>Spam X</p> <p>Jaquelyn Halley to nherlein, bcc: thehorney, bcc: anç show details 9:52 PM (1 hour ago) Reply ▾</p> <p>==== Natural WeightLOSS Solution ===</p> <p>Vital Acai is a natural WeightLOSS product that Enables people to lose wieght and cleansing their bodies faster than most other products on the market.</p> <p>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.</p> <ul style="list-style-type: none">* Rapid WeightLOSS* 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 <p>http://sfsi.kaeconomic.cn http://sivk.kaeconomic.cn</p>	 <p>Spam vs. Not Spam</p>

Web search

The screenshot shows a Google search interface. The search bar contains the query "learning to rank". Below the search bar, a dropdown menu lists several related search terms: "learning to rank", "learning to rank for information retrieval", "learning to rank using gradient descent", and "learning to rank tutorial". To the right of the dropdown is a blue "I'm Feeling Lucky" button. On the left side of the page, there is a sidebar with links for "Search", "Web", "Images", "Maps", "Videos", "News", "Shopping", and "More". Under the "More" section, it says "Manhattan, NY 10012" and "Change location". At the bottom of the sidebar is a link "Show search tools". The main content area displays four search results:

- 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>



Search by image X

Search Google with an image instead of text. Try dragging an image here.

Paste image URL **Upload an image**

No file chosen

Face recognition

- Facebook
 - http://www.youtube.com/watch?v=l4Rn38_vrLQ
- iPhoto
- Cameras, etc.
- Microsoft cognitive services
 - From face, can recognize age, gender, emotions!
 - <https://www.microsoft.com/cognitive-services/>



Scene and text recognition

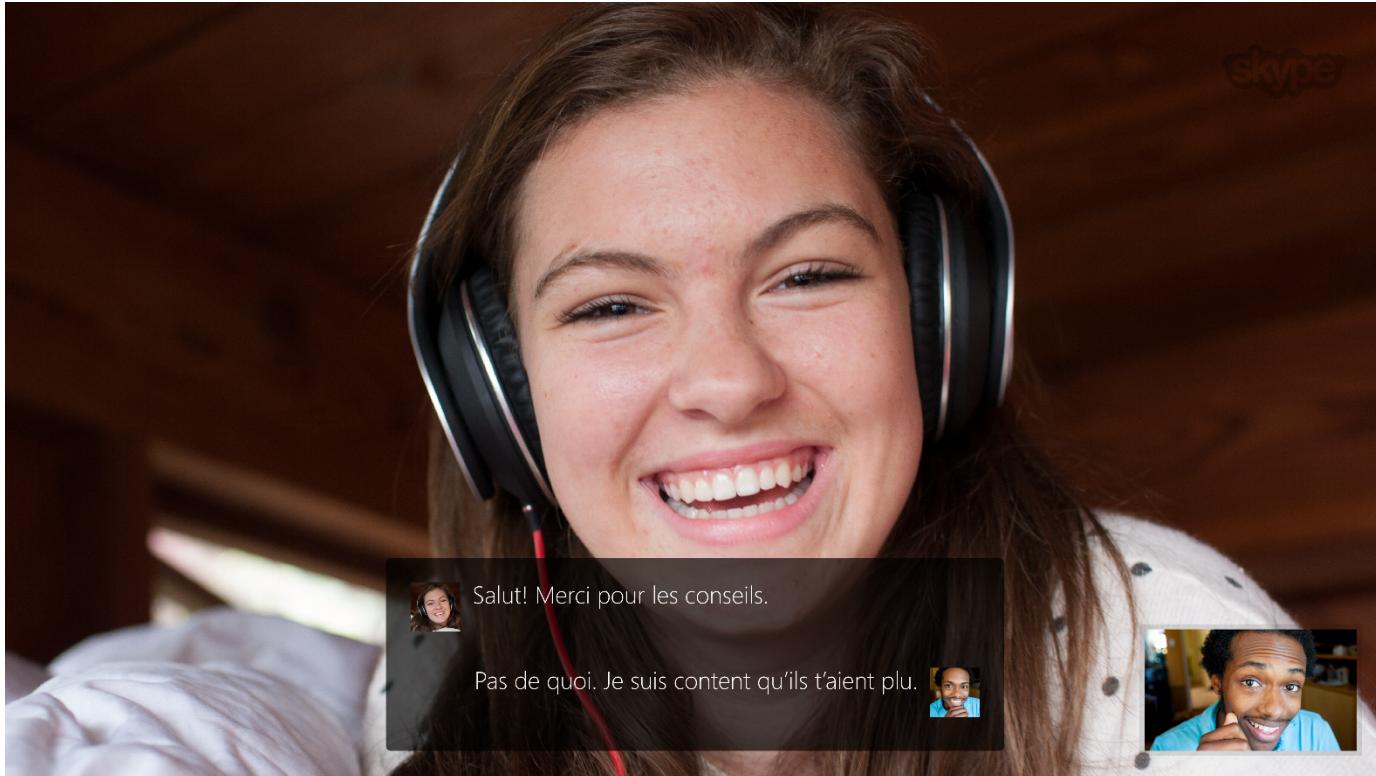
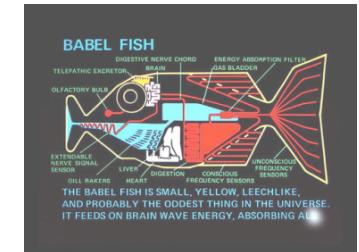
- Microsoft Seeing AI project
 - <https://www.youtube.com/watch?v=R2mC-NUAmMk>



Machine Translation

- Skype real-time translation (Microsoft)

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



Recommender systems

amazon Try Prime

David's Amazon.com | Today's Deals | Gift Cards | Sell | Help

Shop by Department | Search Books | Go | Hello, David Your Account | Try Prime | Cart | Wish List

Daily Lightning Deals
Back-to-School Savings
>Shop now

Your Amazon.com Your Browsing History Recommended For You Amazon Betterizer Improve Your Recommendations Your Profile Learn More

Your Amazon.com > Recommended For You > Books > Subjects > Science & Math > History & Philosophy

Just For Today

[Browse Recommended](#)

Recommendations History & Philosophy

[History of Science](#)
[Philosophy of Biology](#)
[Philosophy of Medicine](#)

These recommendations are based on items you own and more.

view: All | New Releases | Coming Soon

1. **Causality: Models, Reasoning and Inference**
by Judea Pearl (September 14, 2009)
Average Customer Review: ★★★★★ (10)
In Stock

List Price: \$60.00
Price: \$32.49

[Add to Cart](#) [Add to Wish List](#)

I own it Not interested Rate this item

Recommended because you purchased Probabilistic Graphical Models and more ([Fix this](#))

2. **The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century**
by David Salsburg (May 1, 2002)
Average Customer Review: ★★★★★ (26)
In Stock

List Price: \$48.99
Price: \$13.88

[Add to Cart](#) [Add to Wish List](#)

I own it Not interested Rate this item

Recommended because you added **The Theory That Would Not Die** to your Wish List ([Fix this](#))

3. **The Eighth Day of Creation: Makers of the Revolution in Biology, 25th Anniversary Edition**
by Horace Freeland Judson (November 1, 1996)
Average Customer Review: ★★★★★ (10)
In stock on September 4, 2013

List Price: \$56.00
Price: \$36.09

[Add to Cart](#) [Add to Wish List](#)

I own it Not interested Rate this item

Recommended because you purchased **Molecular Biology of the Cell** ([Fix this](#))

4. **The Machinery of Life**
by David S. Goodsell (April 28, 2009)
Average Customer Review: ★★★★★ (41)
In Stock

