Machine Learning

CS-433

Martin Jaggi & Rüdiger Urbanke
EPFL

20th Sept 2016

Course Logistics

Assessment

- **Project 1 (10%)**, due Oct 31st
- **Project 2 (30%)**, due Dec 22nd
- **Final exam (60%)**

Course Logistics

Lectures

tentative
schedule

Date	Topics Covered	Exercises	Projects
20/9	Introduction		
22/9	Linear Regression	Lab 1	
27/9	Cost Functions		
29/9	Optimization	Lab 2	
04/10	Least Squares, ill-conditioning		
06/10	Maximum Likelihood, Overfitting	Lab 3	
11/10	Cross-Validation		
13/10	Bias-Variance decomposition	Lab 4	
18/10	Classification		
20/10	Logistic Regression	Lab 5	
25/10	Generalized Linear Models		
27/10	k-Nearest Neighbor	Lab 6	
01/11	Support Vector Machines		Proj. 1 due 31.10.
03/11	Kernel Regression	Lab 7	
08/11	Unsupervised Learning		
10/11	k-Means	Lab 8	
15/11	Gaussian Mixture Models		
17/11	EM algorithm	Lab 9	
22/11	Matrix Factorizations		
24/11	Recommender Systems	Lab 10	
29/11	SVD and PCA		
01/12	SVD and PCA	Lab 11	
06/12	Neural Networks		
08/12	Multi-Layer Perceptron	Lab 12	
13/12	Neural Networks, CNNs		
15/12	Decision Trees, Random Forests	Lab 13	
20/12	BayesNet and Belief Propagation		

Course Logistics

Lectures

Tuesday 8:15 - 10:00, Room: CE1
Thursday 8:15 - 10:00, Room: CE4

Clicker:
For some active participation in the lectures, please point your browser to this SpeakUp room

We provide PDF lecture notes on our webpage and also on Nota Bene so you can comment & discuss them.

Teaching material mostly by Emtiyaz Khan

(SpeakUp room)

seance.epfl.ch

Room Nr. 89063

Q1: Are you new to EPFL this semester

A. **yes**

B. **no**

(SpeakUp room)

seance.epfl.ch

Room Nr. 89063

Q2: Are you a student of

A. **IC** (Comp Sci.)
B. **SB** (Basic Sciences)
C. **STI** (Engineering)
D. **SV** (Life Sciences)
E. **ENAC** (Arch., Env. Sci.)
F. **CDM** (Management)
G. **CDH** (Humanities)
H. **UNIL**
I. **others**

Course Logistics

Exercises

Thursday 14:15 - 16:00
Rooms: INF119 (A-Ch), INJ218 (Ci-Kj), INM11 (Ko-Pe), INM202 (Pi-Z), according to last-name.

All labs and projects will be in **Python** this year.
See the first lab to get started.

Code Repository for Labs: github.com/epfml/ML_course

Course Logistics

TA team

contact us: epfmlcourse@gmail.com

Mohamad Dia

Ksenia Konyushkova

Victor Kristof

Taylor Newton

Farnood Salehi

Benoît Seguin

Frederik Kunstner

Fayez Lahoud

Tao Lin

Arnaud Miribel

Vidit Vidit

Course Logistics

Projects

- **Project 1 (10%)**, due Oct 31st
- **Project 2 (30%)**, due Dec 22nd

Real-world problems, Python, Groups of 3 Students

What to expect?

- **overview** over ML
- basic **understanding** of most important ML **methods** and **concepts**
- experience how ML is done on a **practical** problem

What **not** to expect?

- becoming an **expert** data scientist / ML researcher / statistician / large-scale ML hacker / ML software expert / domain expert
- learn about this week's newest **hack in deep learning** you saw on blog *xyz*.
- build your own **self-flying** helicopter

What is Machine Learning?

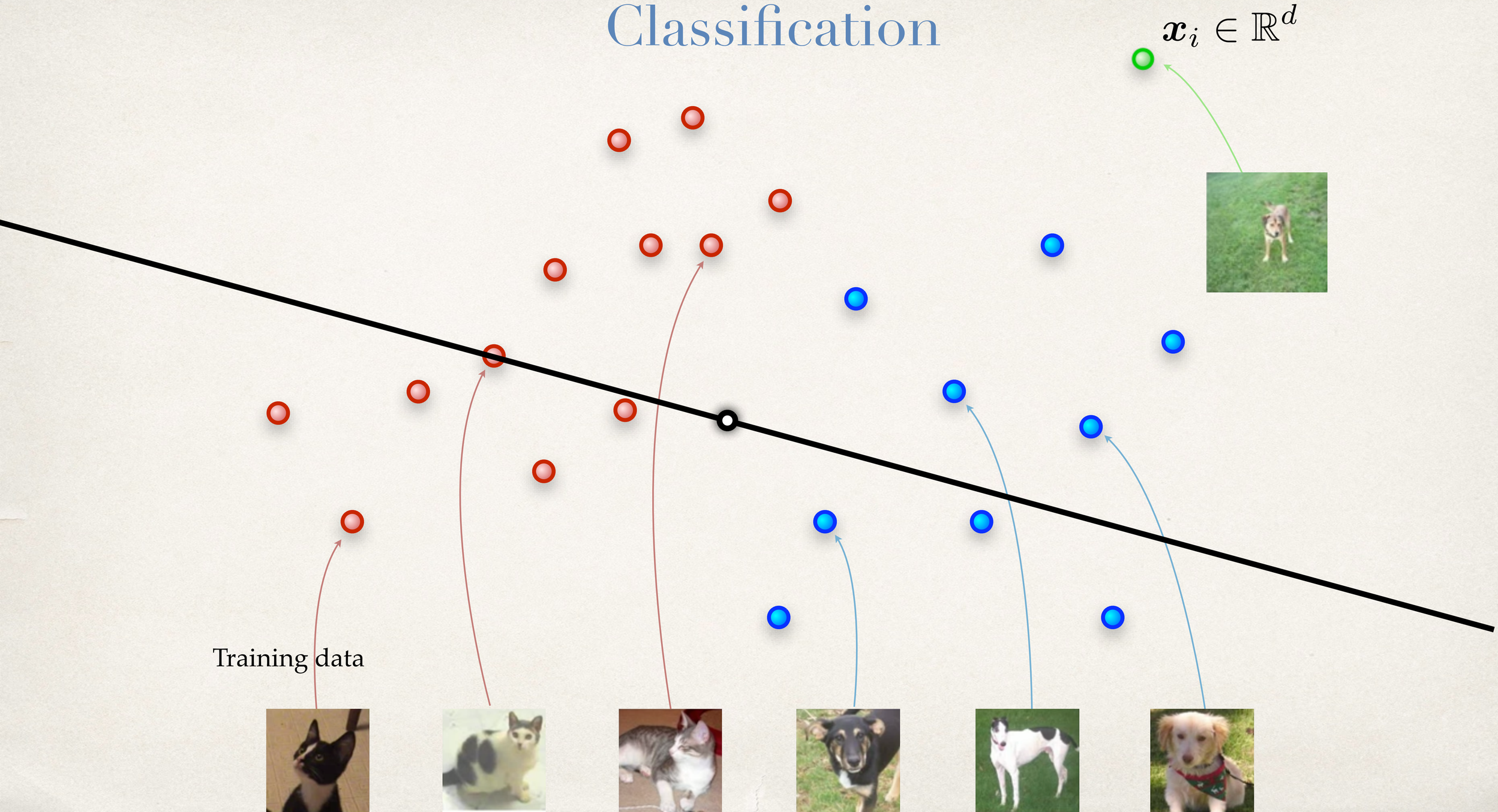
What is Machine Learning?

