Machine Learning

An Introduction

Martin Jaggi
martin.jaggi@epfl.ch

EPFL Machine Learning and Optimization Laboratory
mlo.epfl.ch

28th Feb 2018

What to expect?

- **overview** over ML
- basic **understanding** of most important ML **methods** and **concepts**

Outline

- **Intro**
- **Part I** - Machine Learning Fundamentals
  - Regression, Classification
  - Unsupervised Learning
  - Recommender Systems and Matrix Factorizations
  - Deep Learning
- **Part II** - Big Data Infrastructure
  - Map Reduce

Introduction

What is Machine Learning?

What is Machine Learning?

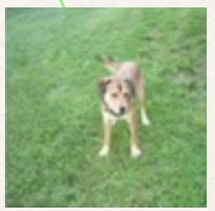
algorithms that can

learn from data

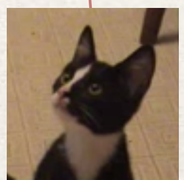
image source

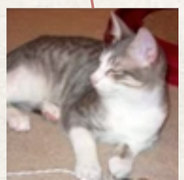
Classification

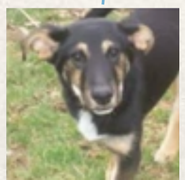
$\boldsymbol{x}_i \in \mathbb{R}^d$

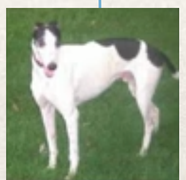


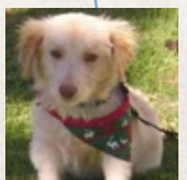
Training data





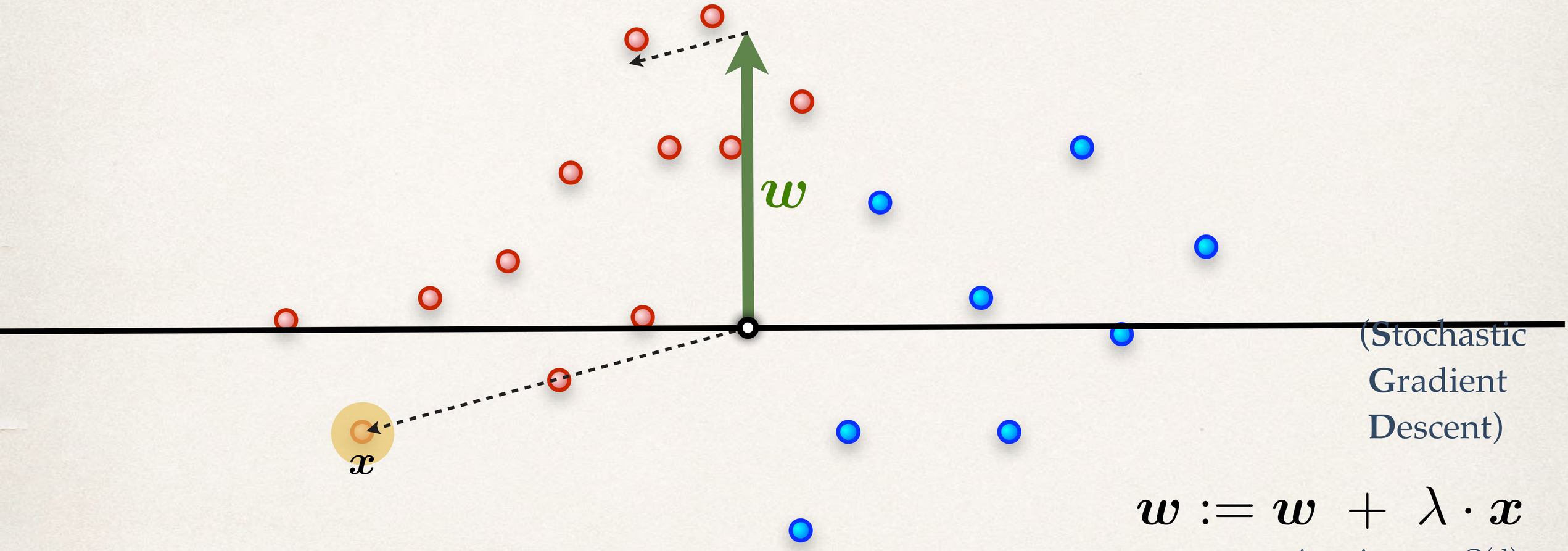






The Learning Algorithm

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



(Stochastic Gradient Descent)

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

iteration cost: O(d)

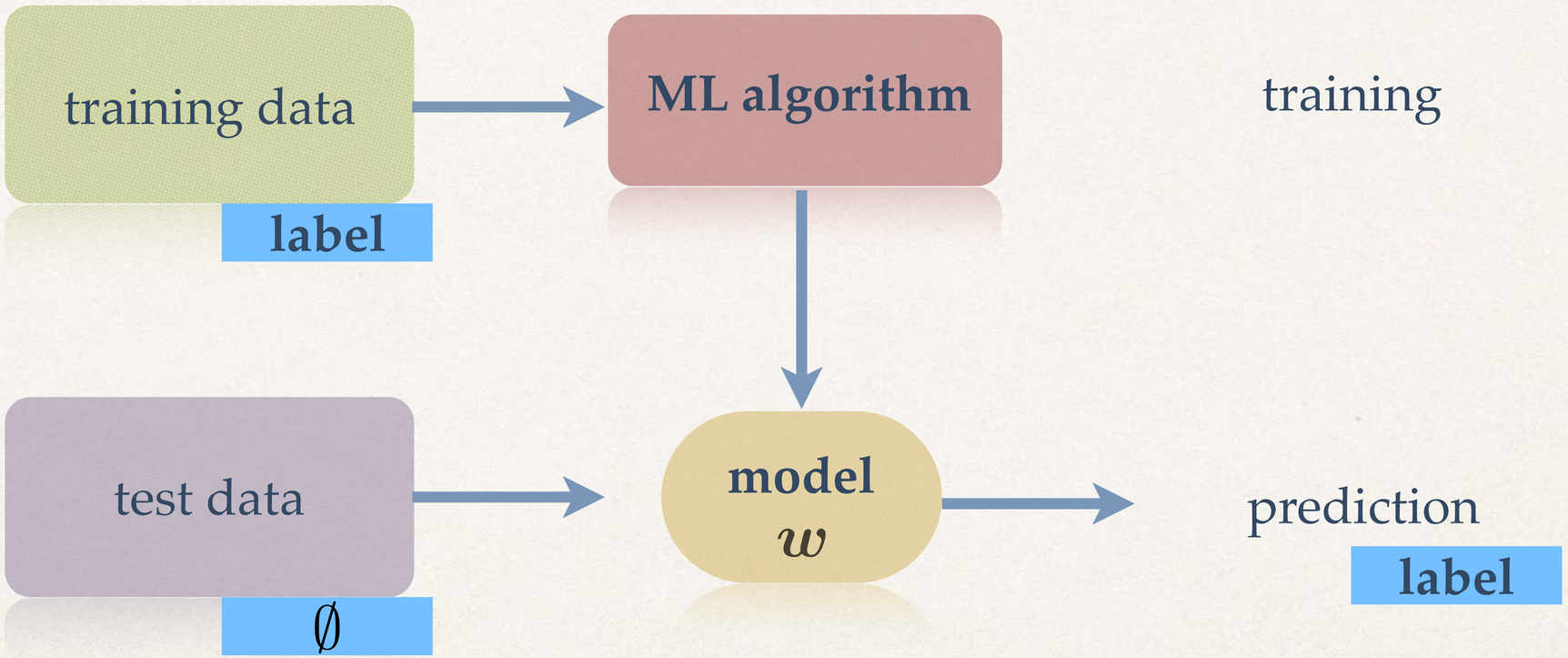
Perceptron

(Rosenblatt 1957)

Support-Vector-Machine

(Cortes & Vapnik 1995)

Machine Learning Fundamentals



what Evaluation Metric?

towards…

understanding intelligence

?

if-then-else

≠

intelligence

towards…

understanding intelligence

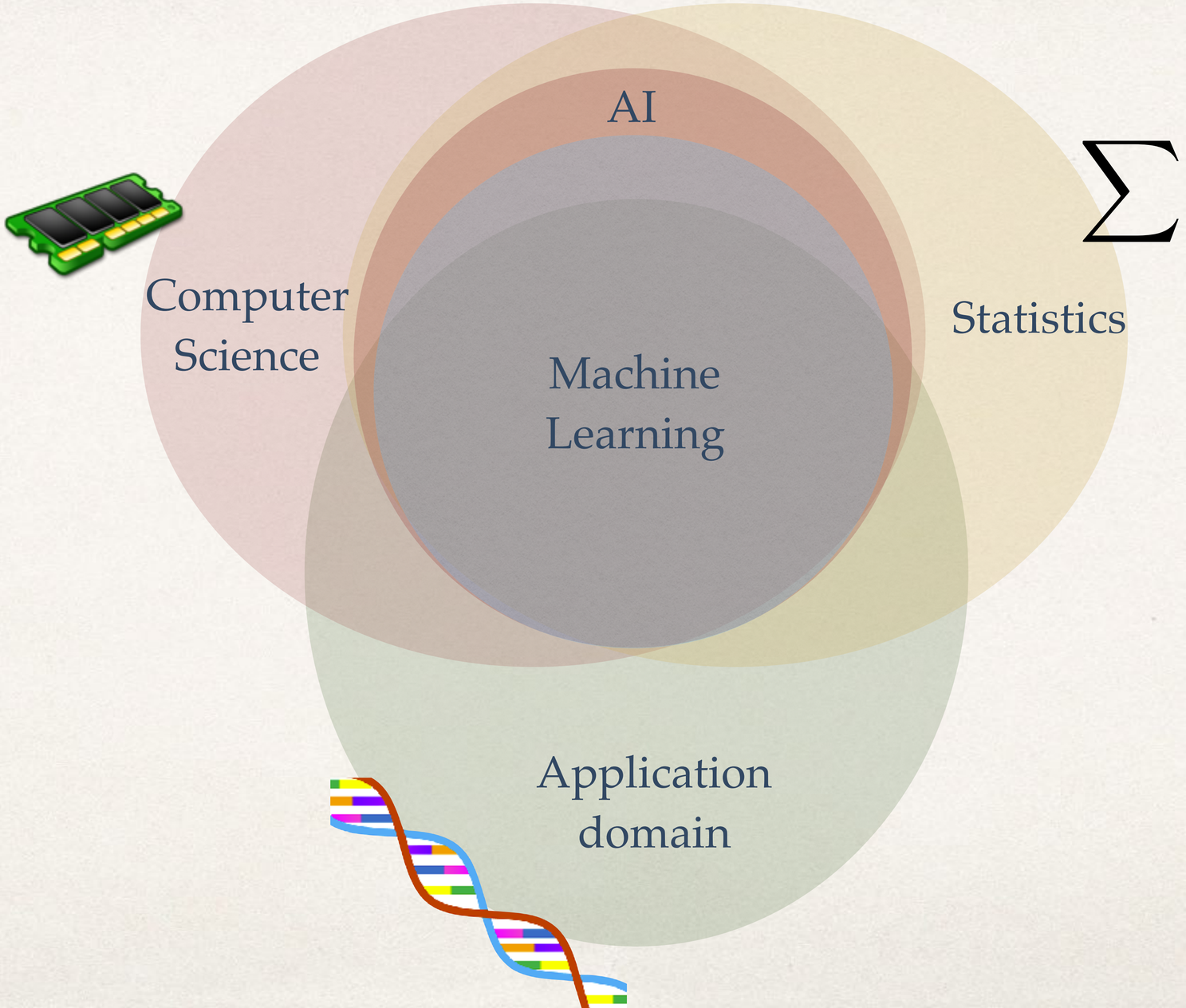
?