List Price: \$25.00
Price: \$17.49

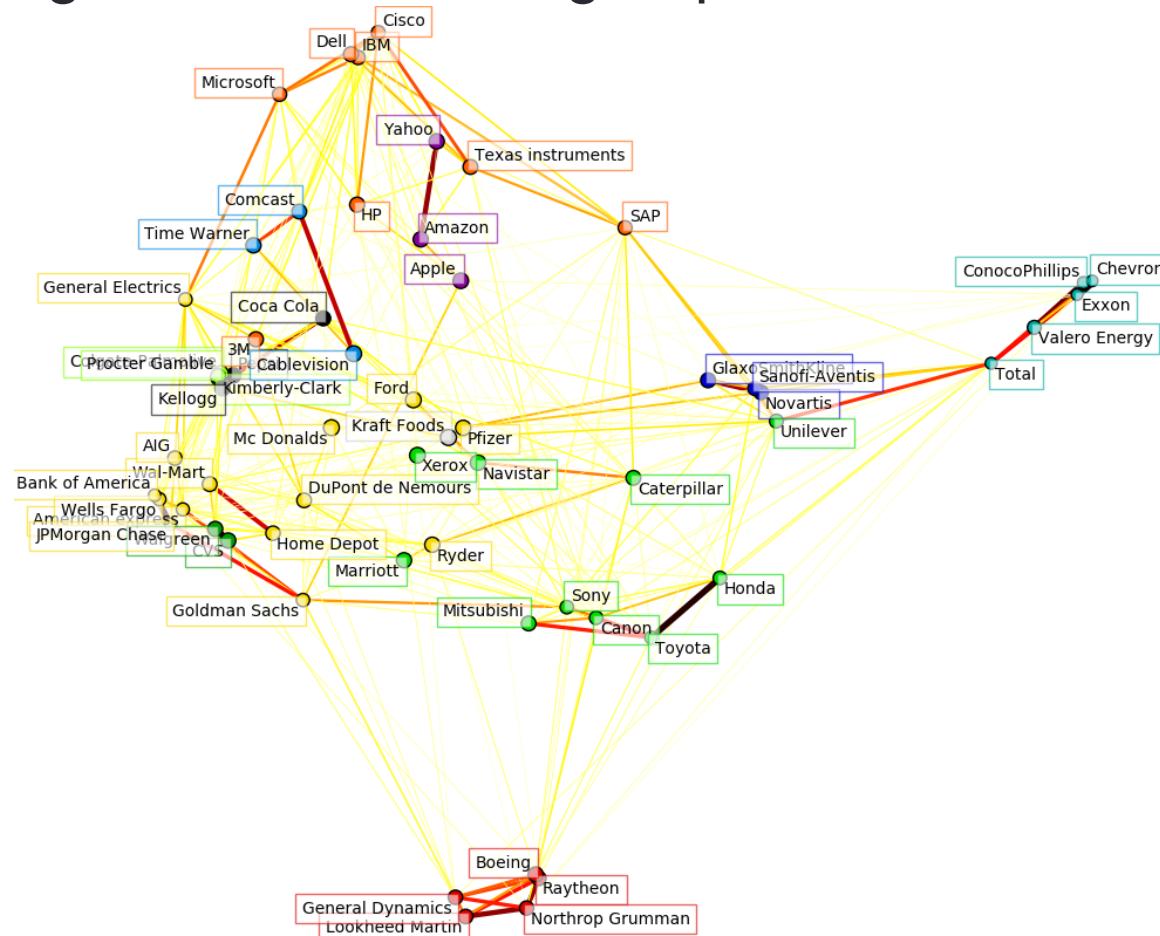
[Add to Cart](#) [Add to Wish List](#)

I own it Not interested Rate this item

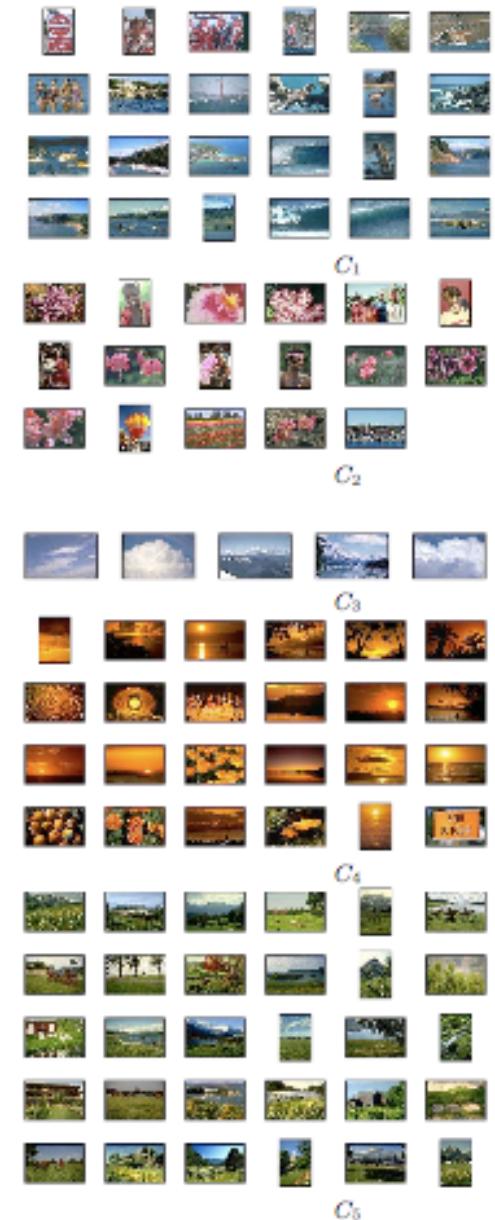
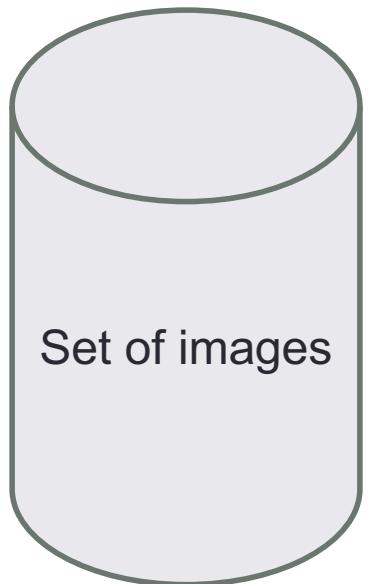
Recommended because you purchased **The Eighth Day of Creation** ([Fix this](#))

Clustering

- Clustering data into similar groups

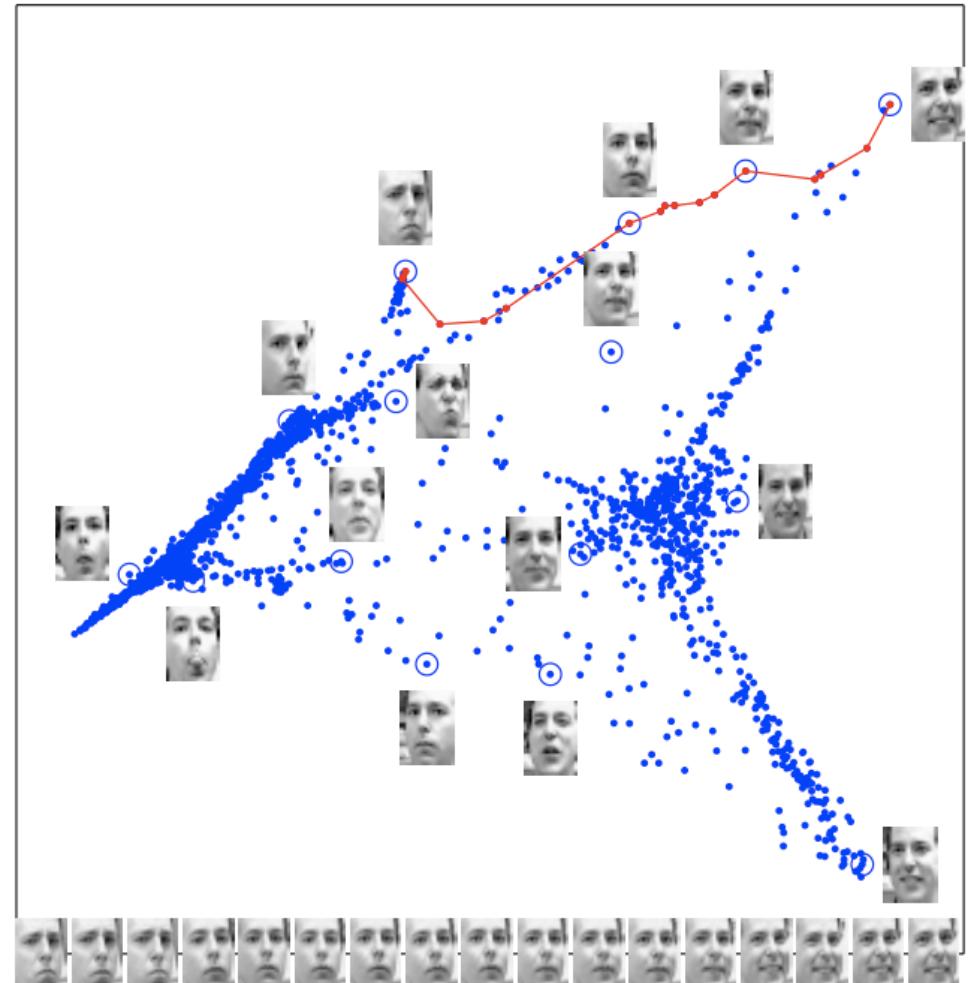


Clustering images



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?