software that can

learn from data

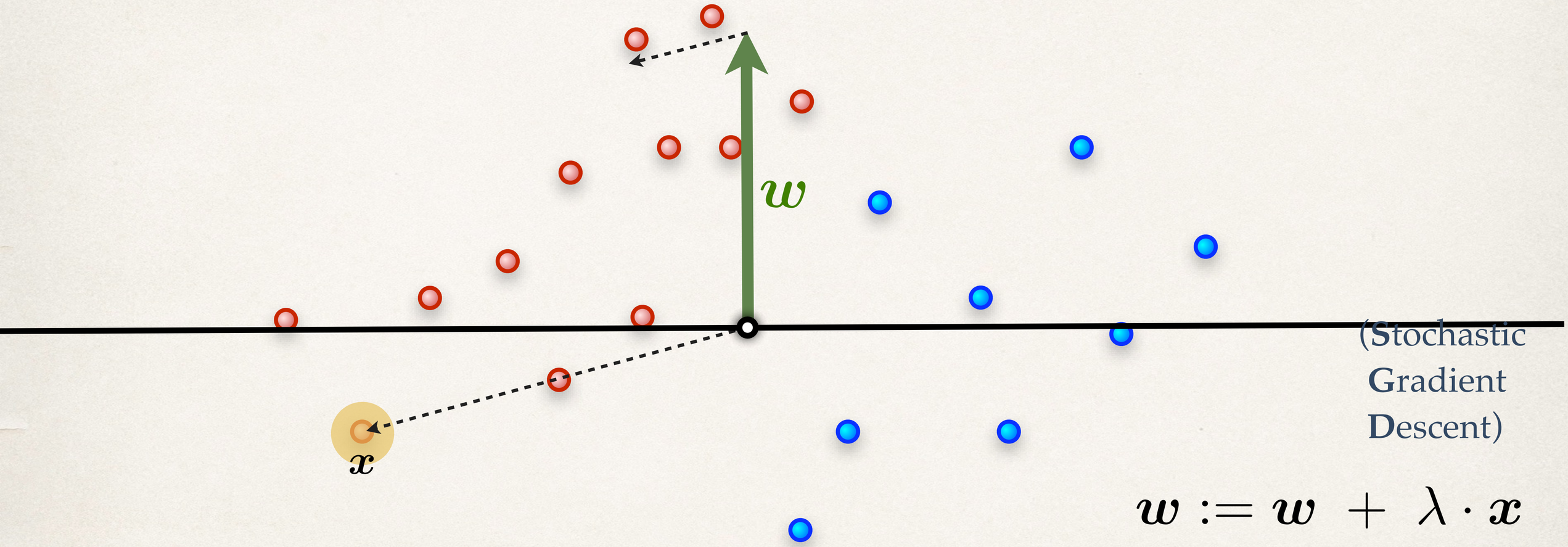
image source

Classification



The Learning Algorithm

$\boldsymbol{x}_i \in \mathbb{R}^d$



(Stochastic Gradient Descent)

$$\boldsymbol{w} := \boldsymbol{w} + \lambda \cdot \boldsymbol{x}$$

iteration cost: O(d)

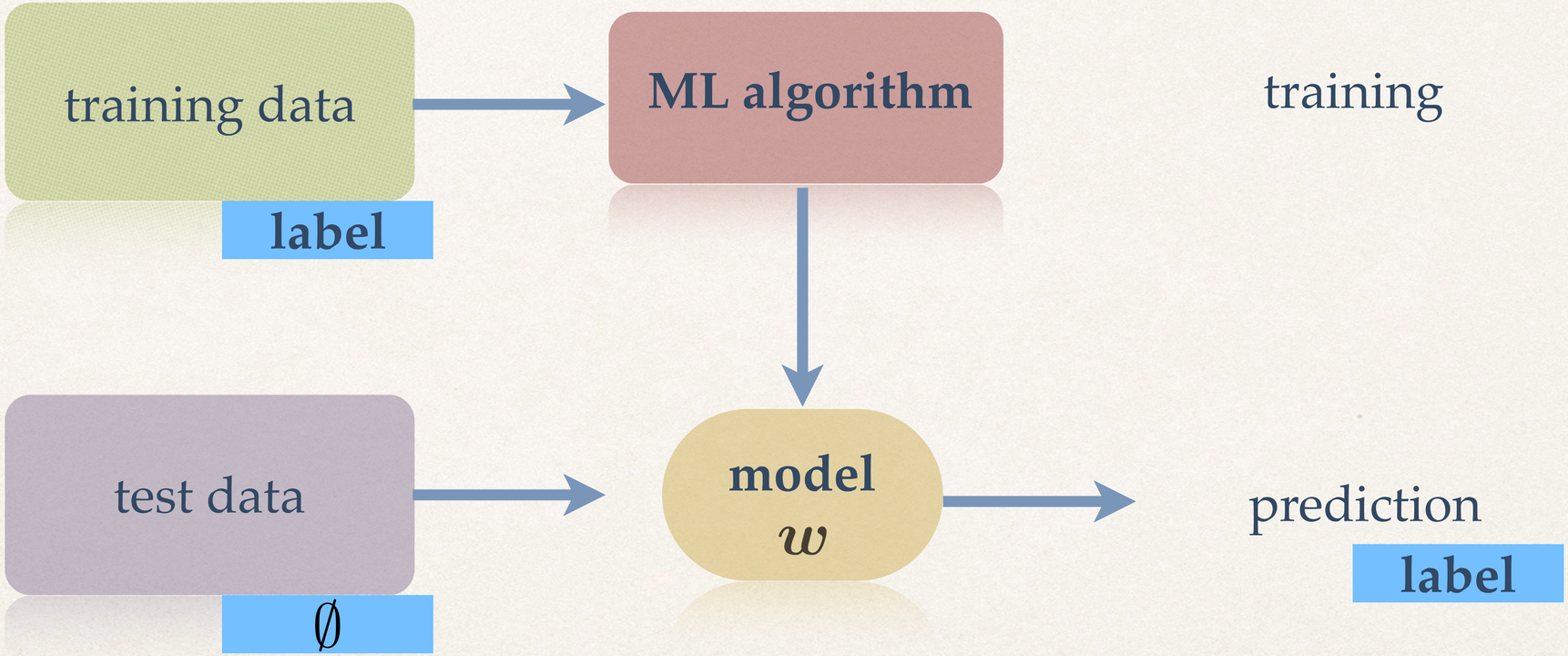
Perceptron

(Rosenblatt 1957)

Support-Vector-Machine

(Cortes & Vapnik 1995)

Machine Learning Fundamentals



what Evaluation Metric?

towards…

understanding intelligence

?

if-then-else

$\neq$

intelligence

towards…

understanding intelligence

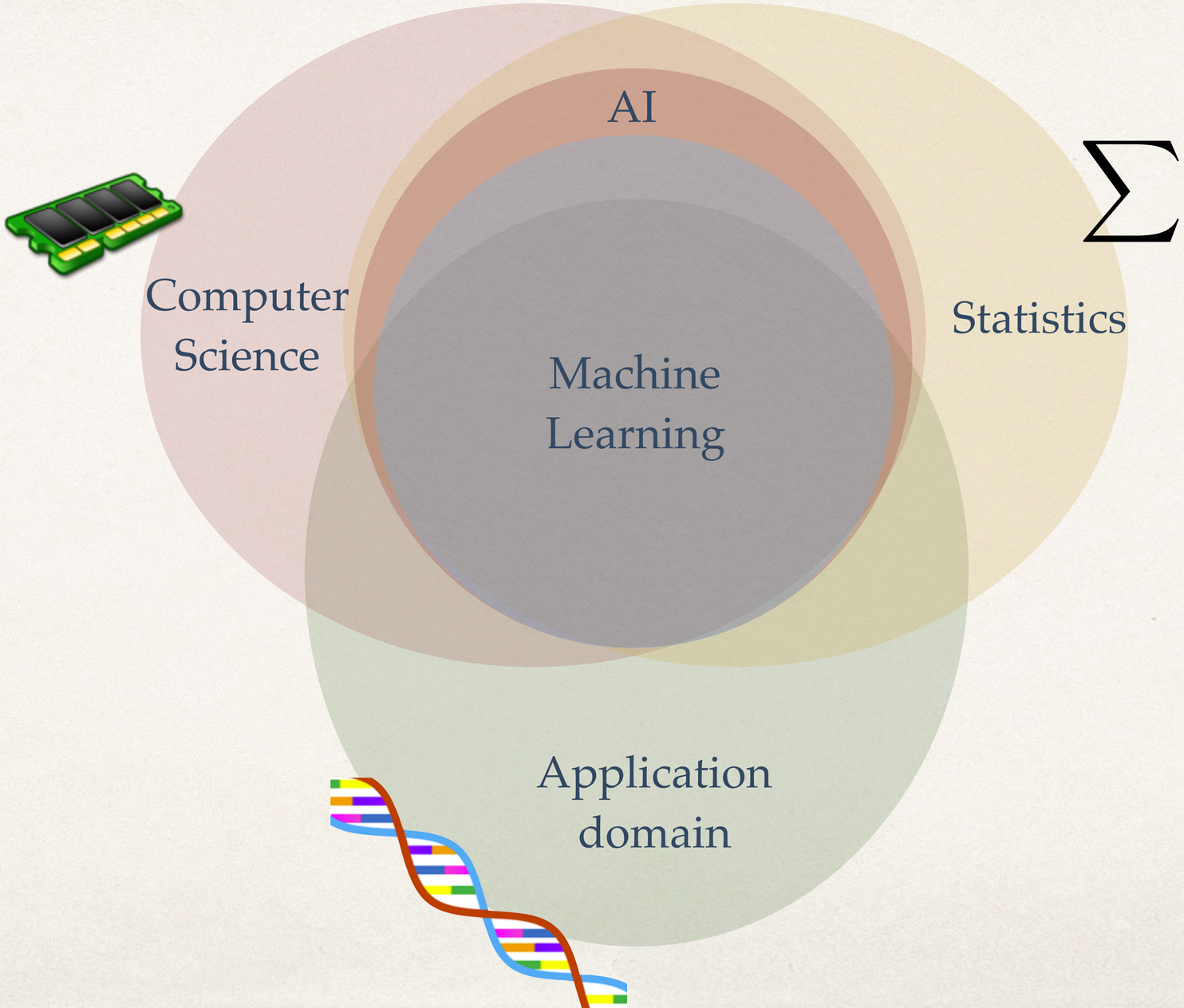
?