Machine Learning

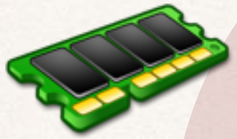
vs

Neuroscience / HBP

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


AI
Computer Science
Statistics
Machine Learning
Application domain

Job Skills



Coding Skills

Machine Learning

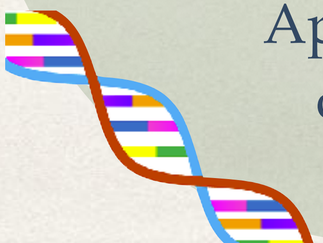
$\sum$

Math & Statistics

Data Scientist

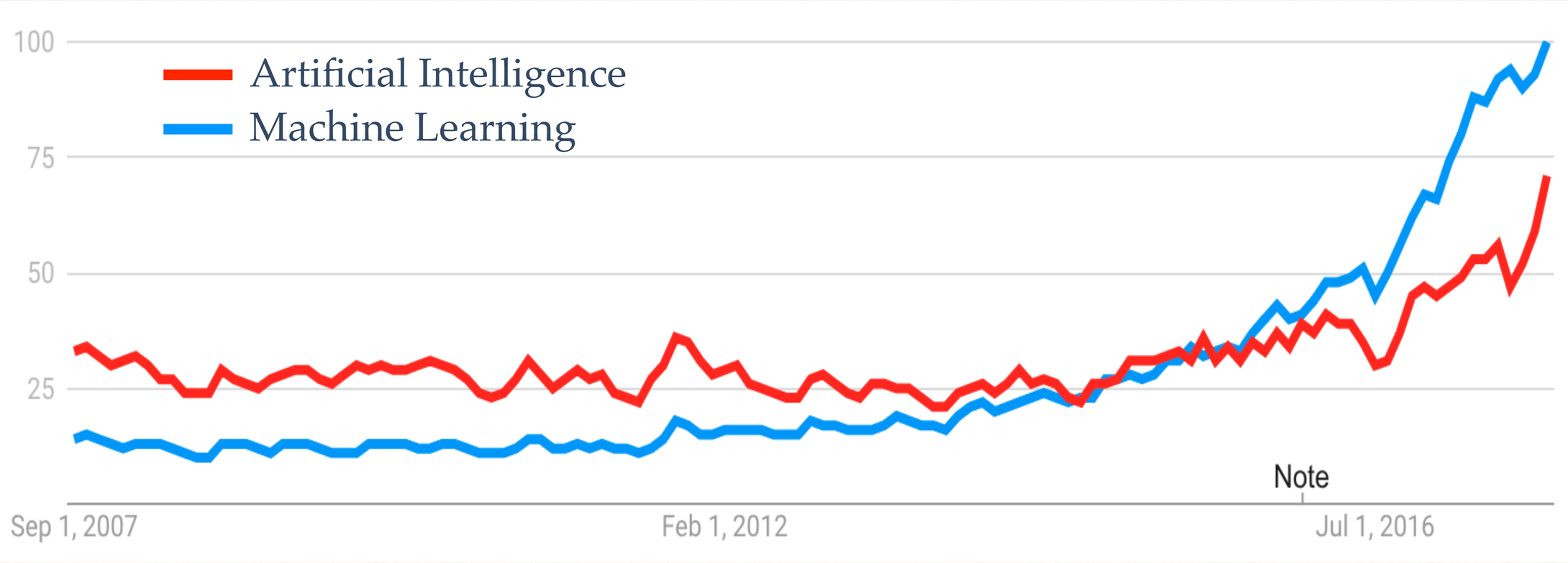
Software Engineer

Traditional Scientist



Application domain

Evolution of Interest



source

why ML?

Applications

Industry Applications

- **not only** the ‘usual suspects’

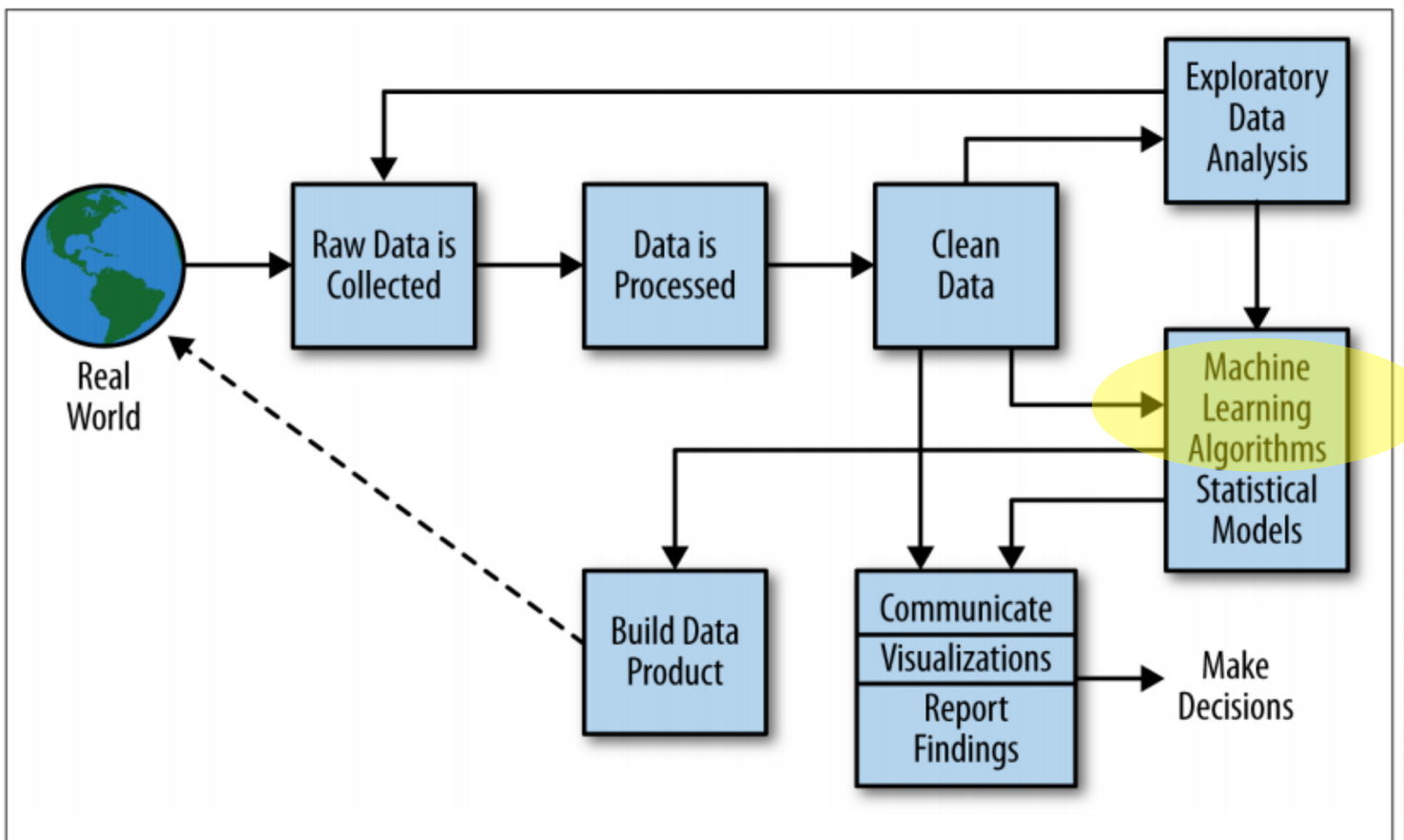
- majority of industries, originally not ‘digital’:
  - agriculture, car & bike sharing, booking.com, manufacturing, sports, personalized health, call centers, …

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

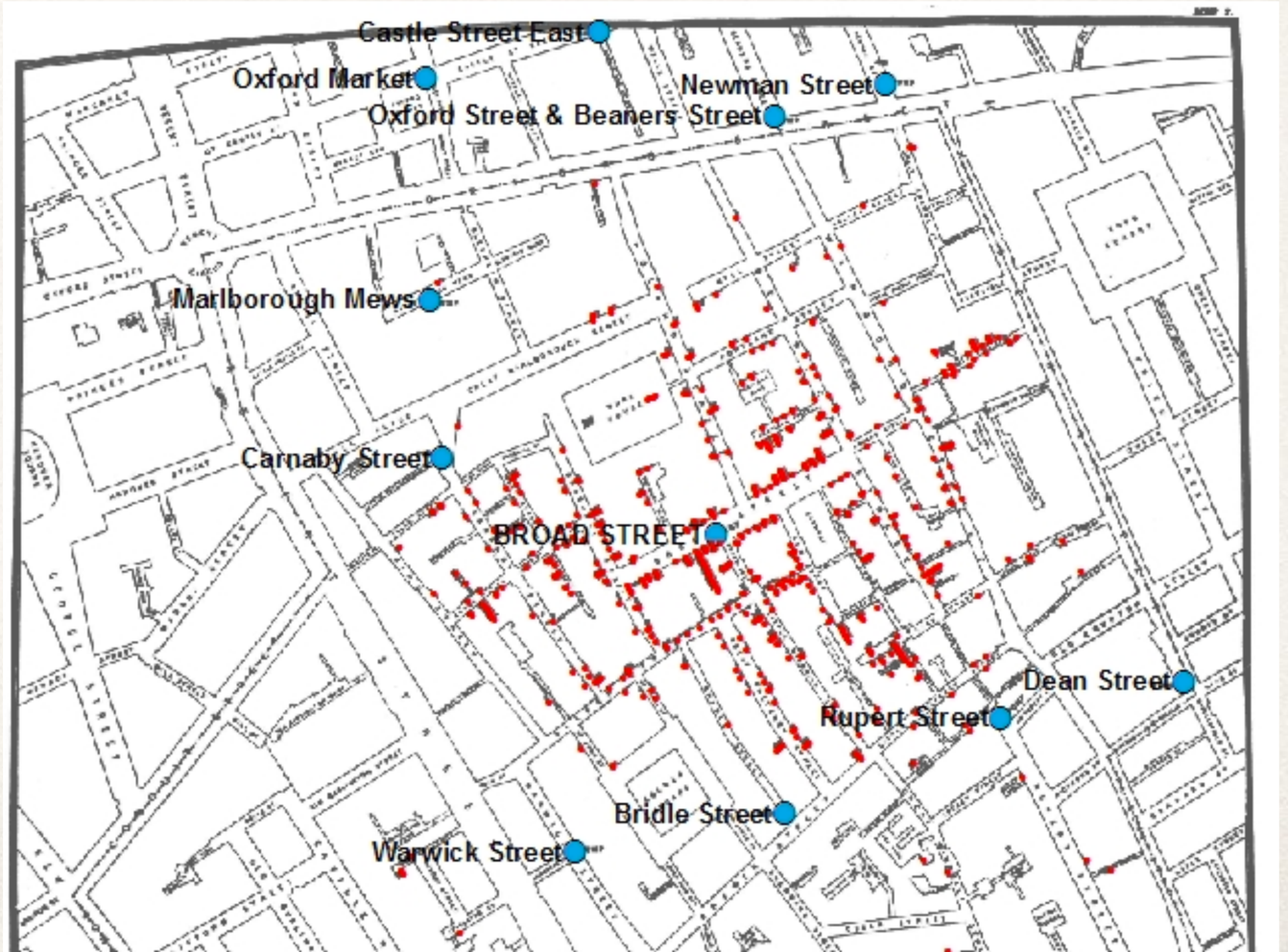
ML is only a small part!