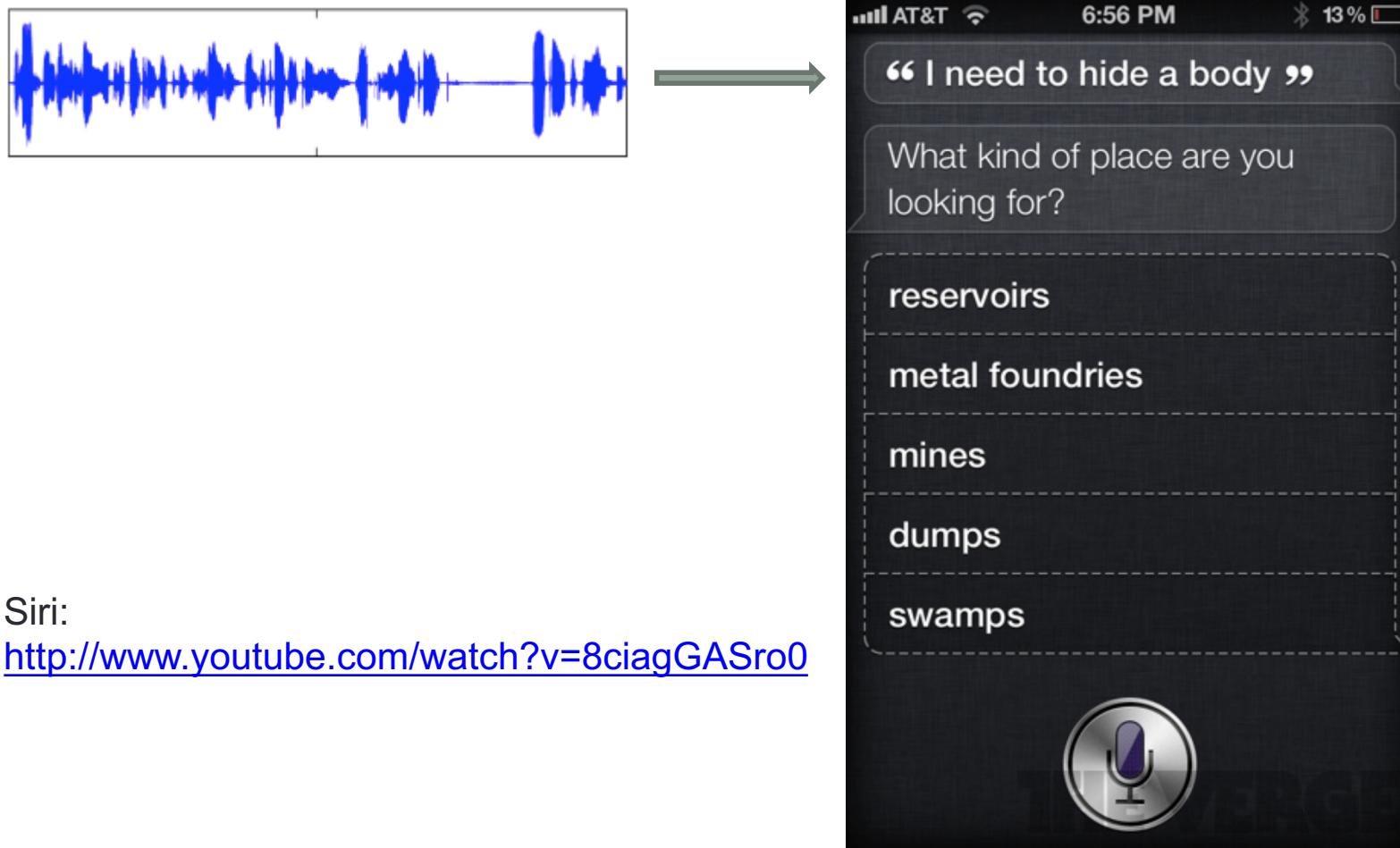


[Saul & Roweis '03]

Growth of Machine Learning

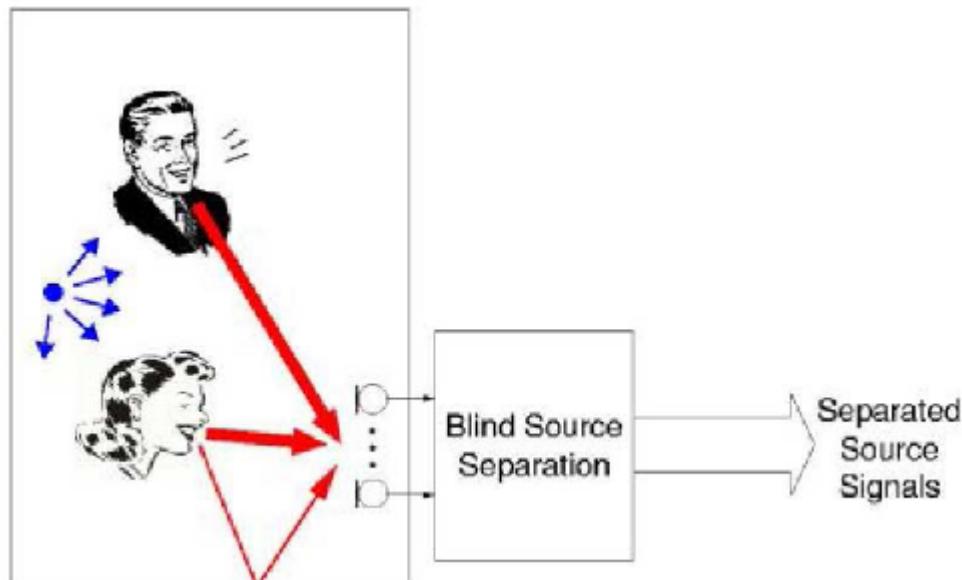
- Preferred approach to
 - Speech recognition, Natural language processing
 - Computer vision
 - Robot control
 - Computational biology
 -
- Accelerating trend
 - Big data (data mining)
 - Improved algorithms
 - Faster computers
 - Availability of good open-source software

Speech recognition

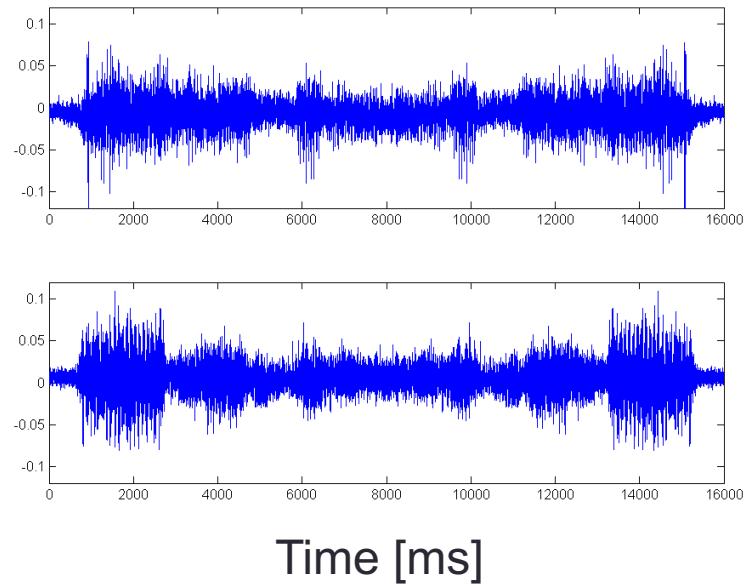


Cocktail party problem

Blind source separation



Independent component analysis



Optical character recognition

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

1 1 5 4 3
7 5 3 5 3
5 5 9 0 6
3 5 2 0 0

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

Clustering web search results

The screenshot shows the Clusty search interface. The top navigation bar includes links for web, news, images, wikipedia, blogs, jobs, and more. A search bar contains the query "race". Below the search bar are links for advanced preferences and a search results link. The main content area displays a cluster titled "Cluster Human" containing 8 documents. The results are listed as follows:

- Race (classification of human beings) - Wikipedia, the free ...**
The term race or racial group usually refers to the concept of dividing humans into populations or groups on the basis of various sets of characteristics. The most widely used human racial categories are based on visible traits (especially skin color, cranial or facial features and hair texture), and self-identification. Conceptions of race, as well as specific ways of grouping races, vary by culture and over time, and are often controversial for scientific as well as social and political reasons. History · Modern debates · Political and ...
[en.wikipedia.org/wiki/Race_\(classification_of_human_beings\)](http://en.wikipedia.org/wiki/Race_(classification_of_human_beings)) - [cache] - Live, Ask
- Race - Wikipedia, the free encyclopedia**
General. Racing competitions The Race (yachting race), or La course du millénaire, a no-rules round-the-world sailing event; Race (biology), classification of flora and fauna; Race (classification of human beings) Race and ethnicity in the United States Census, official definitions of "race" used by the US Census Bureau; Race and genetics, notion of racial classifications based on genetics. Historical definitions of race; Race (bearing), the inner and outer rings of a rolling-element bearing. RACE in molecular biology "Rapid ... General · Surnames · Television · Music · Literature · Video games
en.wikipedia.org/wiki/Race - [cache] - Live, Ask
- Publications | Human Rights Watch**
The use of torture, unlawful rendition, secret prisons, unfair trials, ... Risks to Migrants, Refugees, and Asylum Seekers in Egypt and Israel ... In the run-up to the Beijing Olympics in August 2008, ...
www.hrw.org/backgrounder/usa/race - [cache] - Ask
- Amazon.com: Race: The Reality Of Human Differences: Vincent Sarich ...**
Amazon.com: Race: The Reality Of Human Differences: Vincent Sarich, Frank Miele: Books ... From Publishers Weekly Sarich, a Berkeley emeritus anthropologist, and Miele, an editor ...
www.amazon.com/Race-Reality-Differences-Vincent-Sarich/dp/0813340861 - [cache] - Live
- AAPA Statement on Biological Aspects of Race**
AAPA Statement on Biological Aspects of Race ... Published in the American Journal of Physical Anthropology, vol. 101, pp 569-570, 1996 ... PREAMBLE As scientists who study human evolution and variation, ...
www.physanth.org/positions/race.html - [cache] - Ask
- race: Definition from Answers.com**
race n. A local geographic or global human population distinguished as a more or less distinct group by genetically transmitted physical
www.answers.com/topic/race-1 - [cache] - Live
- Dopefish.com**
Site for newbies as well as experienced Dopefish followers, chronicling the birth of the Dopefish, its numerous appearances in several computer games, and its eventual take-over of the human race. Maintained by Mr. Dopefish himself, Joe Siegler of Apogee Software.
www.dopefish.com - [cache] - Open Directory

On the left sidebar, there are sections for clusters, sources, sites, and a list of all results (238). The "clusters" section is currently selected. Other sections include sources (22), sites (14), and a definition section (13). The bottom left also has a "find in clusters" input field and a "Find" button.

Netflix

- Machine learning competition with a \$1 million prize



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

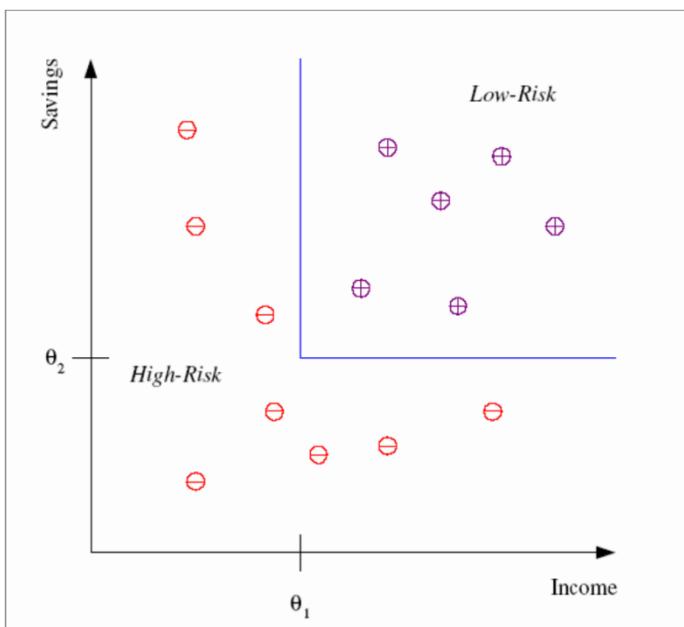
- Reinforcement learning (not in this course)
 - Given: Feedback (reward/cost) during trial-and-error episodes
 - Goal: Maximize Reward/minimize cost
 - Examples: learning to control a robot/car/helicopter etc.
 - see Master's course "Autonomously Learning Systems"

Learning from examples (data)

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

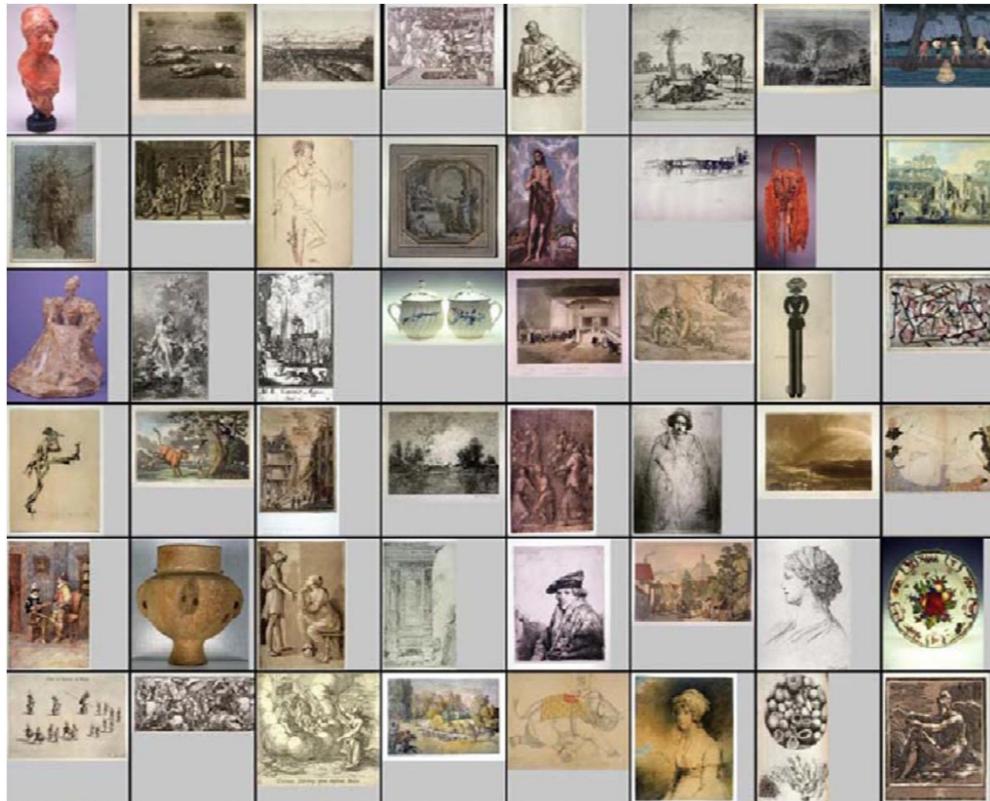


Applications:

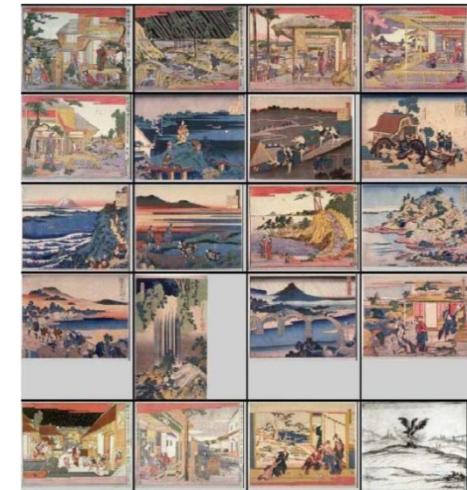
- *Spam filters*
- *Character recognition*
- *Speech recognition*
- *Collaborative filtering (predicting if a customer will be interested in an advertisement ...)*
- *Medical diagnosis*
- *...*

Unsupervised Learning: Example

- Find patterns and structure in data



Clustering art



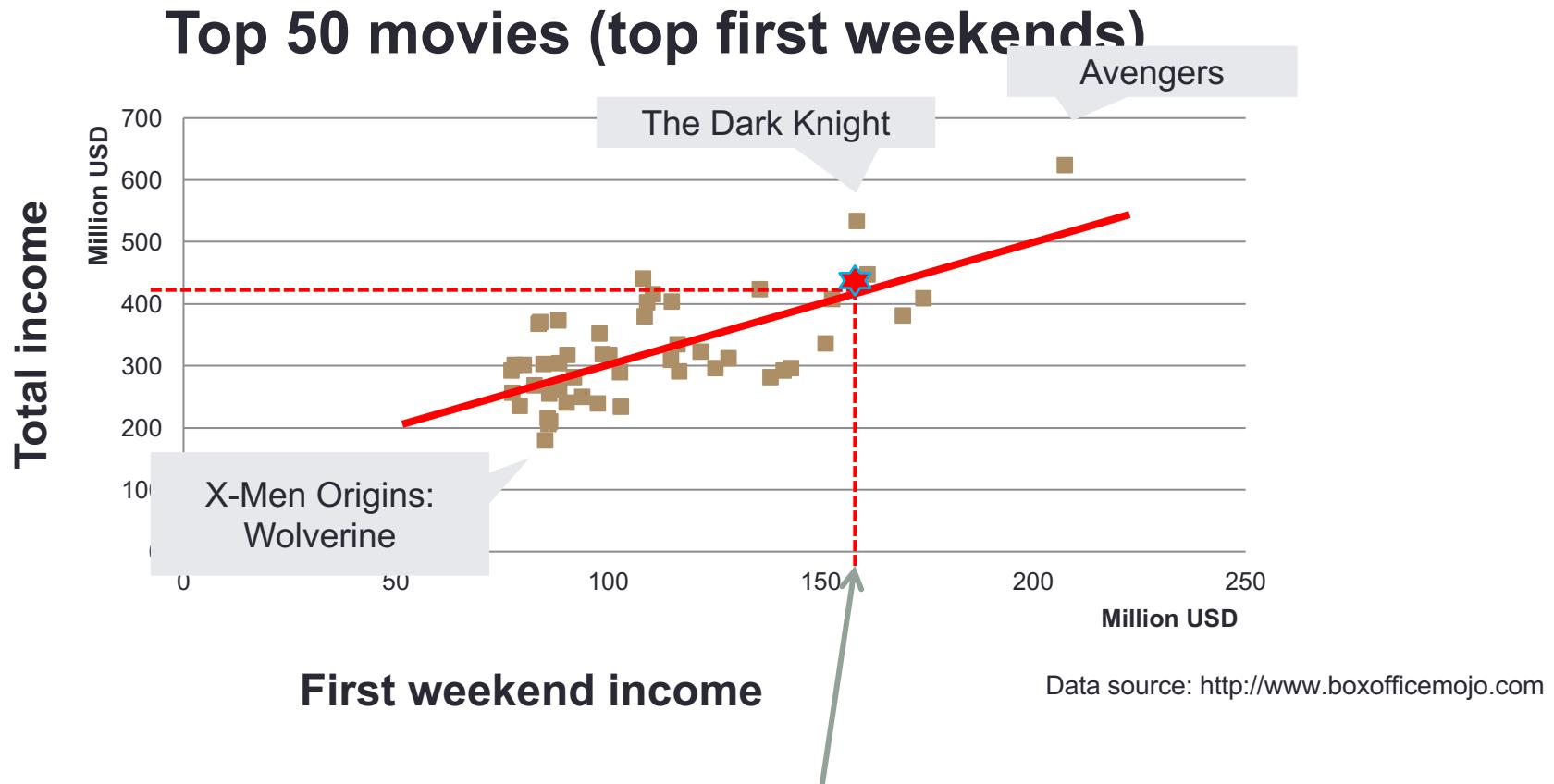
Unsupervised Learning: Applications

- 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



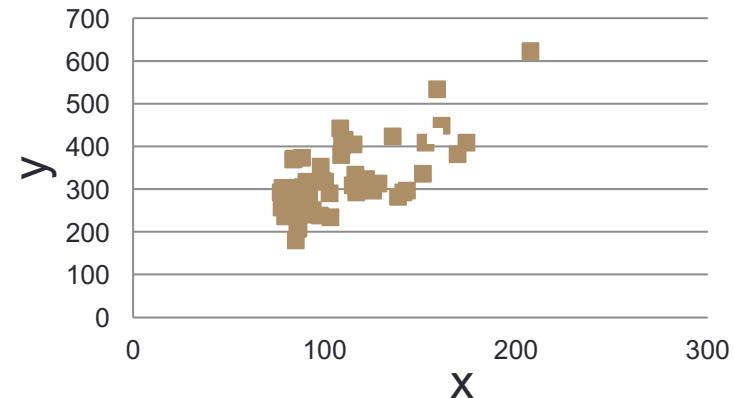
„The Hunger Games: Catching Fire“: 158 Mio. USD on opening weekend.
How much in total?

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

Simple regression example (cont'd)