Machine Learning

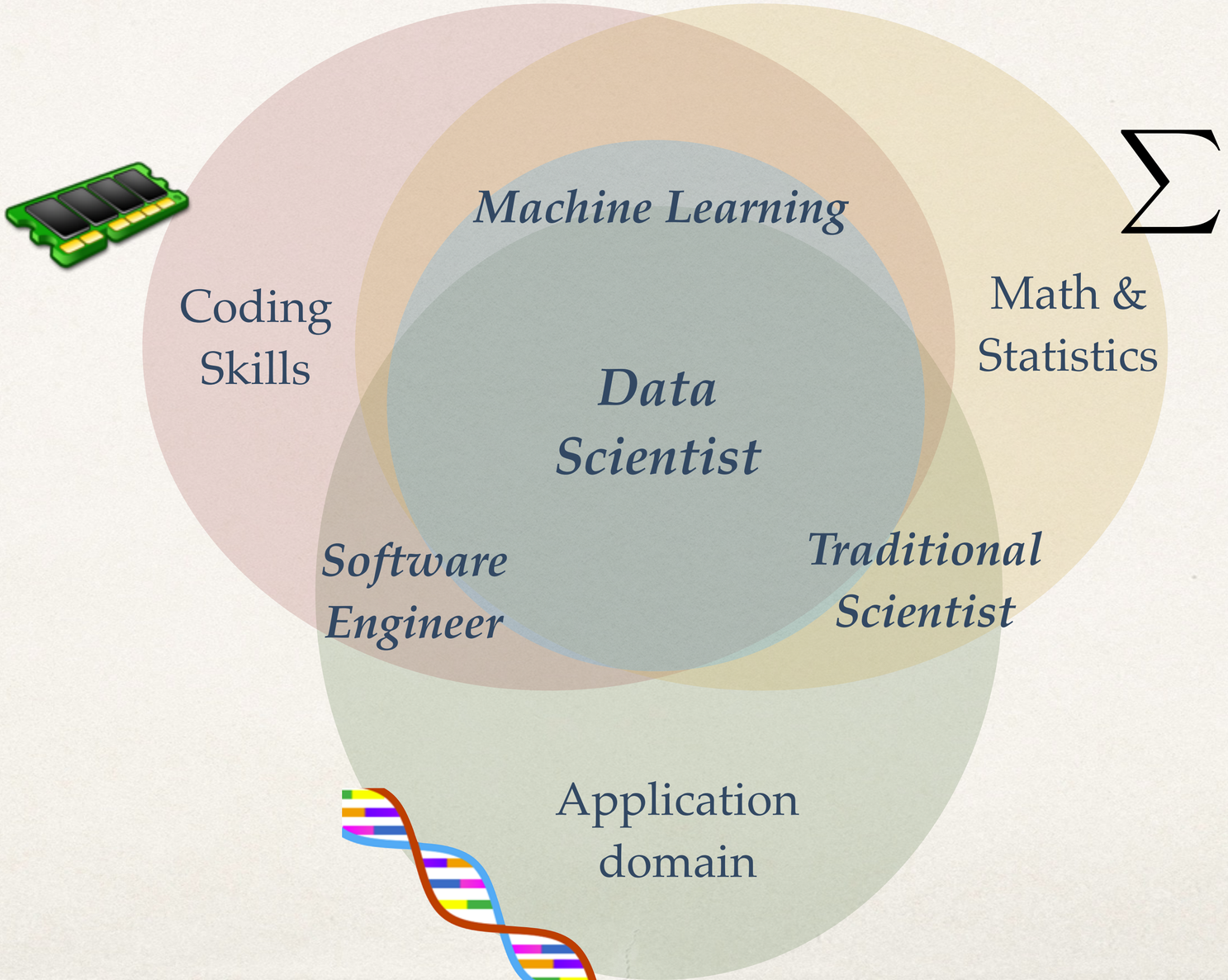
vs

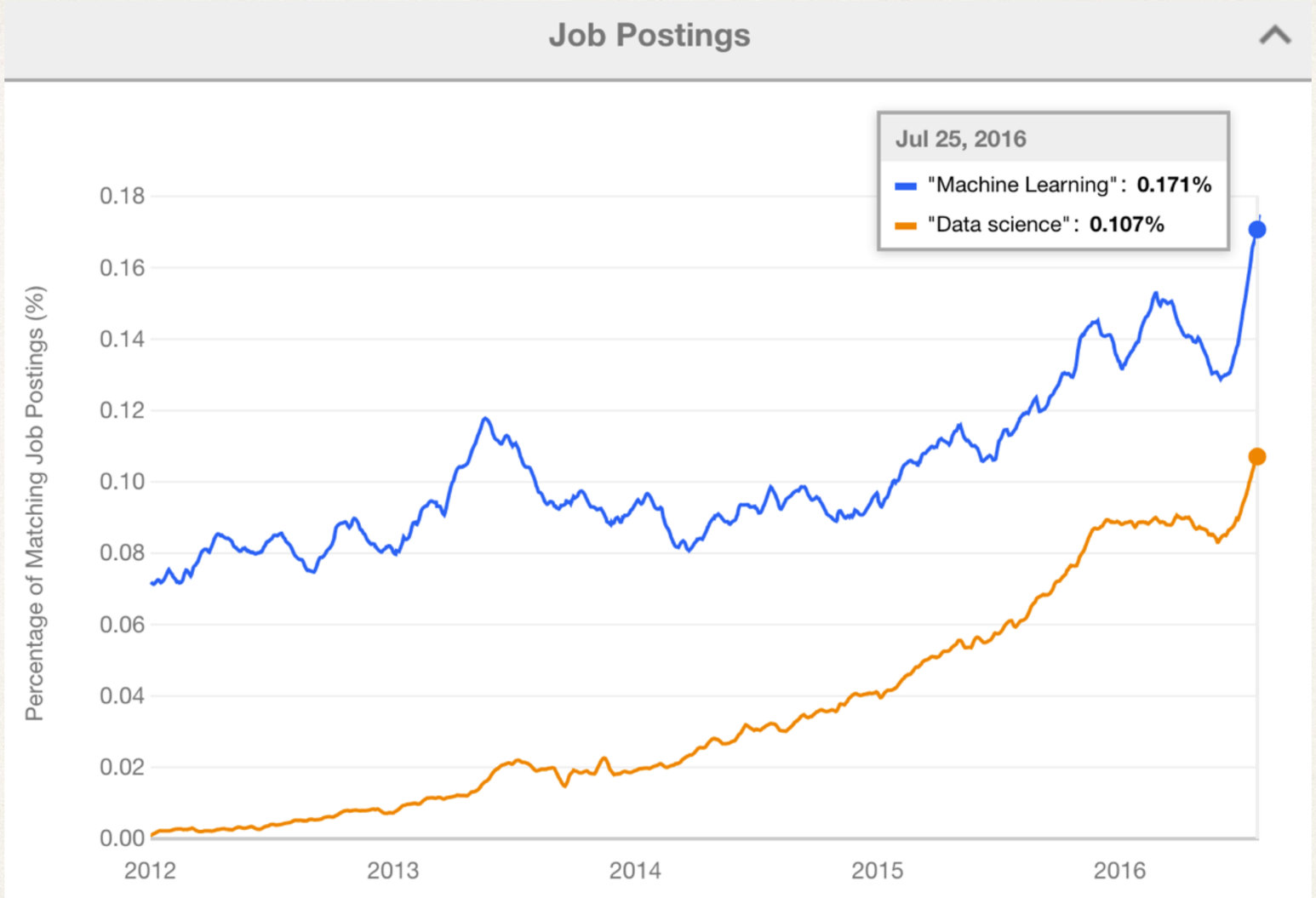
Neuroscience / HBP

What is the difference between Artificial Intelligence, Data Mining, Statistics, Machine Learning?