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

2017: 10^{17} 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

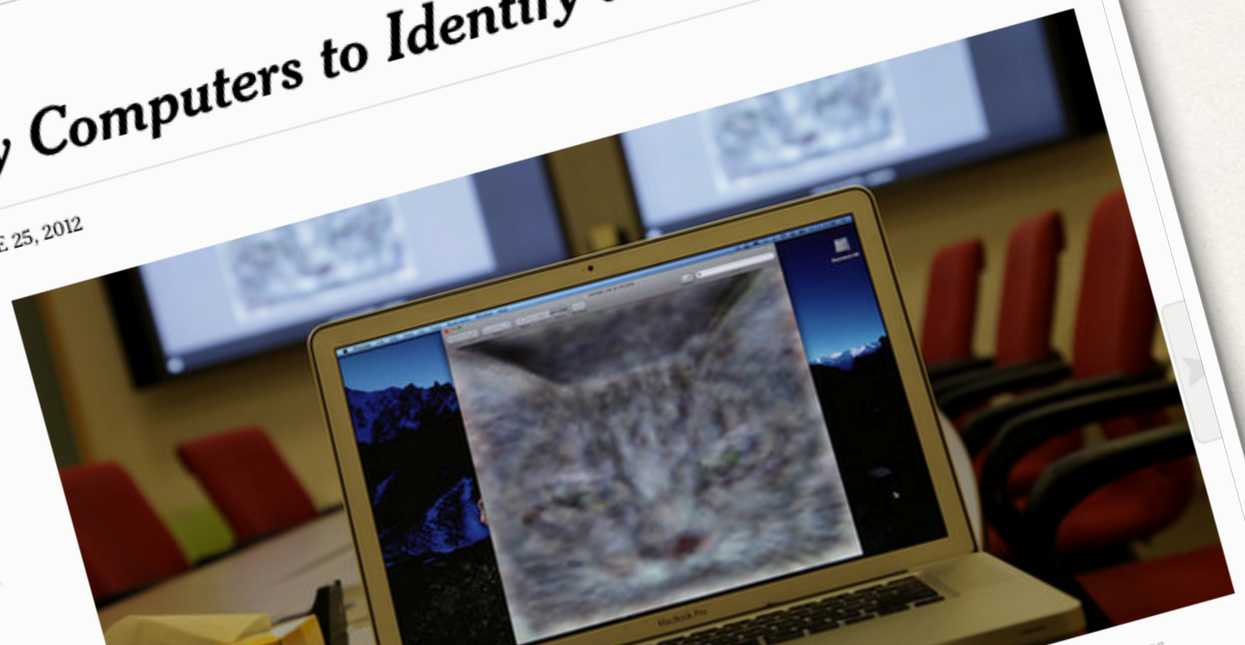
The New York Times

SECTIONS

TECHNOLOGY

How Many Computers to Identify a Cat? 16,000

By JOHN MARKOFF JUNE 25, 2012



Email

Share

Tweet

Challenges

- Hype
  - cycles of AI popularity
- Data Ethics, Privacy, Fairness

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

★★★★½ (12)

EUR 5,22

Sturmhaube 3-loch Acryl schwarz von Unbekannt

★★★½☆ (13)

EUR 1,38

Pfefferspray KO-FOG 40ML von Pfeffer KO

★★★★½ (17)

EUR 5,19

KO-CS Verteidigungsspray 40 ml von KO-CS

★★★★½ (33)

EUR 4,22

Challenges

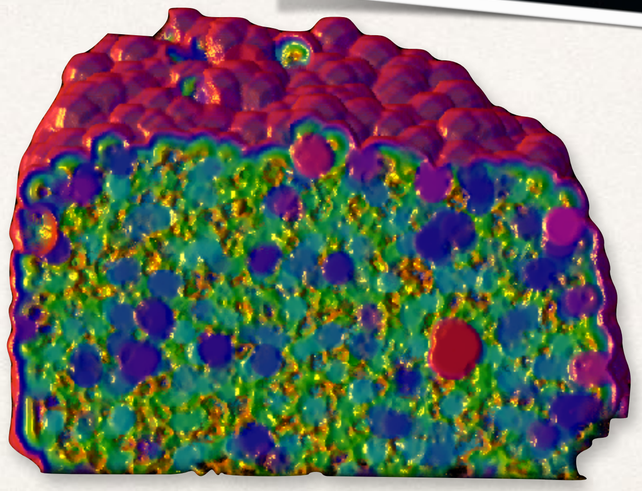
- Hype
  - cycles of AI popularity
- Data Ethics, Privacy, Fairness
- Lack of Interpretability
  - example: medical applications of deep learning
- Social Implications,
  Threats from Super-human AI?
  - see Nick Bostrom

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

Text Data

- Spam Detection
- User Content
- Medical Data
- Machine Translation

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

signing that was scheduled for tomorrow at the Books A Million

oks like it! Was after El Clasico on Sunday. I didn't like her lol

ent for the 2nd time today!

asketball Game tomorrow at 6:00 pm Then Football Senior nigh

@Young__Assassin VS @jamievarner set for TUF 16 Finale on t

lide thru sometime this weekend ill have somethin yu can sip c

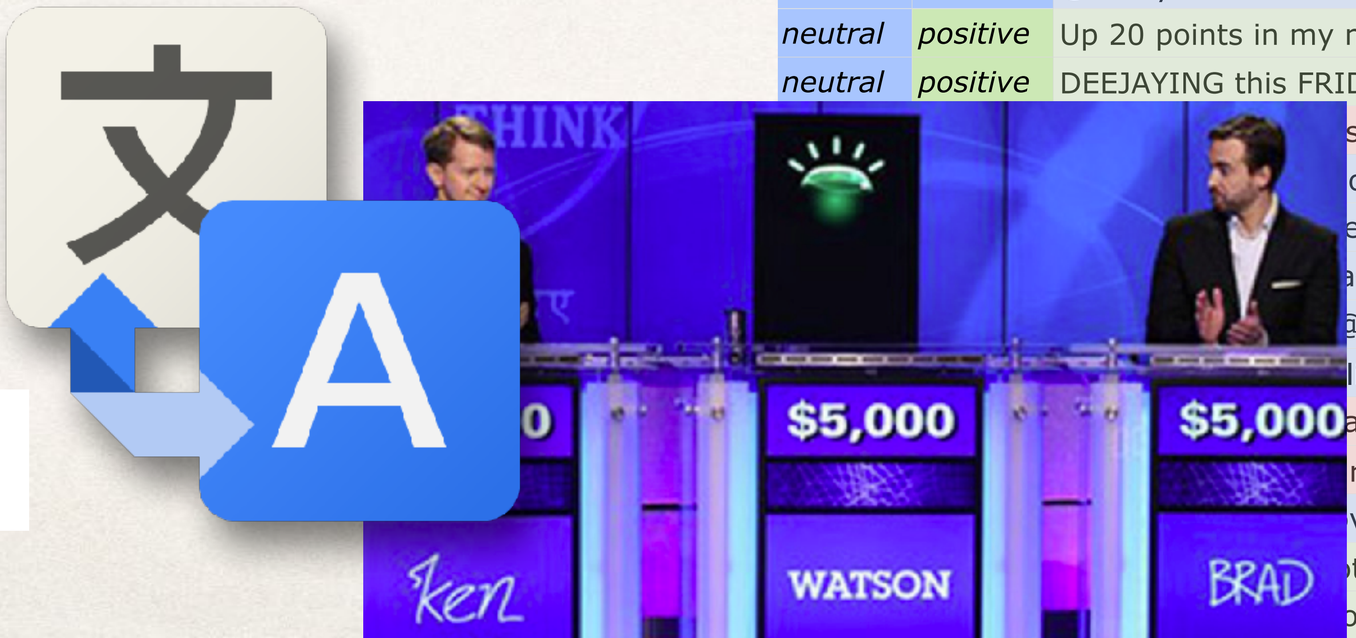
absolutely-- I meant out of the Bachmann, Perry, Santorum, H

re Levein discussion on Wed. Can't keep changing boss, but he

vember 02, 1958 Elvis gave a party at his hotel before going o

t to then kick back n party everyday like its Fri

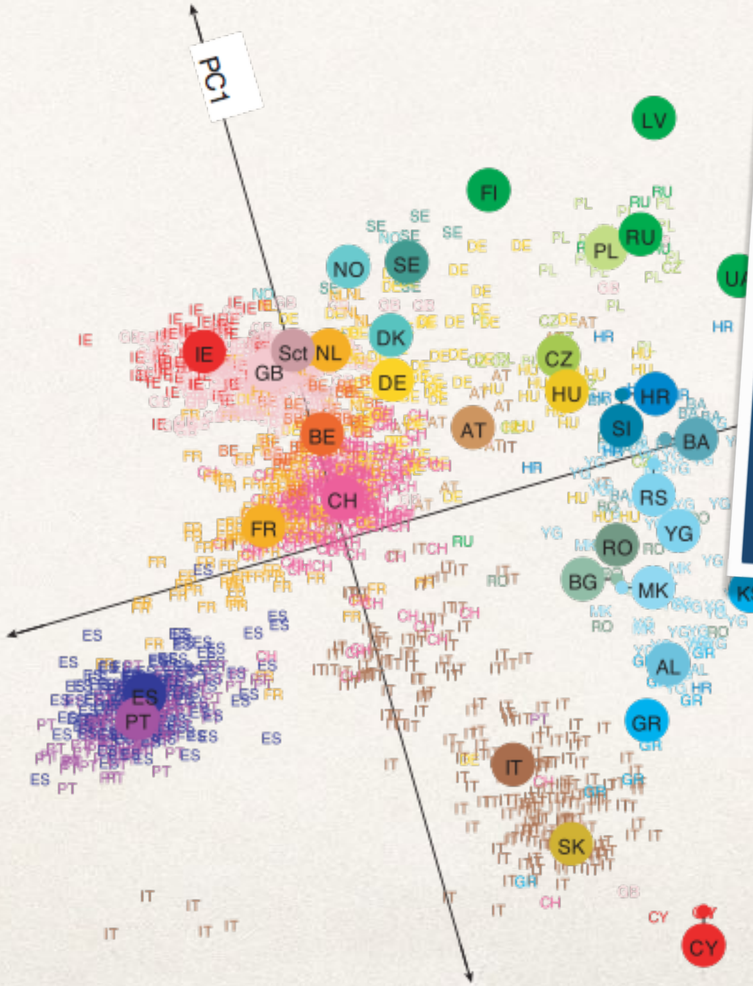
o exited about Vancouver tomorrow! I'm like a kid at Christma