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

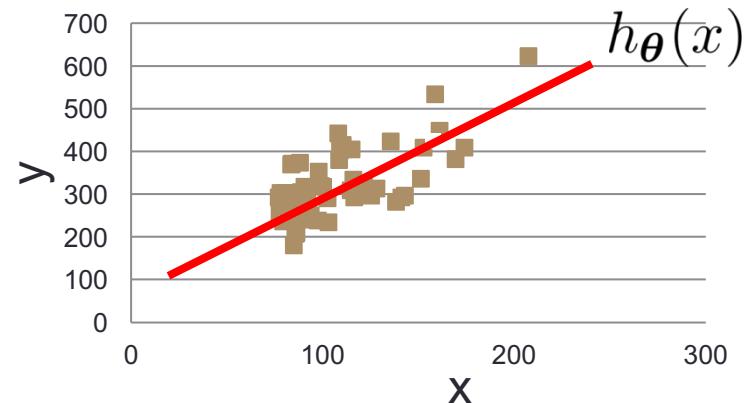
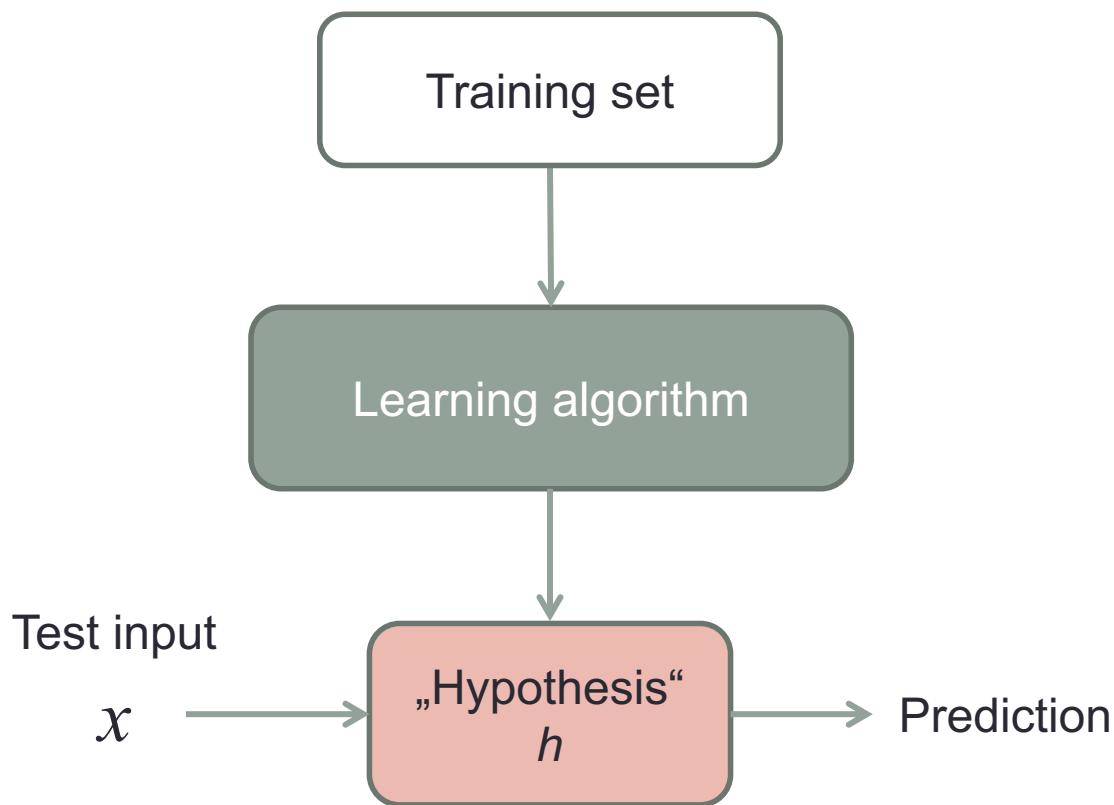
	i	First weekend $x^{(i)}$	Total $y^{(i)}$
Avengers	1	207	623
Iron Man 3	2	174	409
Harry Potter and the Deathly...	3	169	381
The Dark Knight Rises	4	161	449
The Dark Knight	5	158	533
	



m data points
(data samples)

Simple regression example (cont'd)

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



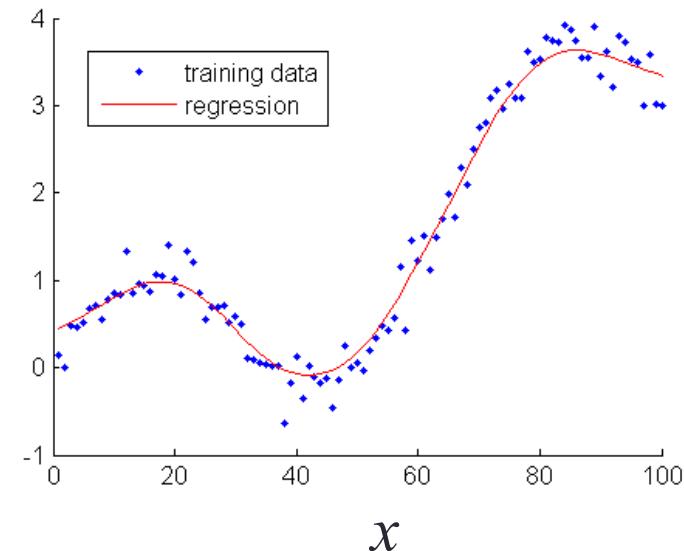
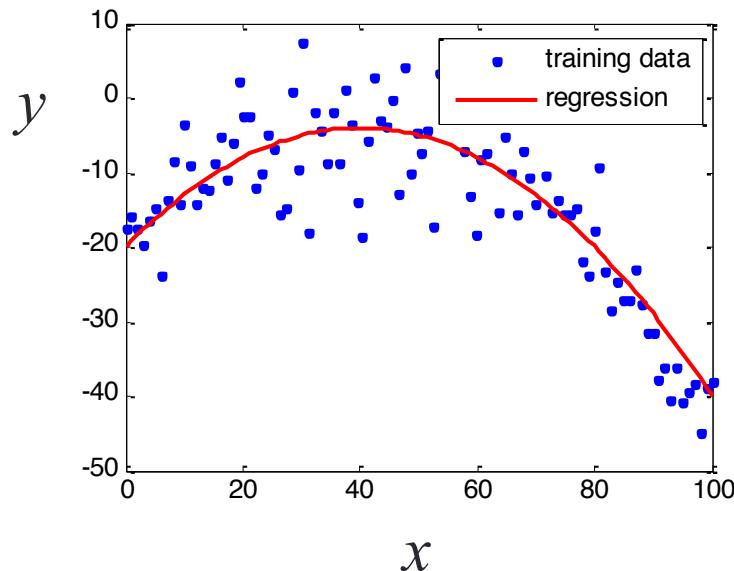
Hypothesis