AI
Computer Science
Statistics
Machine Learning
Application domain

Job Skills


Machine Learning
Coding Skills
Math & Statistics
Data Scientist
Software Engineer
Traditional Scientist
Application domain



"…salary of a top AI researcher
~ top NFL quarterback"

Peter Lee, Microsoft

www.indeed.com/jobtrends

why ML?

Applications

Industry Applications

- **not only** the 'usual suspects'

facebook

amazon.com

Google

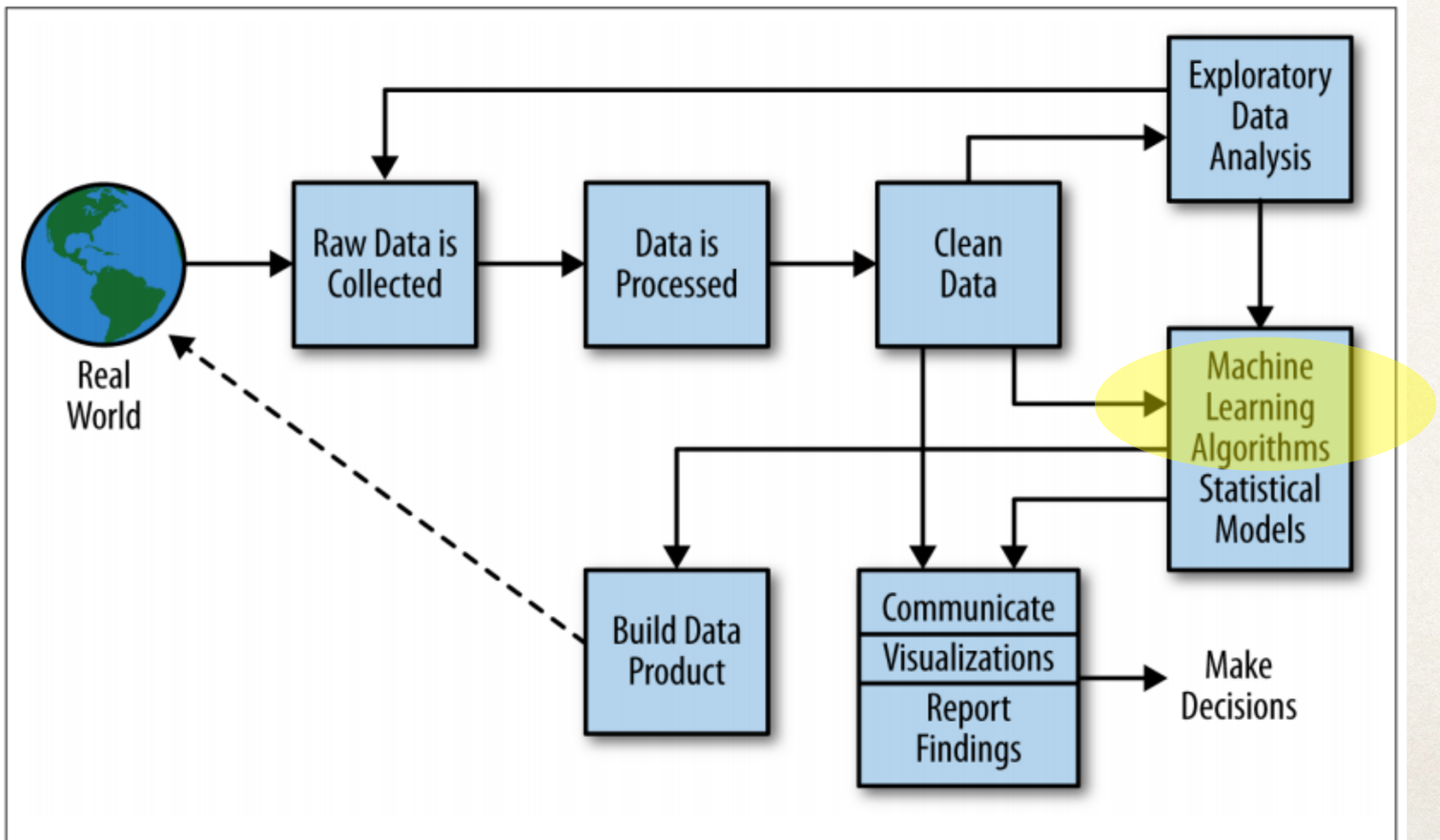
- majority of industries, originally not 'digital':
  - agriculture, car & bike sharing, booking.com, manufacturing, sports, personalized health, call centers, …

Applications in Other Sciences

- increasingly data driven
  - … Psychology, Economy, Social sciences
  - science of X ➔ *digital* science of X

en.wikipedia.org/wiki/Index_of_branches_of_science

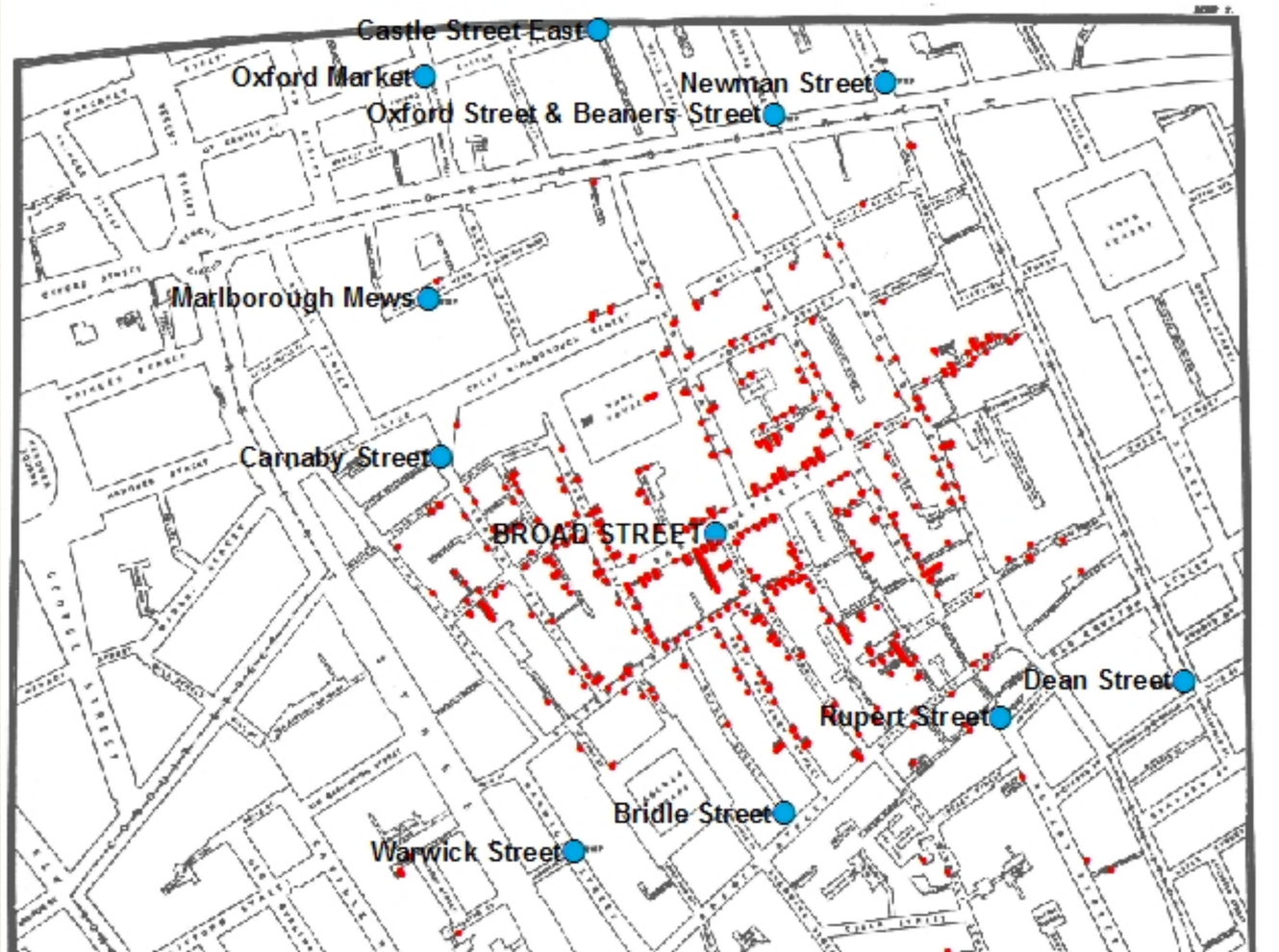
ML is only a small part!