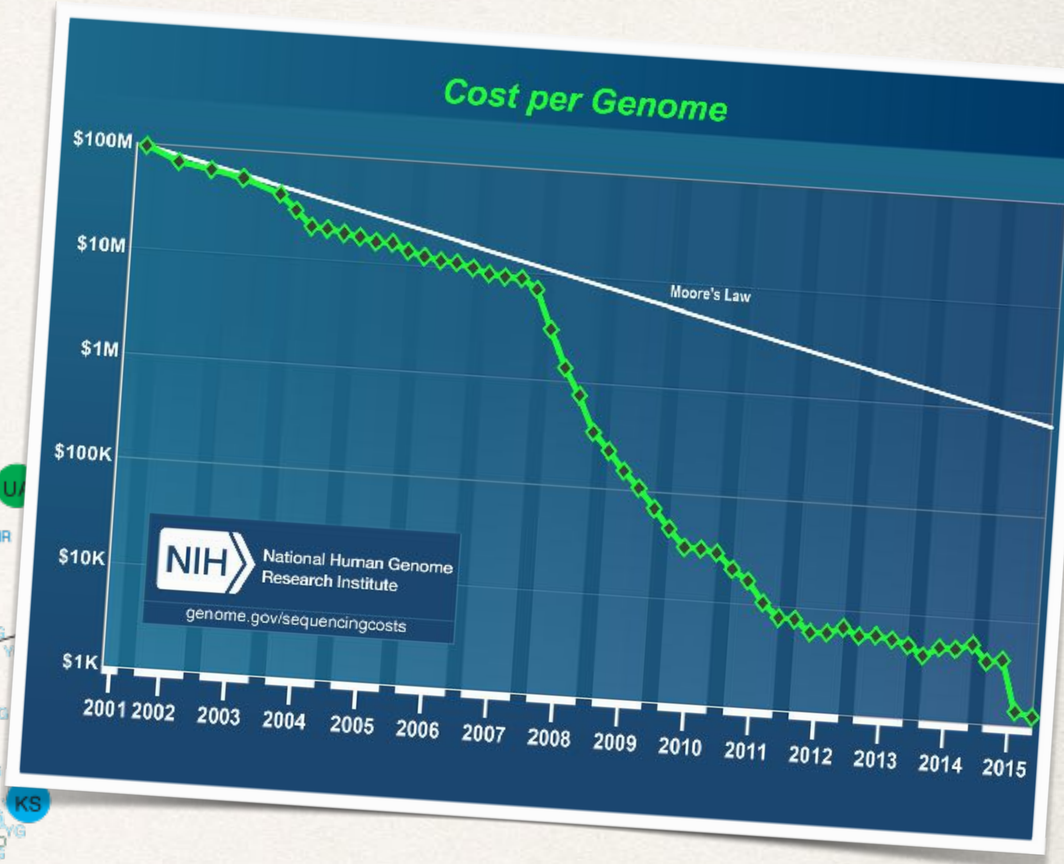


DeepL

Medical: Genetic Data

721290	C	G
723819	A	T
723891	C	G

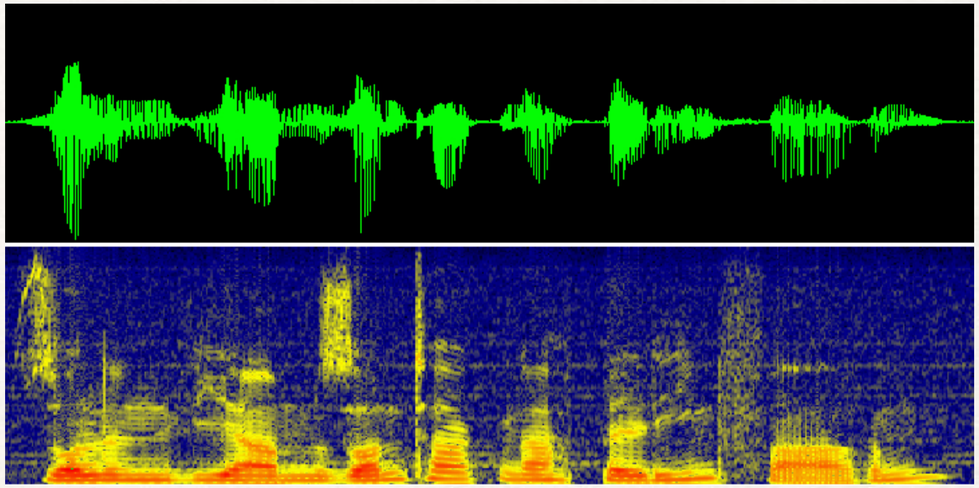




ref

Audio Data

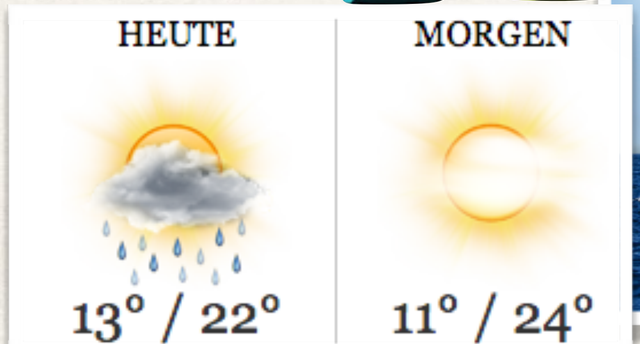
- Hearing aids
- Voice Recognition
- Automatic Translation



LYREBIRD

Numerical / Sensor Data

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



Games & Simulations

- Immediate Feedback
- Chess, Go
- Physical World

Internet Data

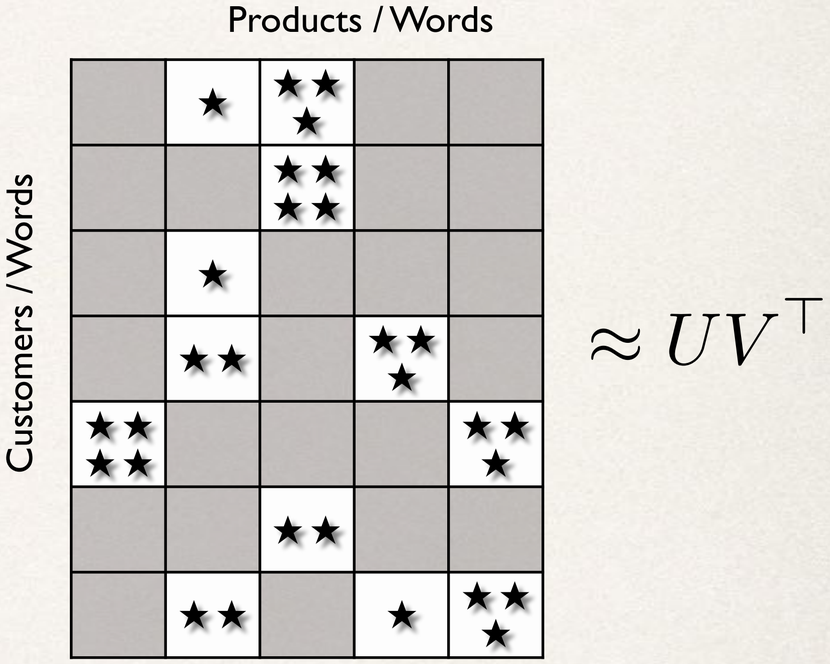
WIKIPEDIA

NETFLIX

facebook

amazon.com

Google



Prediction Markets

Insurance, Finance, Legal

- Business-Analytics
- Targeted Advertizing
- Fraud Detection
- Customer Relations
- Marketing
- Prediction of Risk / Return / other metrics
- Prison Bail link

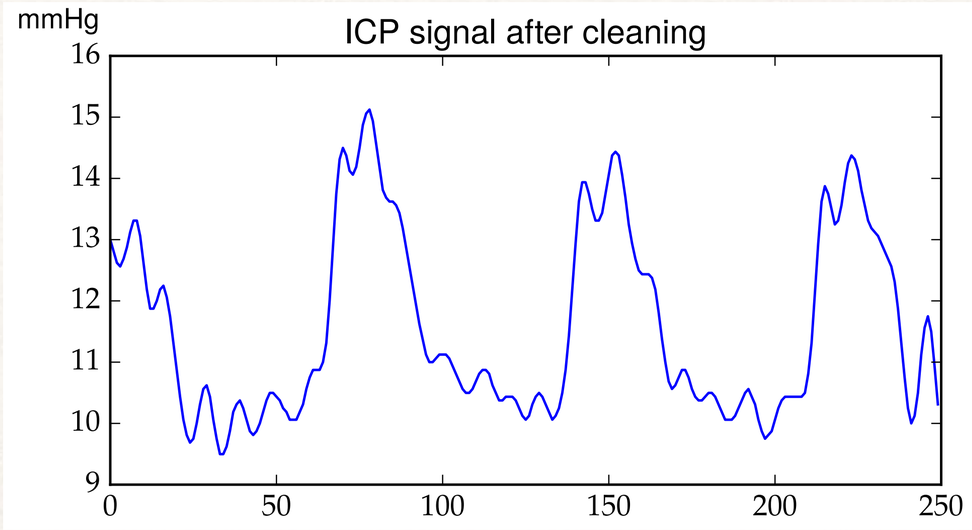
ML Applications
by Master Students

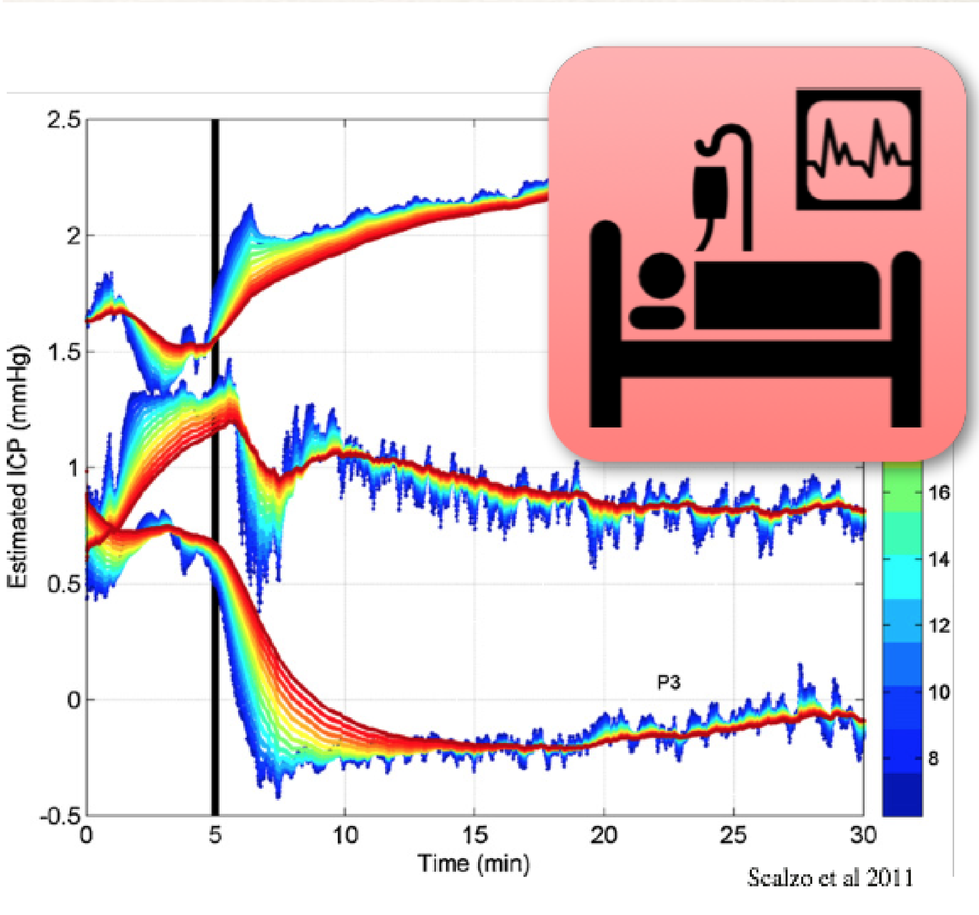
Generating Maps from Satellite Images

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 c
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