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

Parameters

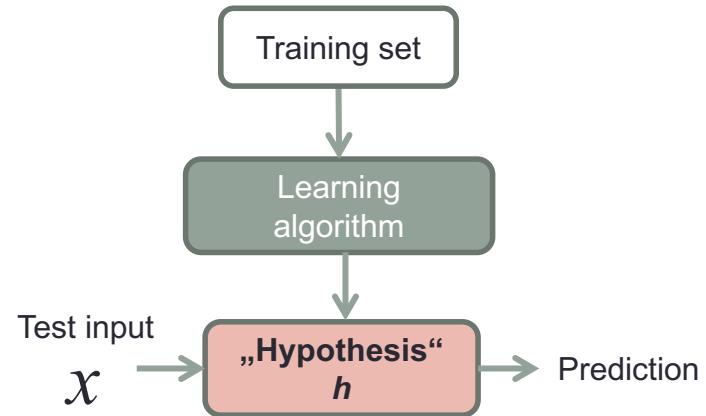
$$\theta = (\theta_0, \theta_1)$$

Non-linear regression

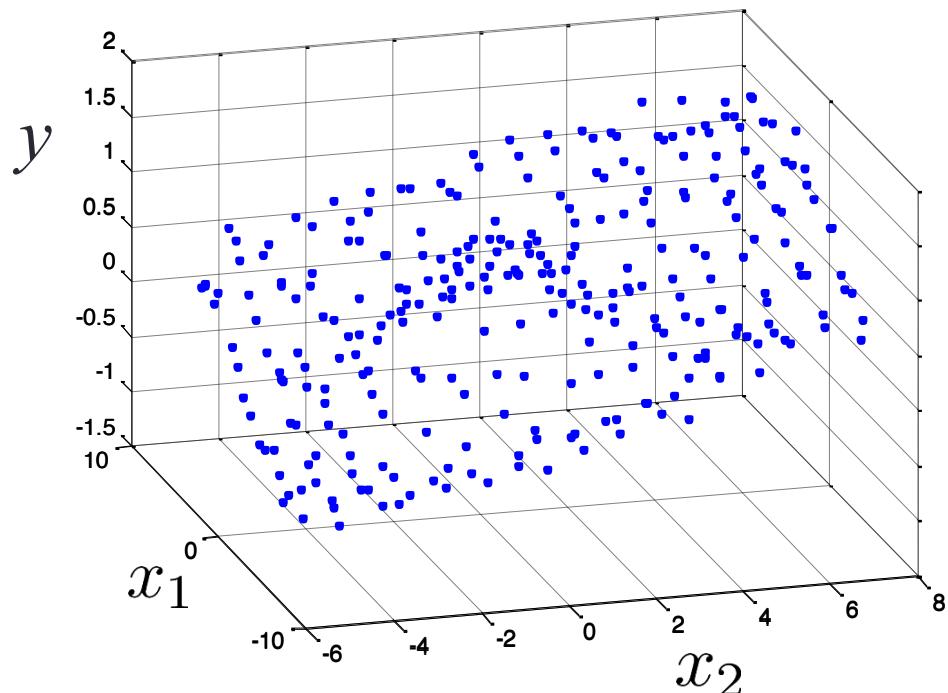


Non-linear hypothesis, for example

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

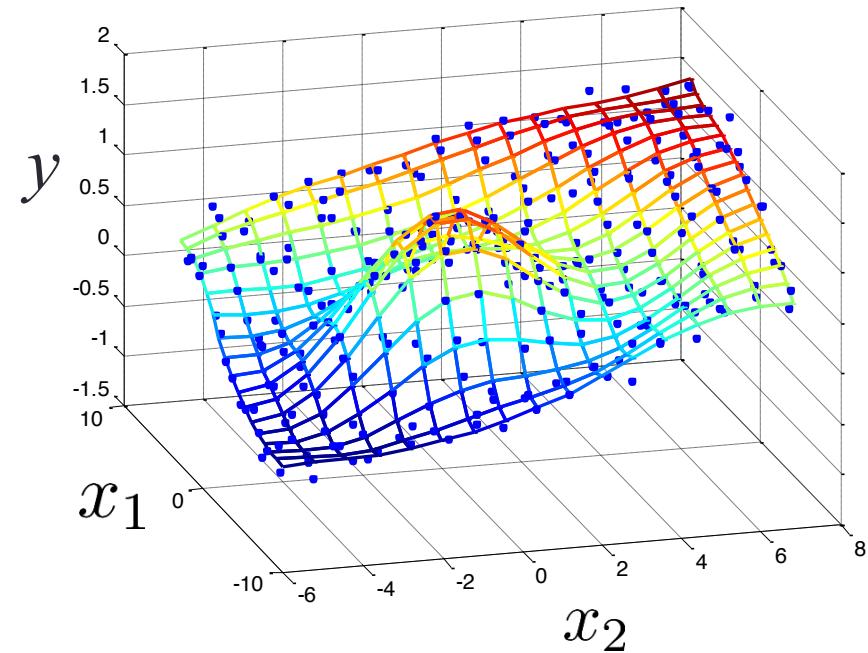


Regression with multiple inputs



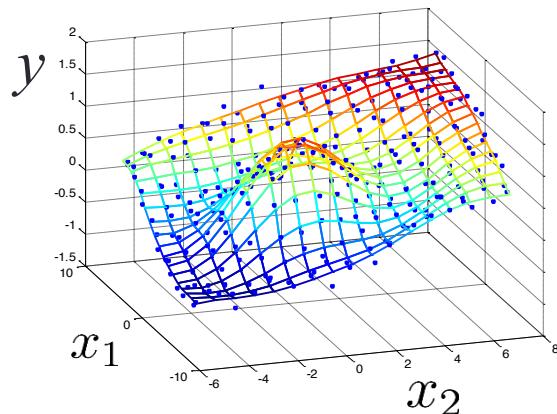
linear hypothesis

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

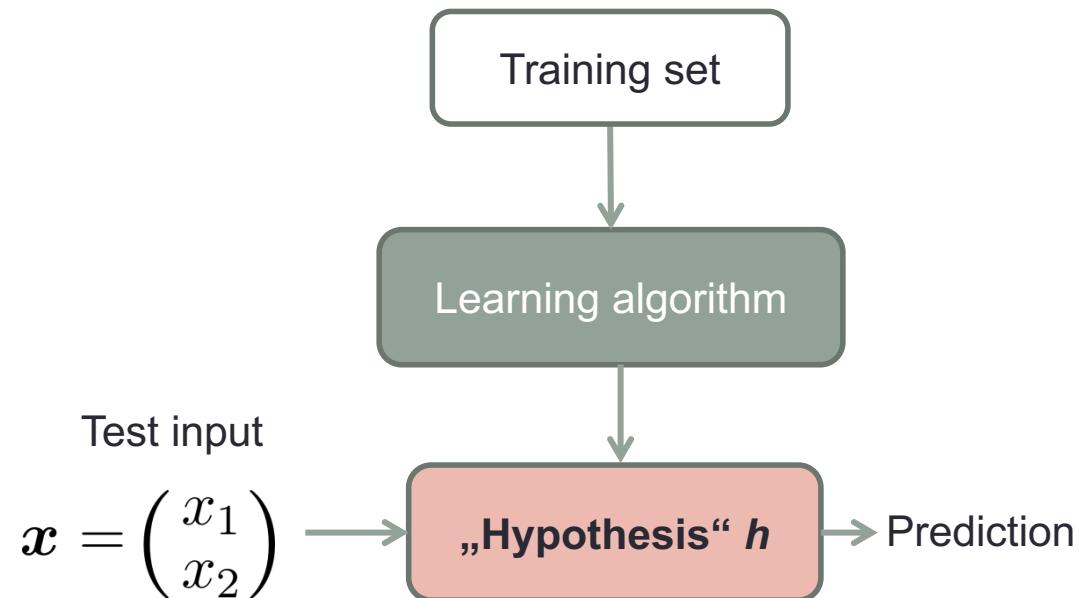


non-linear hypothesis

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

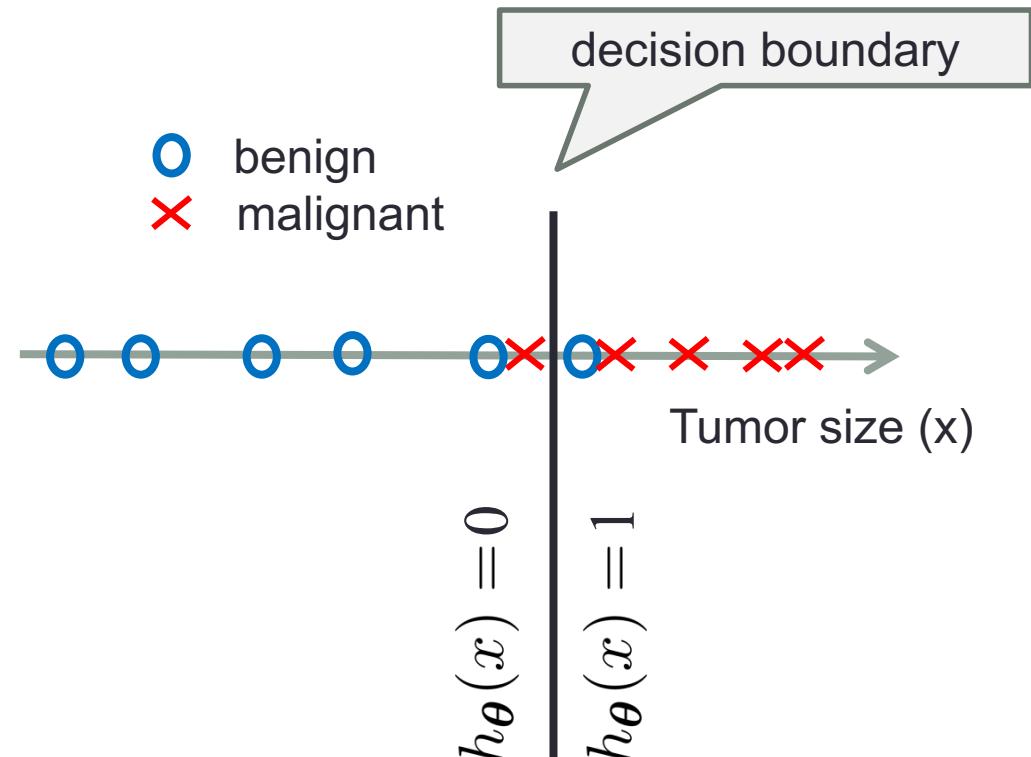


Simple classification example

„labeled data“

i	Tumor size (mm)		Malignant?
	x	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