History

- ML is **not new!**

London 1854
cholera outbreak



History

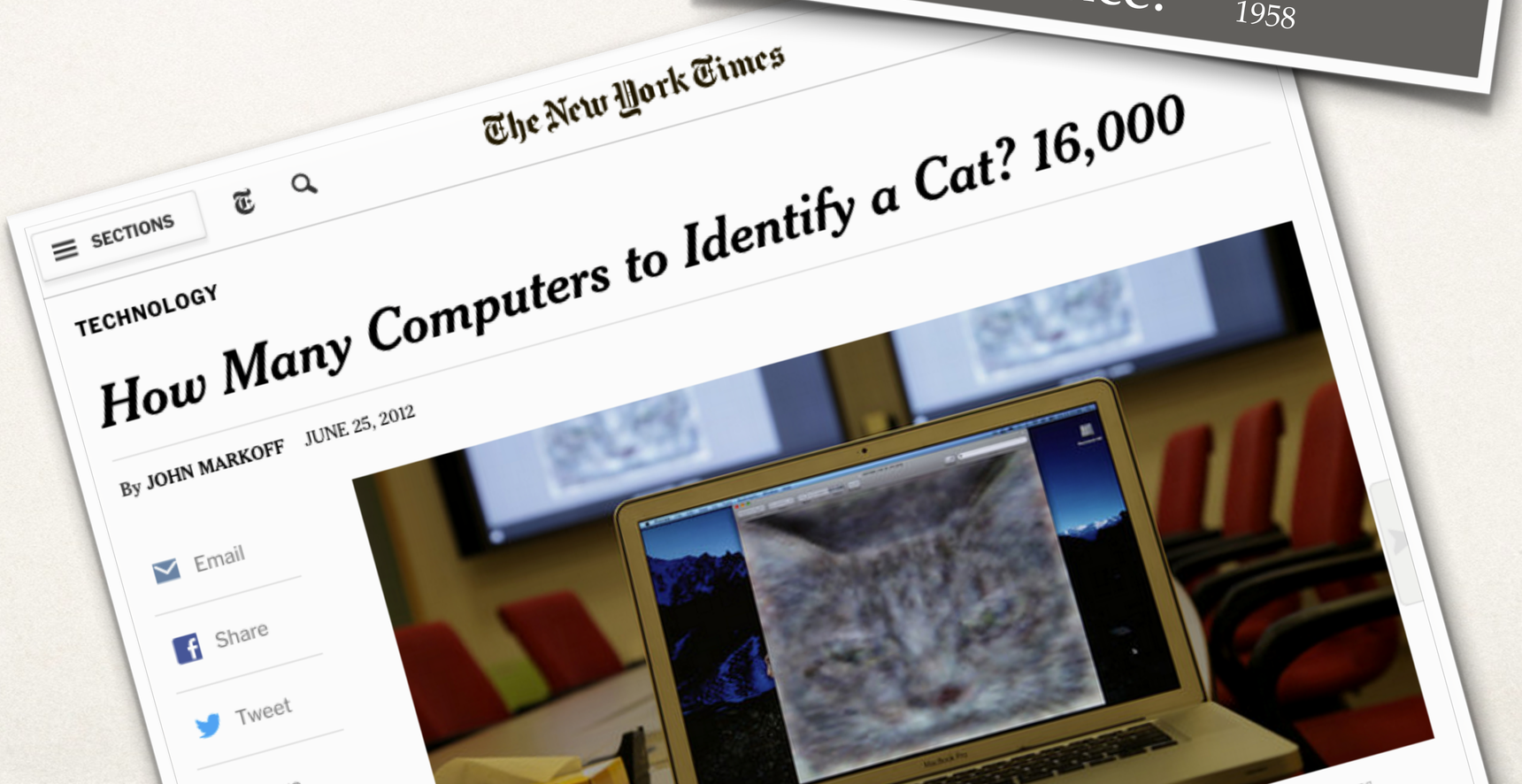
- ML is **not new!**
- the early days - 1950ies and 1960ies
  - Neural networks
  - Turing

What has changed?

1950s: 10^3 FLOPS

2010s: 10^{15} FLOPS

"the embryo of an electronic computer that … will be able to walk, talk, see, write, reproduce itself and be conscious of its existence." *1958*



Challenges

- Hype
  - cycles of AI popularity
- Data Privacy

amazon.de

Hallo! Melden Sie sich an, um persönliche Empfehlungen zu erhalten. Neukunde? Bitte hier starten.

Mein Amazon | Sonderangebote | Wunschzettel | Gutscheine | Geschenke

Alle Kategorien ansehen | Suche | Sport & Freizeit

Sport & Freizeit | Erweiterte Suche | Camping & Outdoor | Fitness | Fußball | Golf | Pulsuhren & GPS | R

Größeres Bild

Für Kunden: Stellen Sie Ihre eigenen Bilder ein.

Baseballschläger Aluminium "32" von Tysonz

von Unbekannt

Noch keine Kundenrezensionen vorhanden: Schreiben Sie die erste!

Preis: EUR 21,90

Alle Preisangaben inkl. MwSt.

Auf Lager.

Verkauf und Versand durch **knockout24**. Für weitere Informationen klic
Verkäufernamen.

Noch 4 Stück auf Lager.

2 neu ab EUR 21,90

Kunden, die diesen Artikel gekauft haben, kauften auch

Pfeffer 2442 KO Verteidigungsspray JET 40 ml von Pfeffer KO

(12)

EUR 5,22

Sturmhaube 3-loch Acryl schwarz von Unbekannt

(13)

EUR 1,38

Pfefferspray KO-FOG 40ML von Pfeffer KO

(17)

EUR 5,19

KO-CS Verteidigungsspray 40 ml von KO-CS

(33)

EUR 4,22

Challenges

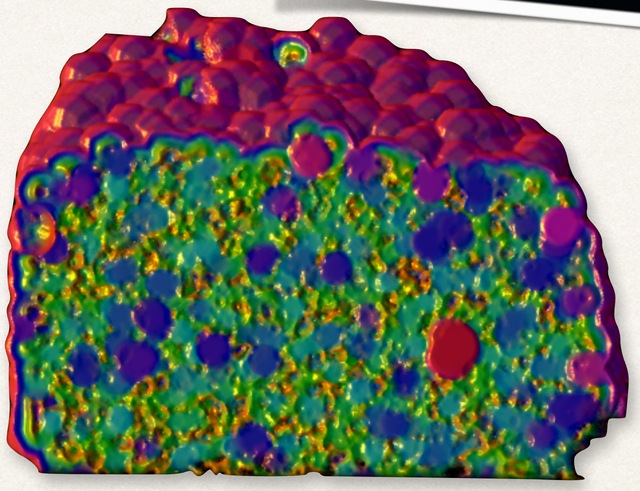
- Hype
  - cycles of AI popularity
- Data Privacy
- Lack of Interpretability
  - example: medical applications of deep learning
- Threats from Super-human AI?
  - see Nick Bostrom
  - social implications

need: Scientific Method, Reproducible Research, Open Source and Open Data

Applications, cont.