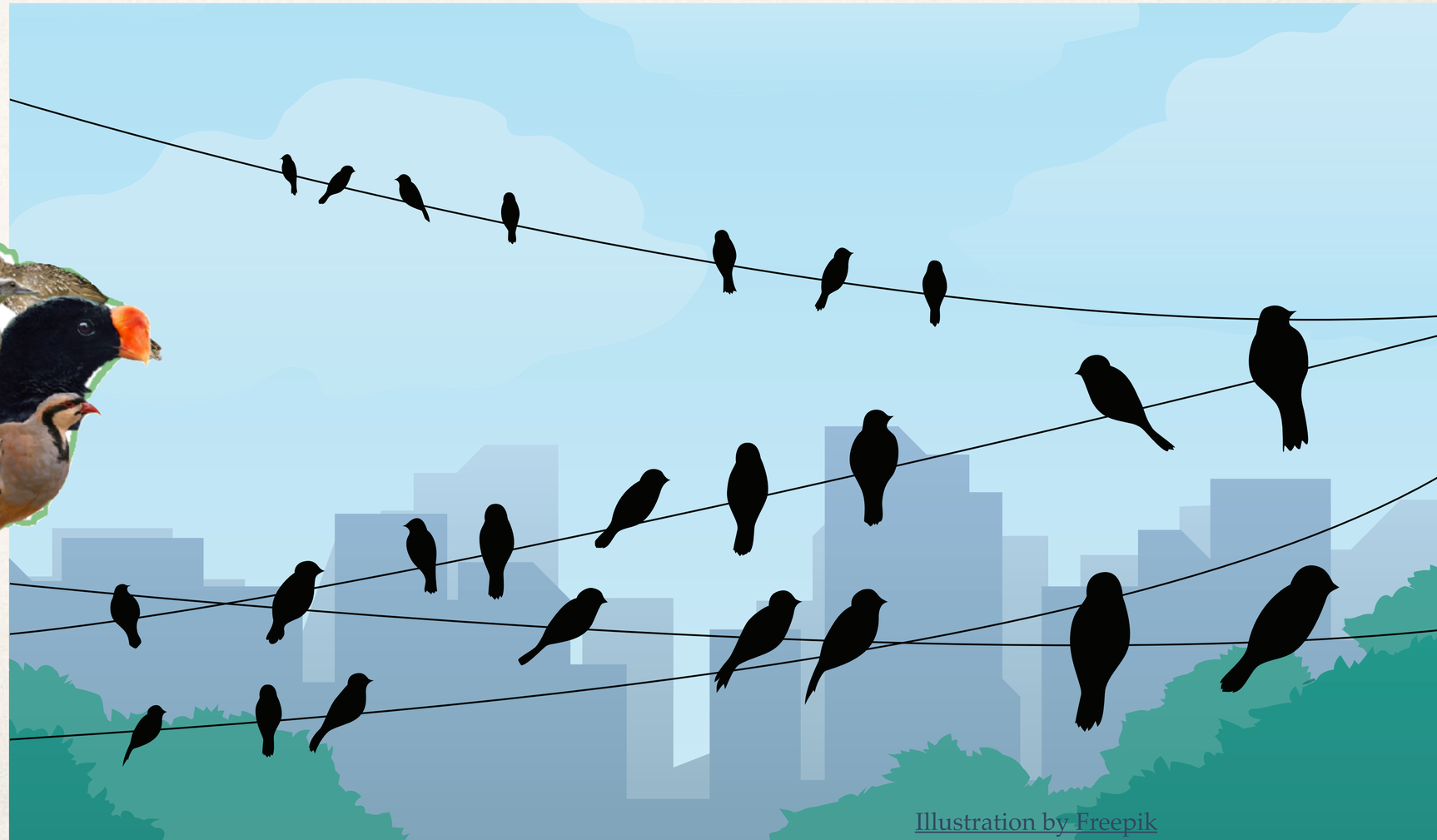
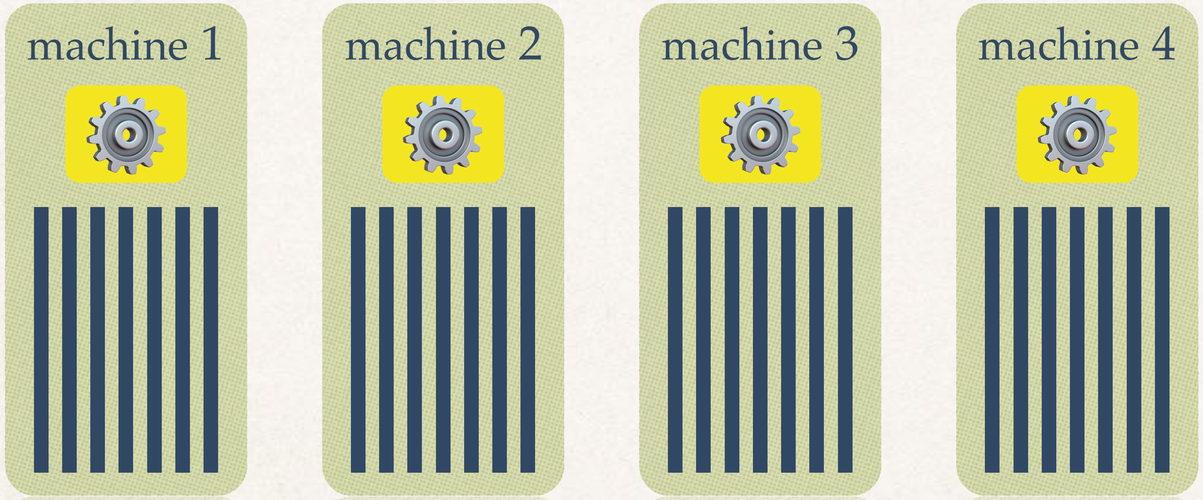


Illustration by Freepik

MSc Thesis Project
Elias Sprengel

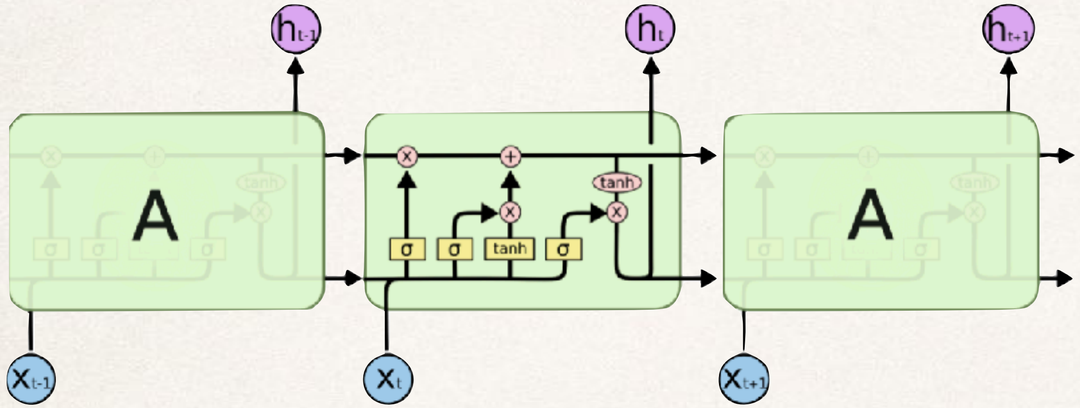
Distributed Training

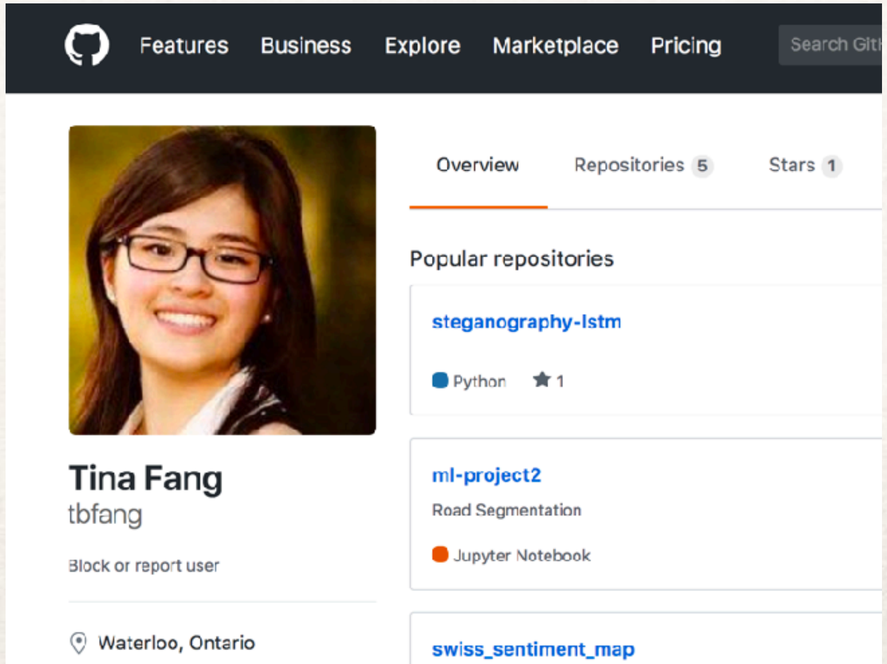


MSc Thesis Project
Simone Forte

google cloud ML

Text Steganography with Neural Networks





steganographic tweets:

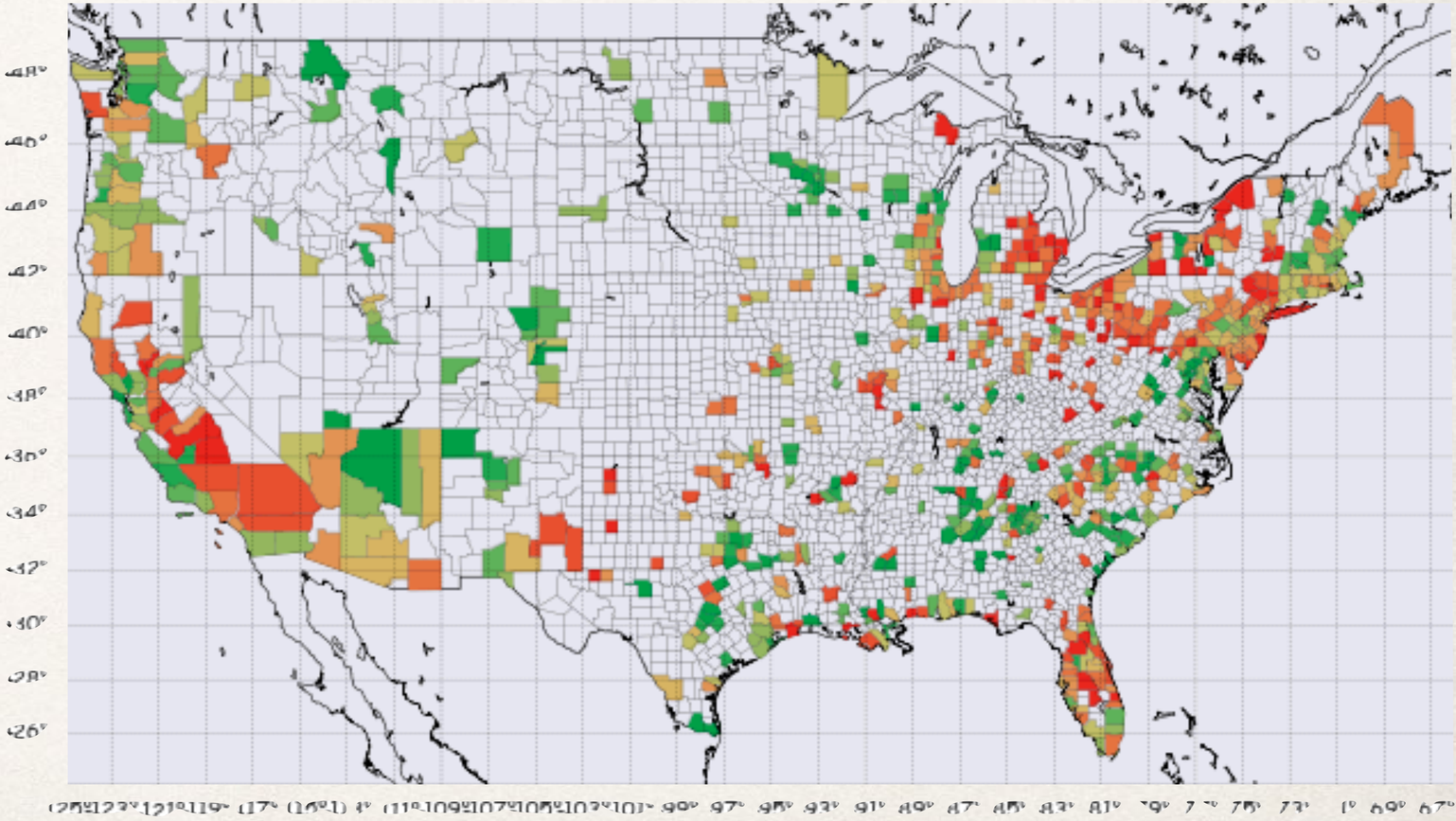
- *I can't take the whole day to the car, the people I suddenly didnt understand.*
- *"where else were u making?... i feel fine? - e? lol" * does a voice for me & take it to walmart?*
- *Don't think I can trust anymore... but I know is my fault.*
- *I can't believe he is a freshman!*

Semester Project
Tina Fang

Heart Attacks from Tweets

positively correlated

negatively correlated



<u>Semester Project</u>
Arno Schnewly

other materials

- **introductory courses online**
- **open source code examples online**
- **EPFL ML course**

  github.com/epfml/ML_course

monthly / bimonthly

Lausanne Machine Learning

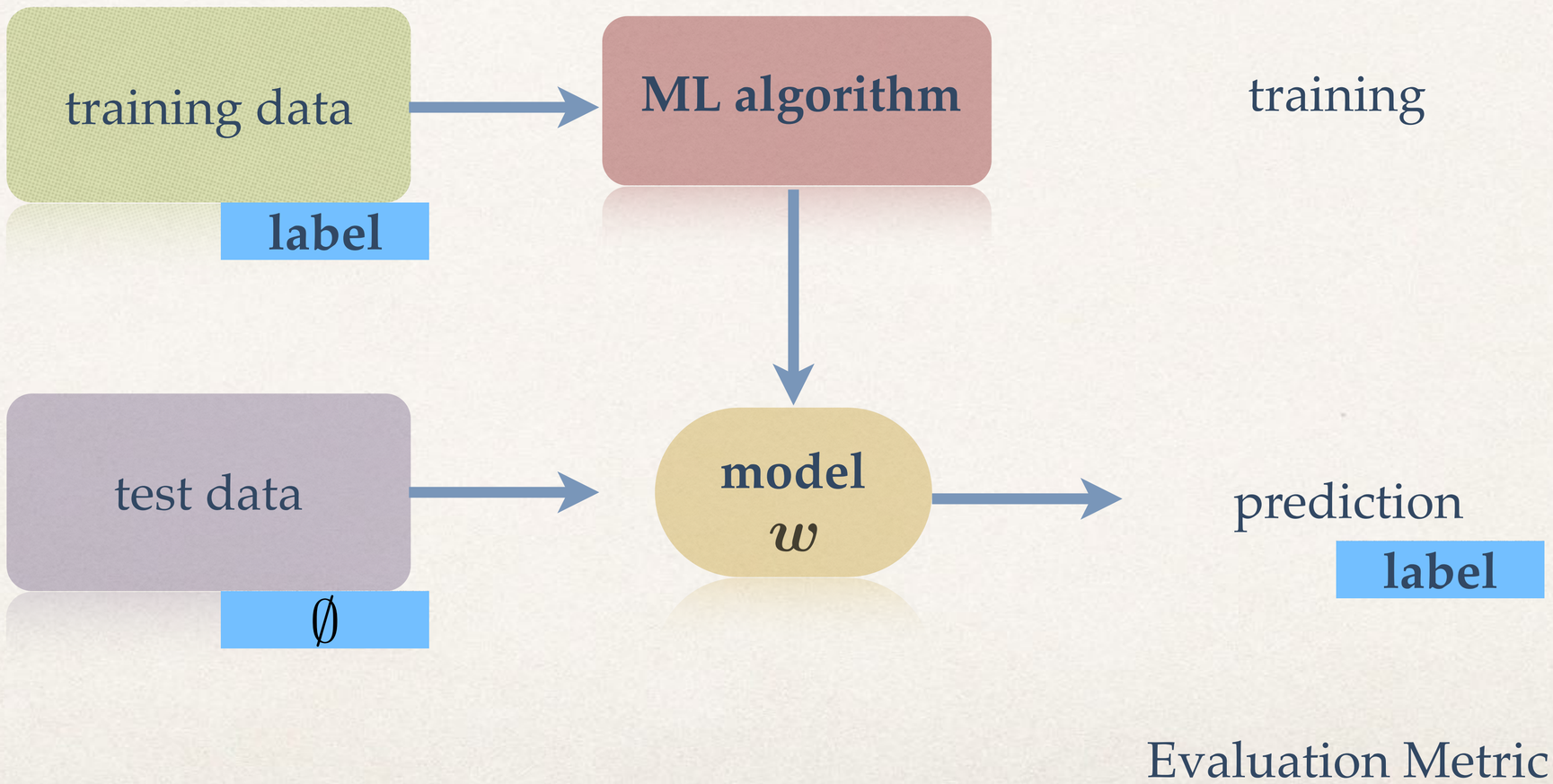
[Link to Website]

up next:

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

Part I

Machine Learning Fundamentals



The Most Common Methods

- Supervised Methods
  - *(Linear)* Classification & Regression
  - Kernel Methods, Decision Trees & Random Forests
  - Deep Learning
- Unsupervised Methods
  - Clustering, Dimensionality Reduction
  - Text, word vectors, Deep Learning

Regression & Classification Models

$$\min_{w \in \mathbb{R}^d} \left[P(\boldsymbol{w}) := \frac{\lambda}{2}\|\boldsymbol{w}\|^2 + \frac{1}{n}\sum_{i=1}^{n} \ell_i(\boldsymbol{w}^T \boldsymbol{x}_i) \right]$$

regularizer

loss function

L1,
L2,
elastic net,
…

least squares,
hinge,
logistic,
…

Regression & Classification Models

Classification

- Support Vector Machine *(SVM) (L1,L2)*
- Logistic Regression *(L1,L2)*

Structured Prediction *(L1,L2)*

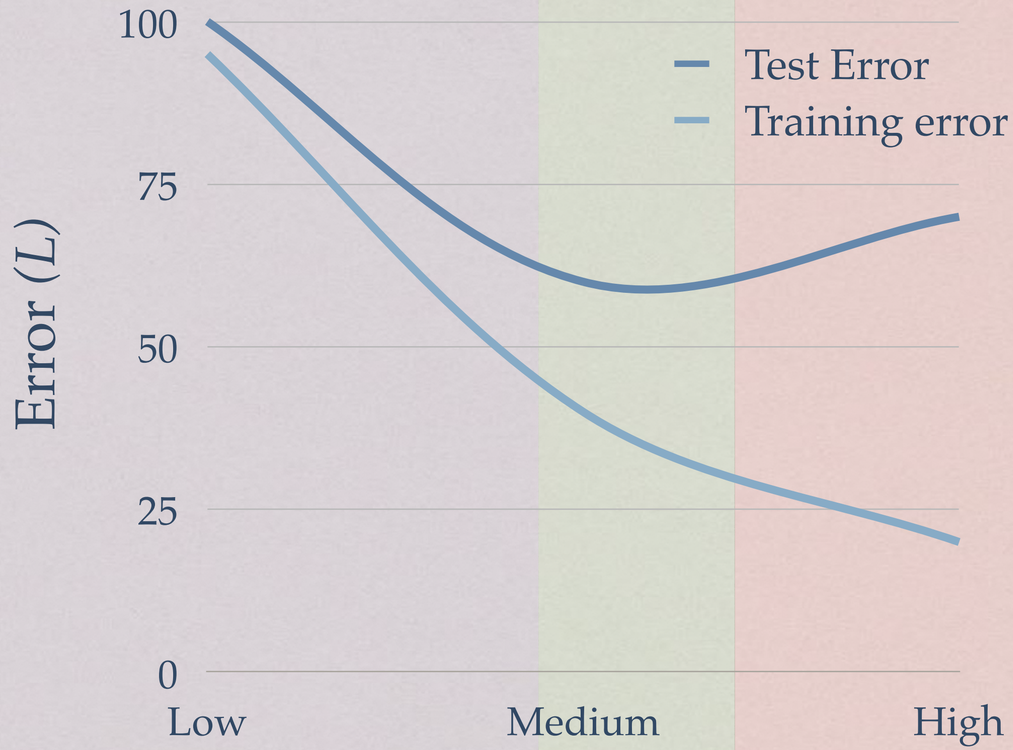
Regression

- Least Squares variants *(L1,L2):*
- Ridge Regression
- Lasso, Elastic-Net *(Feature Selection, Compressed Sensing)*
- Structured Norm