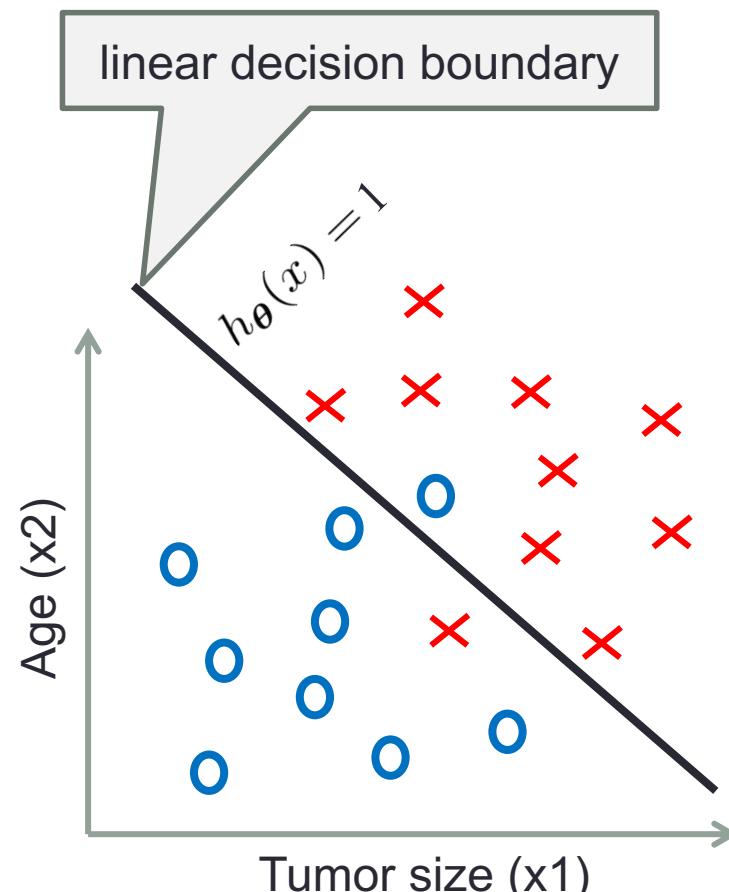


Example hypothesis:
 $h_{\theta}(x) = 1 \quad \text{if } x > \theta_0$

Classification with multiple inputs

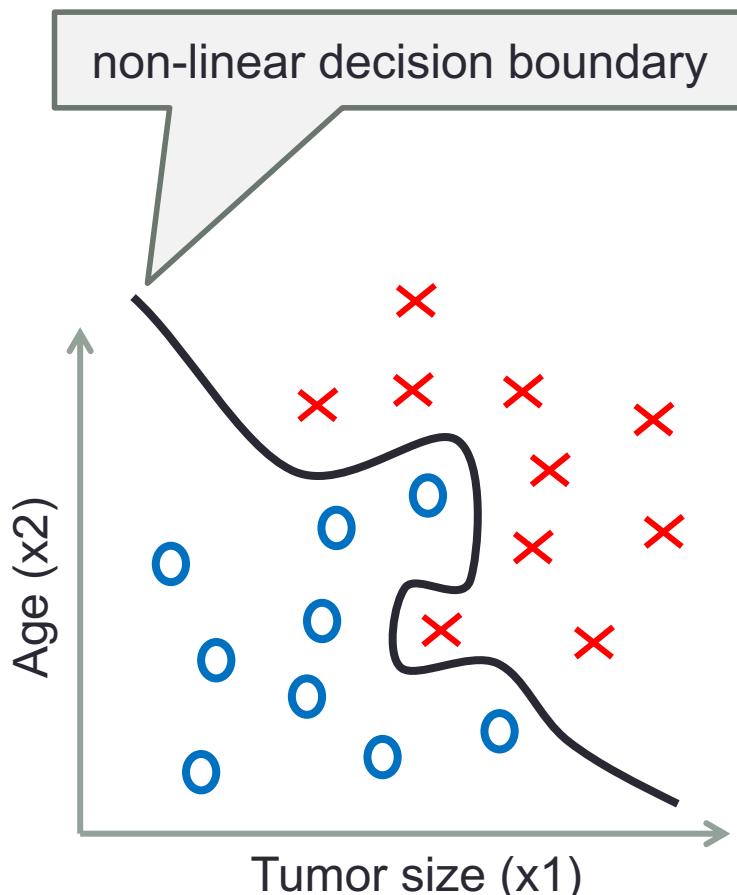
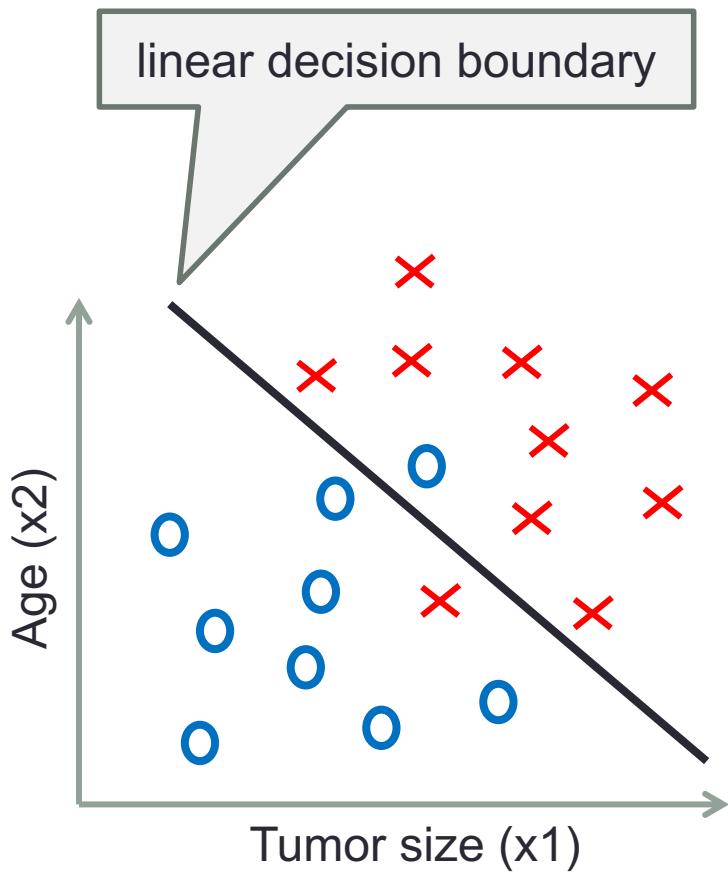
i

	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)



○ benign
✗ malignant

Non-linear classification



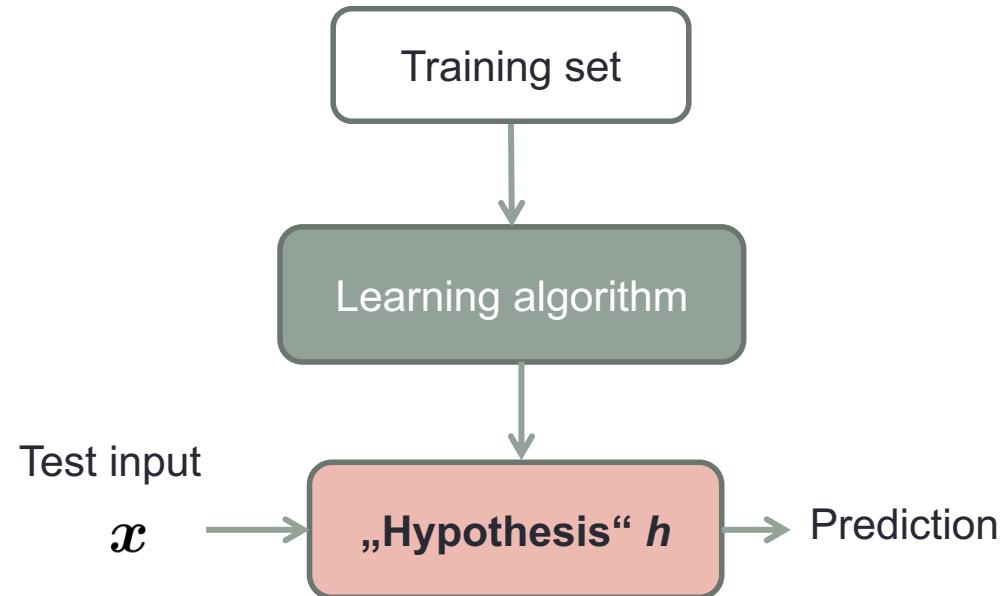
Both hypotheses fit the data quite well.
Which one fits the **training data** better?
Which one would you trust more for **prediction**?

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?

- Linear regression
- Gradient descent
- Non-linear basis functions