Image Data

- Astronomy
- Face recognition
- 2D + 3D medical imaging
- OCR
- self-driving cars



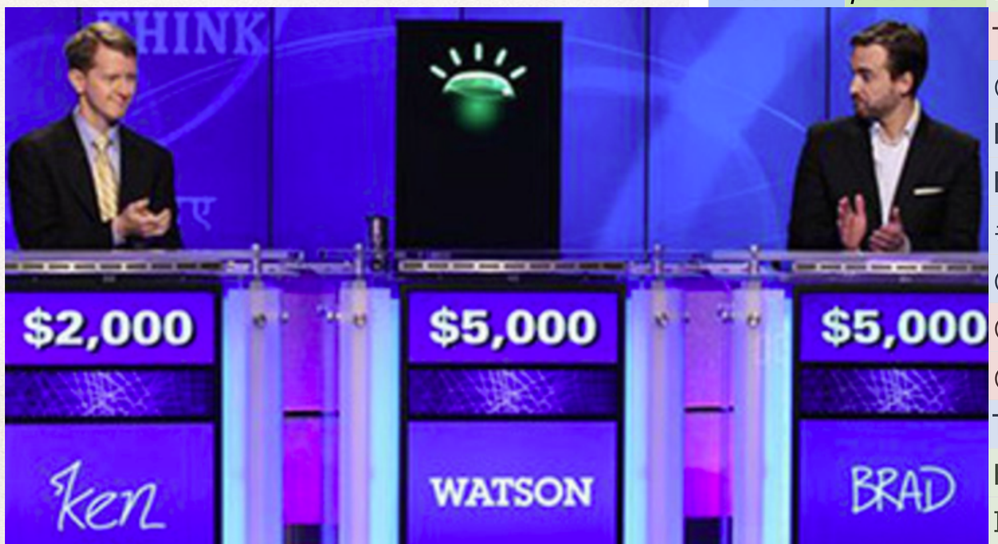
how-old.net

How-Old.net
HOW OLD DO I LOOK? #HowOldRobot

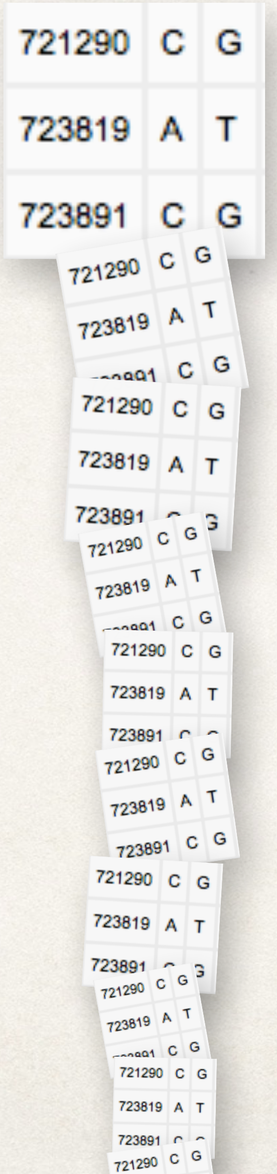
Text Data

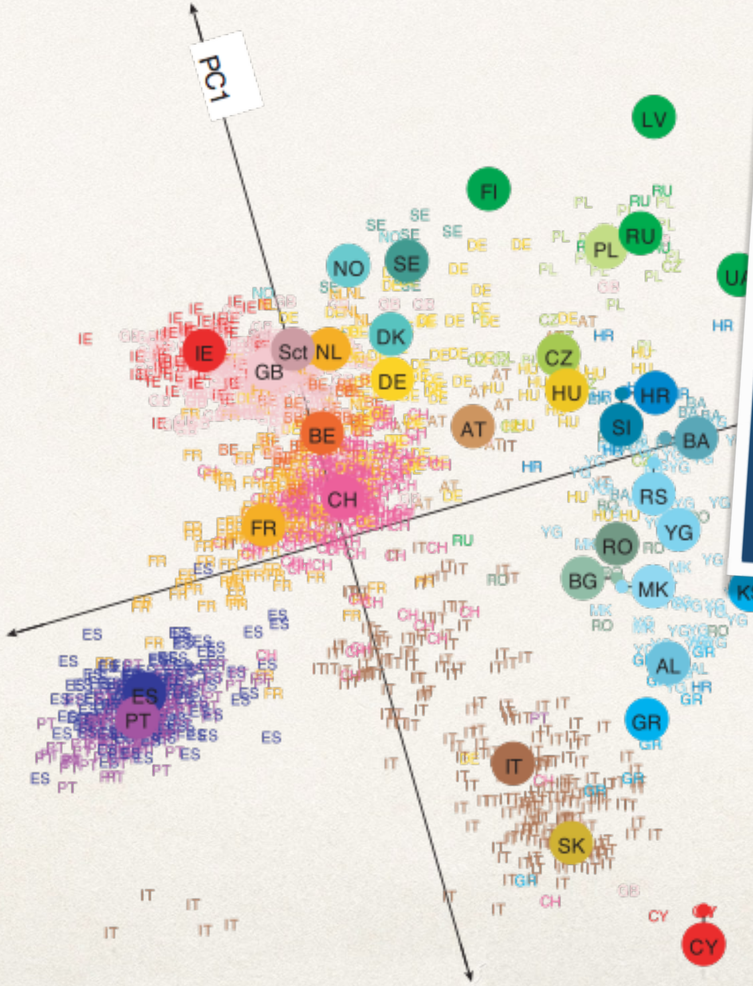
- Spam Detection
- User Content
- Medical Data

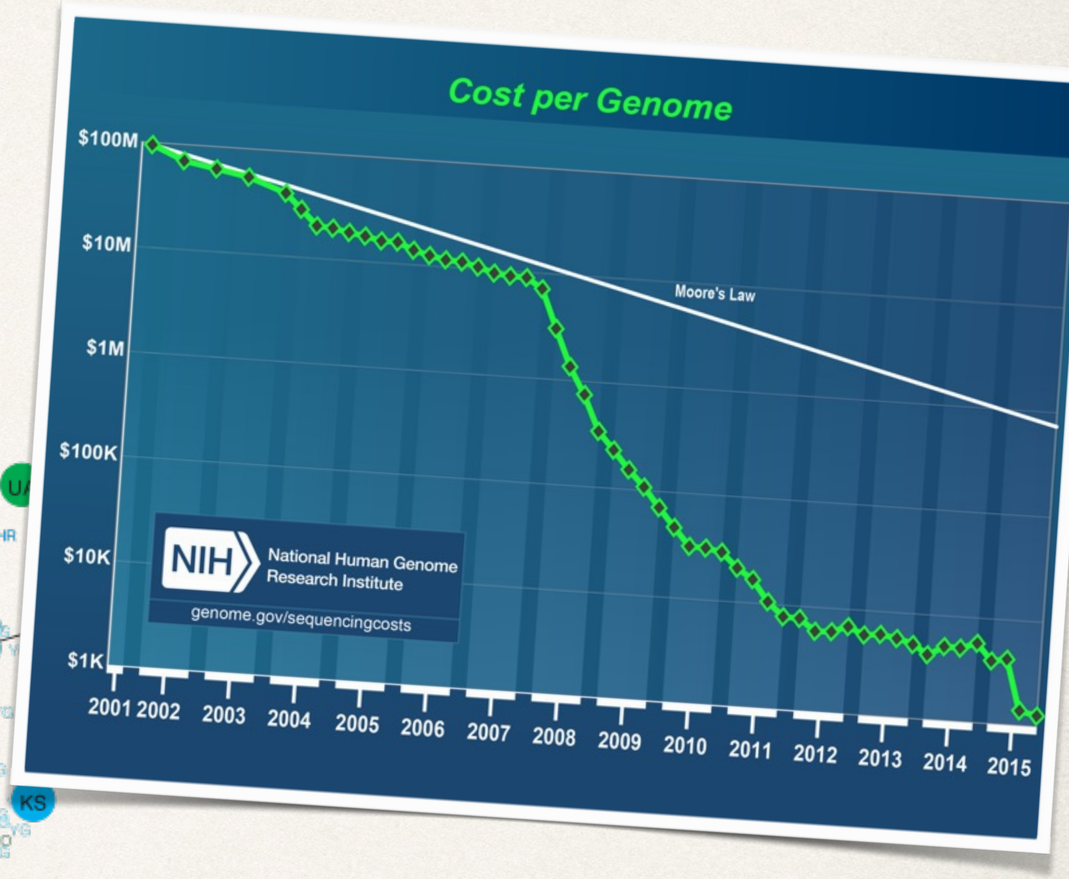
negative	*neutral*	But i wanna wear my Concords tomorrow though but i don't feel like it
positive	*neutral*	Gonna watch Grey's Anatomy all day today and tomorrow(:
negative	*neutral*	@CoachVac heey do you know anything about UVA's fallll fest loll they invited me
neutral	*neutral*	@DustyEf when that sun is high in that Texas sky, I'll be buckin it to county fair. A
neutral	*positive*	Up 20 points in my money league with Vernon Davis and L. Fitz still to go tomorro
neutral	*positive*	DEEJAYING this FRIDAY in THE FIRST CHOP it's CHRIS actual SMITH with a smash
		The Rick Santorum signing that was scheduled for tomorrow at the Books A Million
		@dreami9 lol yep looks like it! Was after El Clasico on Sunday. I didn't like her lol
		Back in Stoke on Trent for the 2nd time today!
		First Girls Varsity Basketball Game tomorrow at 6:00 pm Then Football Senior nigh
		#UFC lightweights @Young__Assassin VS @jamievarner set for TUF 16 Finale on t
		@OOOOO_WEEEE slide thru sometime this weekend ill have somethin yu can sip o
		@DannyB618 Sure absolutely-- I meant out of the Bachmann, Perry, Santorum, H
		@RichardGordon48 re Levein discussion on Wed. Can't keep changing boss, but he
		Today In History November 02, 1958 Elvis gave a party at his hotel before going o
		Hustle cause you got to then kick back n party everyday like its Fri
		I can't sleep. Way too exited about Vancouver tomorrow! I'm like a kid at Christma