The Main Trade-Offs

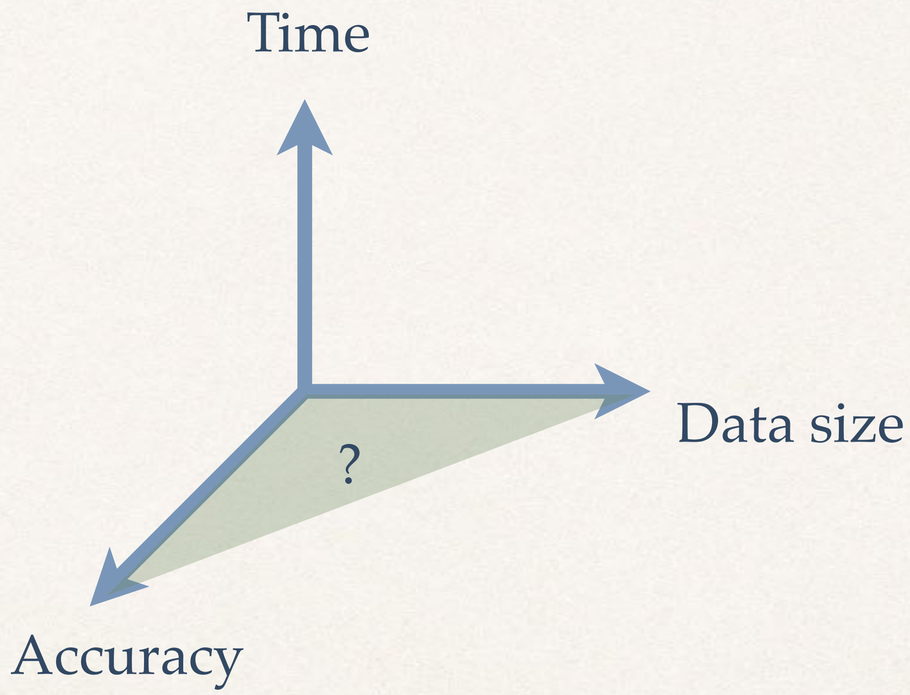
Underfitting

Overfitting



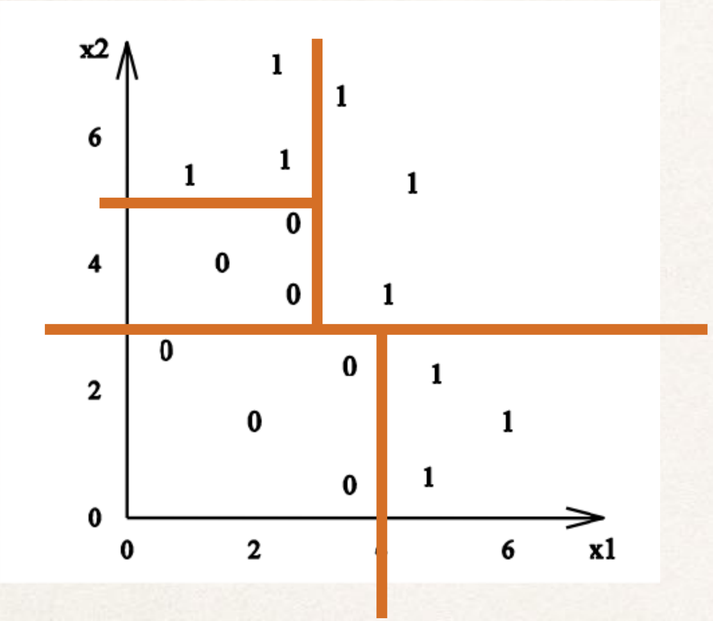
$$\min_{\boldsymbol{w} \in \mathbb{R}^d} \; \lambda R(\boldsymbol{w}) + \sum_i L(\boldsymbol{w}, \boldsymbol{x}_i)$$

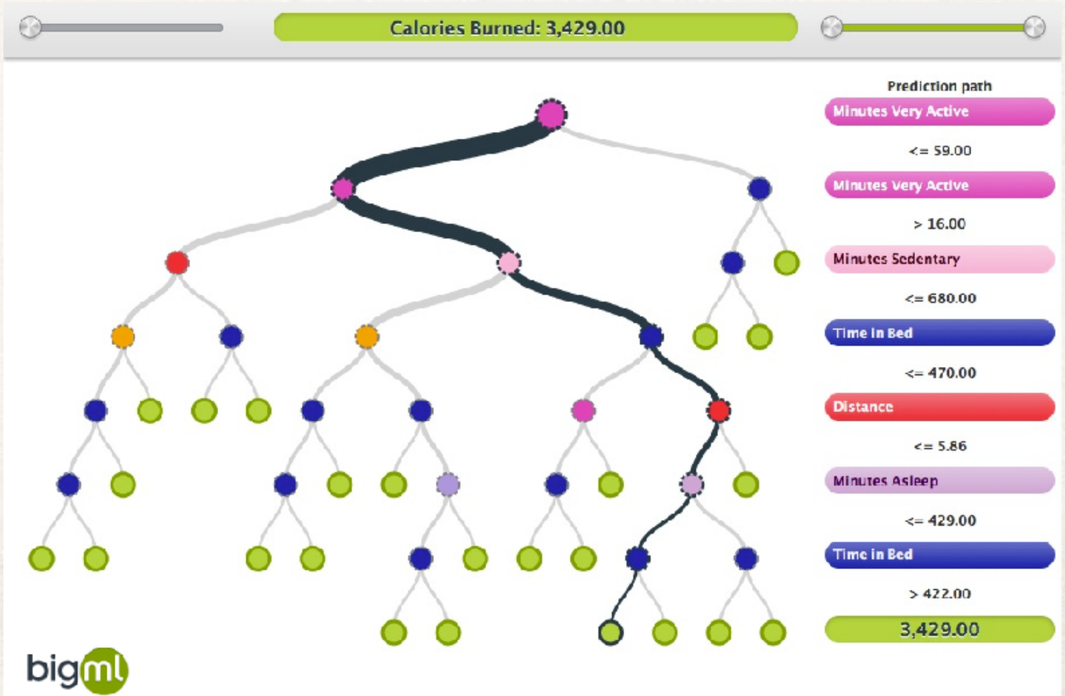
The Main Trade-Offs



Decision Trees & Random Forests

Idea: recursively split the feature space





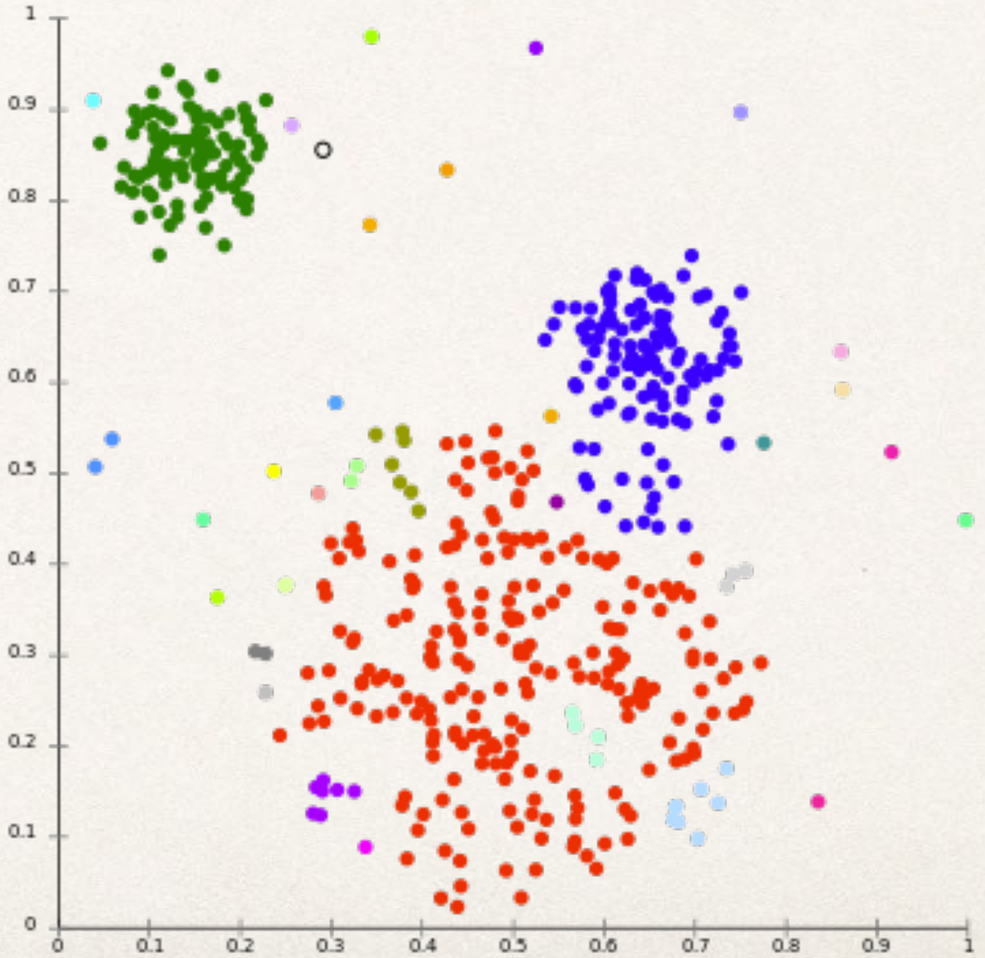
Unsupervised Methods

Clustering

k-means

Dimensionality Reduction

Semantic Embeddings



Part II

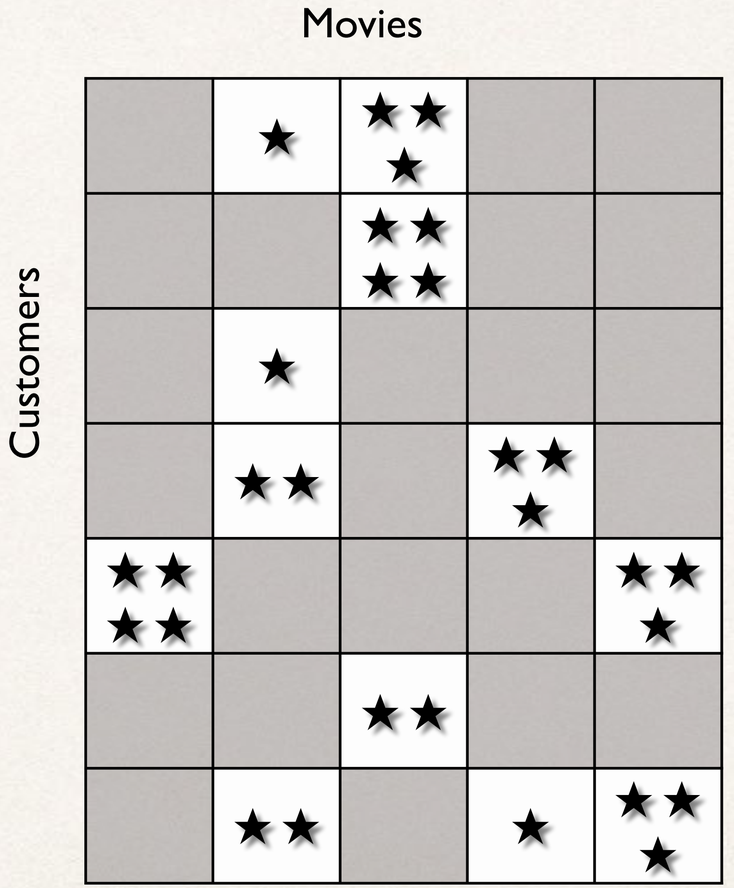
Representation Learning

Idea: Feature engineering:

How to represent a data example for use in an ML system?