Medical: Genetic Data



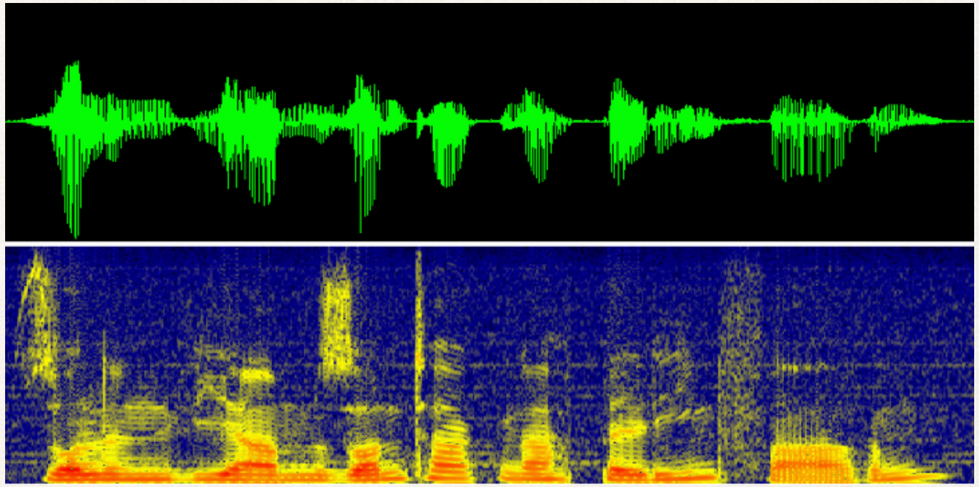




ref

Audio Data

- Hearing aids
- Voice Recognition
- Automatic Translation



Numerical / Sensor Data

- Cern
- SKA, Telescopes
- Fitness Trackers
- Weather Forecast
- Robotics
- Kinect

Games & Simulations

- Immediate Feedback
- Chess, Go
- Physical World

Internet Data

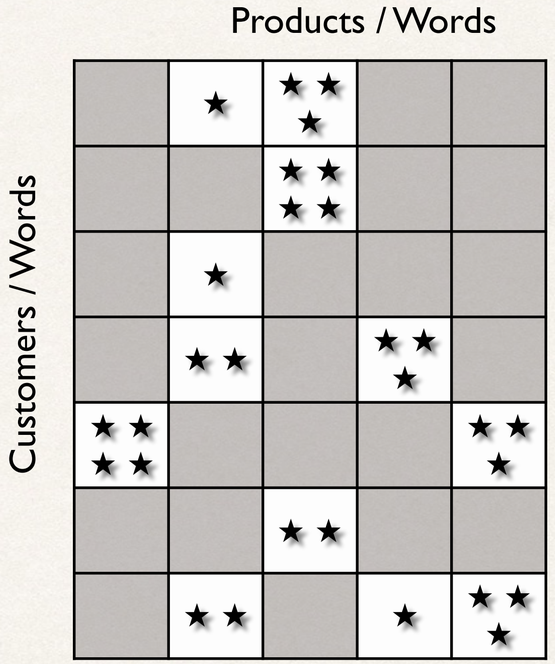
WIKIPEDIA

NETFLIX

facebook

amazon.com

Google



$\approx UV^T$

Prediction Markets

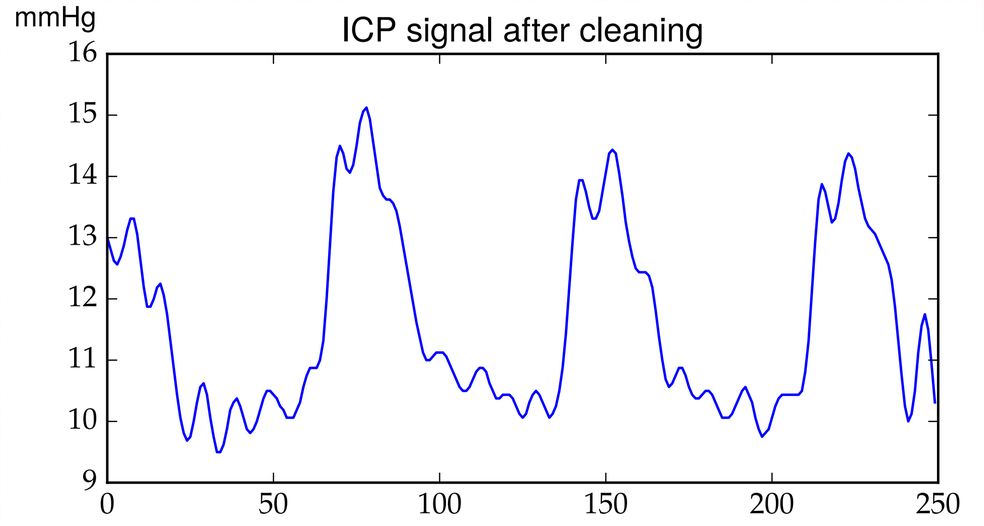
ML Applications by Master Students

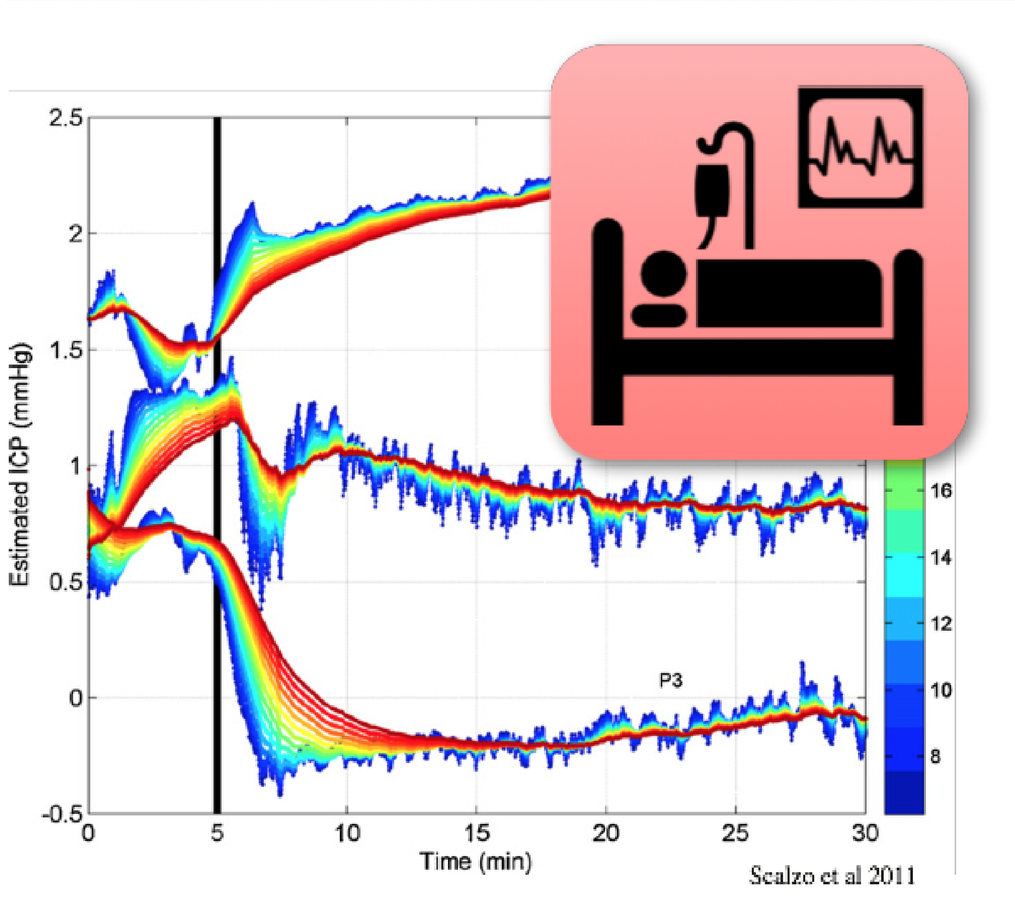
Generating Maps from Satellite Images

MSc Thesis Project
Pascal Kaiser

Medical Applications in Intensive Care

Forecasting of intracranial hypertension



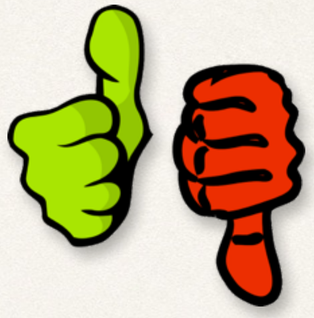


43 GB of raw CSV text data
67 days of ICP signal
sampling rate 125 Hz

MSc Thesis Projects
Matthias Hüser & Adrian Kündig

Text Understanding

sentiment of tweets



SemEval 2016:

winner out of 34

teams, >20 countries

negative	*neutral*	But i wanna wear my Concords tomorrow though but i don't feel like it
positive	*neutral*	Gonna watch Grey's Anatomy all day today and tomorrow(:
negative	*neutral*	@CoachVac heey do you know anything about UVA's fallll fest loll they invited me
neutral	*neutral*	@DustyEf when that sun is high in that Texas sky, I'll be buckin it to county fair. A
neutral	*positive*	Up 20 points in my money league with Vernon Davis and L. Fitz still to go tomorro
neutral	*positive*	DEEJAYING this FRIDAY in THE FIRST CHOP it's CHRIS actual SMITH with a smash
negative	*negative*	The Rick Santorum signing that was scheduled for tomorrow at the Books A Million
positive	*neutral*	@dreami9 lol yep looks like it! Was after El Clasico on Sunday. I didn't like her lol
neutral	*neutral*	Back in Stoke on Trent for the 2nd time today!
neutral	*neutral*	First Girls Varsity Basketball Game tomorrow at 6:00 pm Then Football Senior nigh
neutral	*neutral*	#UFC lightweights @Young__Assassin VS @jamievarner set for TUF 16 Finale on t
neutral	*neutral*	@OOOOO_WEEEE slide thru sometime this weekend ill have somethin yu can sip o
negative	*negative*	@DannyB618 Sure absolutely-- I meant out of the Bachmann, Perry, Santorum, H
negative	*negative*	@RichardGordon48 re Levein discussion on Wed. Can't keep changing boss, but he
neutral	*neutral*	Today In History November 02, 1958 Elvis gave a party at his hotel before going o
neutral	*positive*	Hustle cause you got to then kick back n party everyday like its Fri
positive	*positive*	I can't sleep. Way too exited about Vancouver tomorrow! I'm like a kid at Christma

MSc Thesis Projects
Jan Deriu & Maurice Gonzenbach

Audio: Bird Songs & Ecology

another kind of tweets

winner of

BirdCLEF 2016

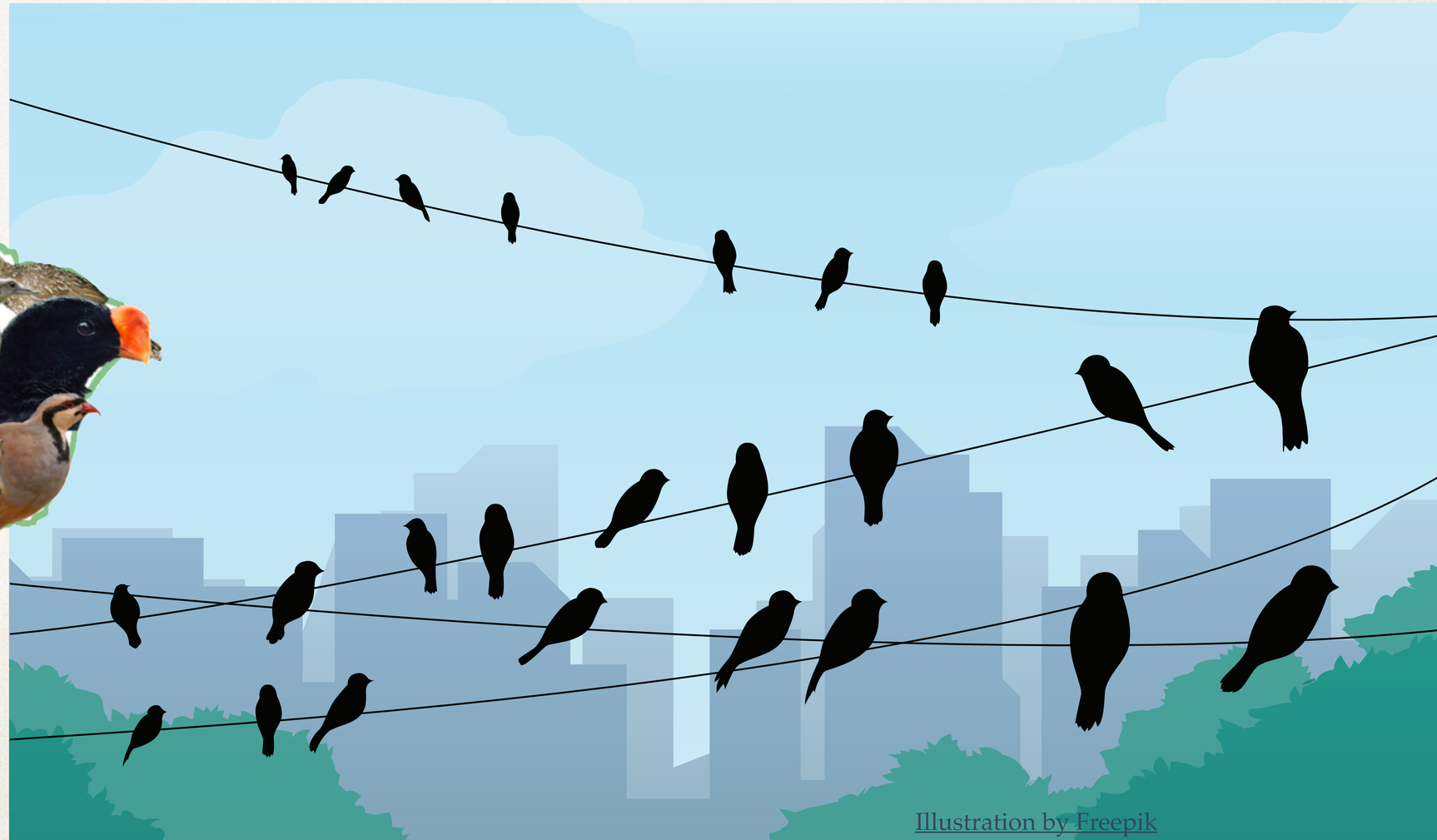


Illustration by Freepik

MSc Thesis Project
Elias Sprengel

up next:

- Regression
- Linear Regression
- Classification
- … fundamental concepts of ML