Recommender Systems

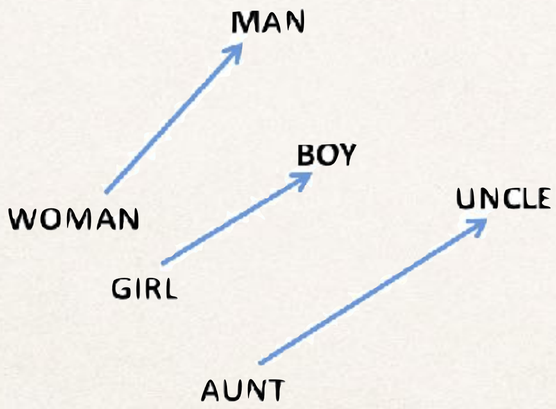
17’000 movies
500’000 customers
100’000’000 ratings



$$\approx UV^T$$

Semantic Text Representations

word2vec



Word \ Word					
		1	1		
			3		
		1			
		2		1	
	1				1
			1		
		1		1	1

$$\approx UV^T$$

spinningbytes.com/demos

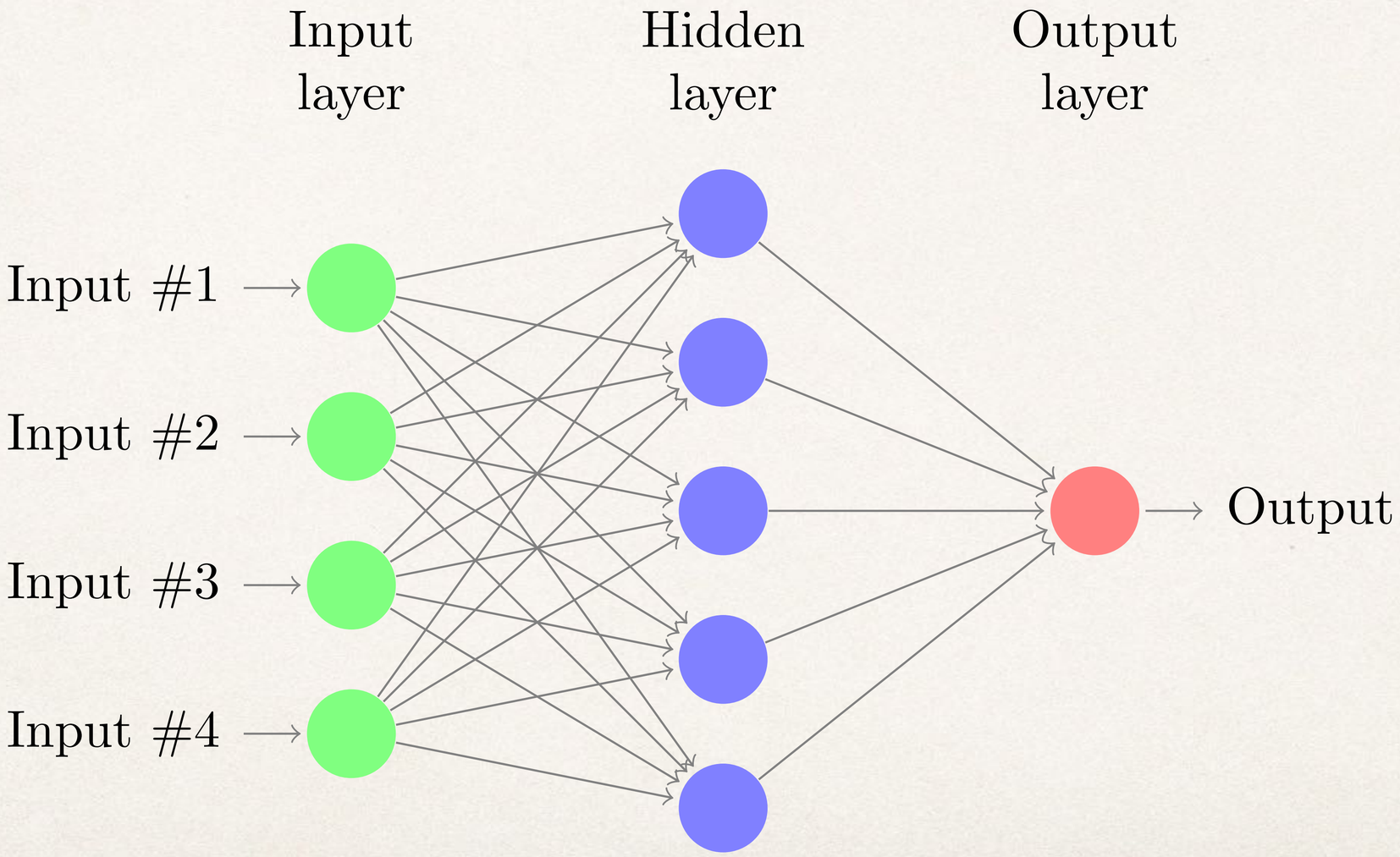
Deep Learning

The return of the neural networks

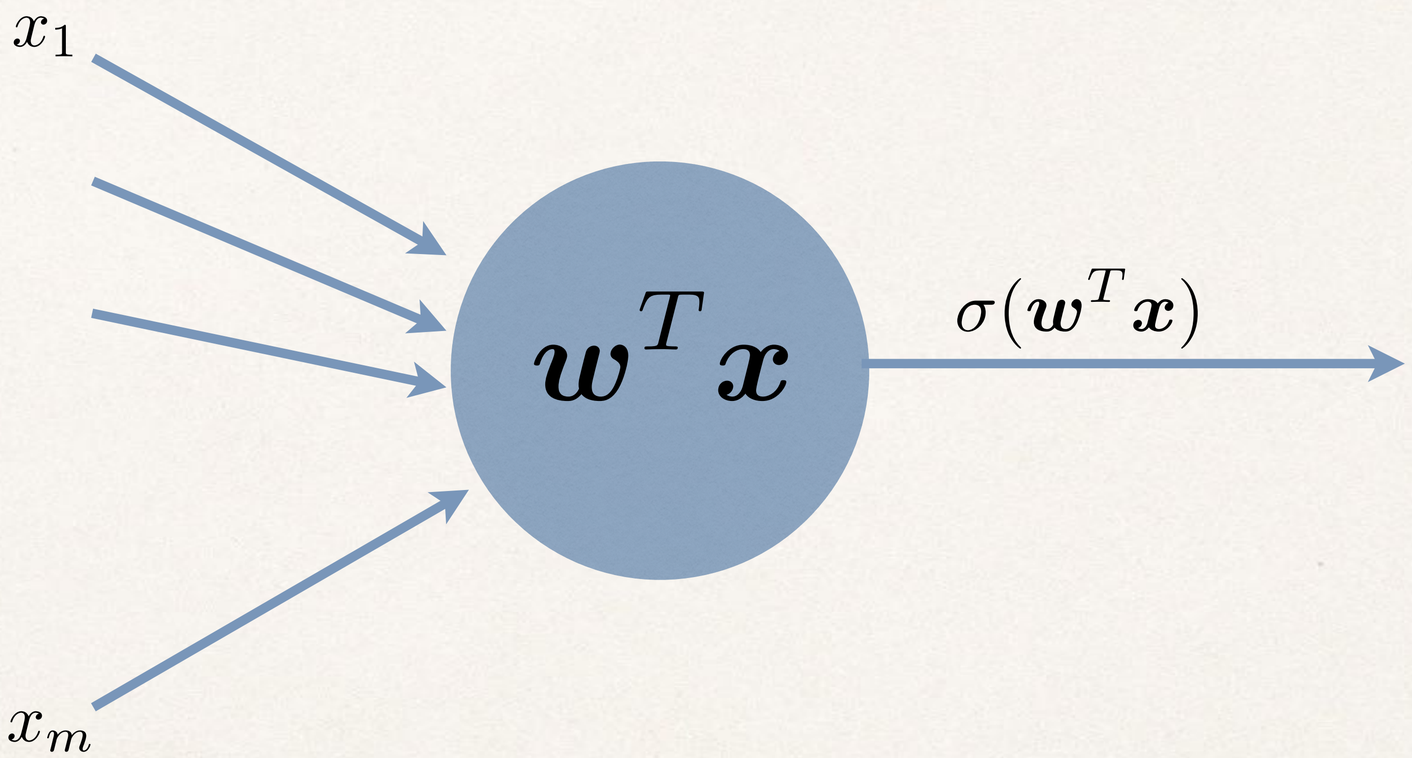
Andrew Ng Intro
https://youtu.be/n1ViNeWhC24

Bengio, Y., Courville, A., & Vincent, P. (2013).
Representation Learning: A Review and New Perspectives.
IEEE Transactions on PAMI. http://doi.org/10.1109/TPAMI.2013.50

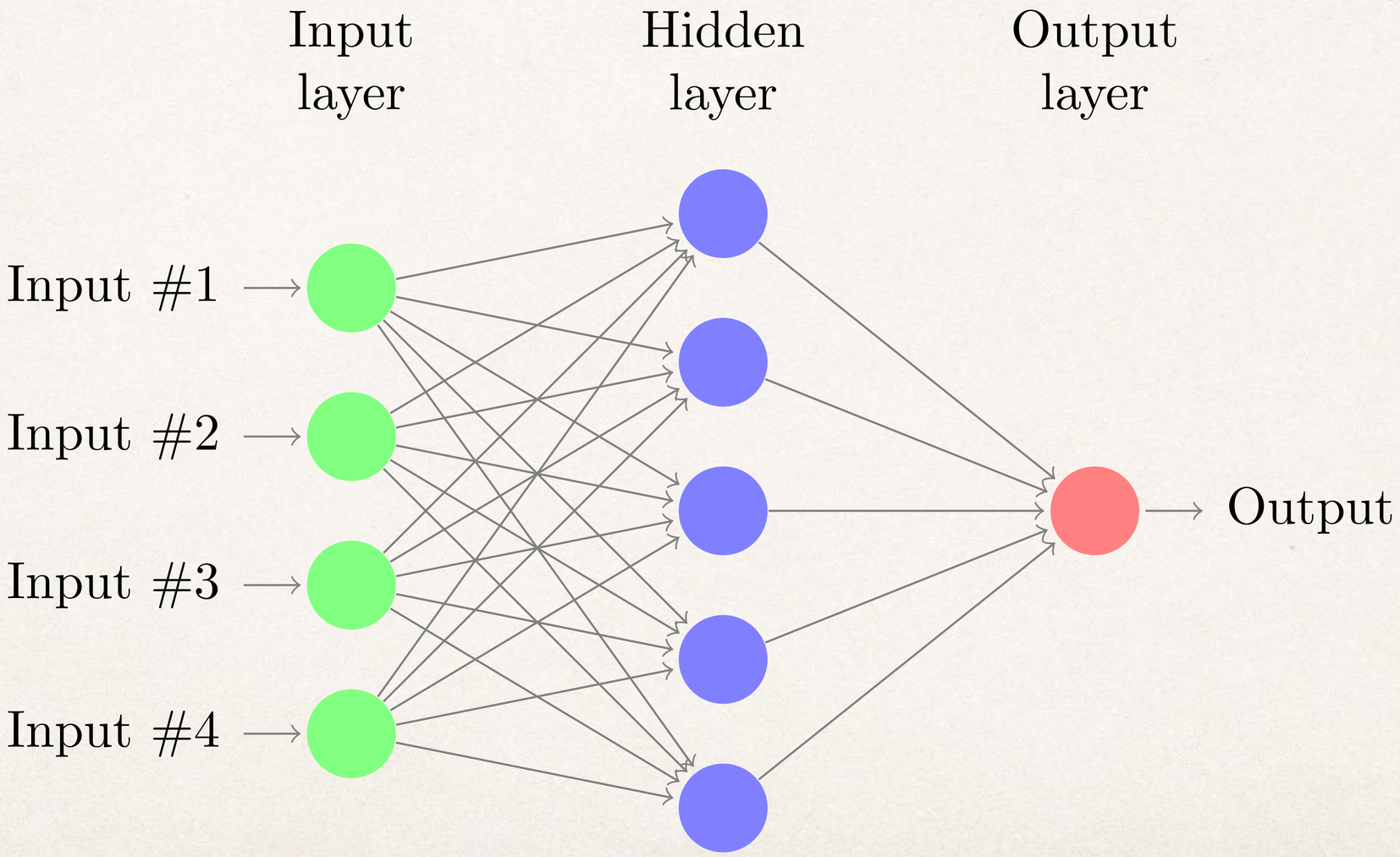
Neural Networks



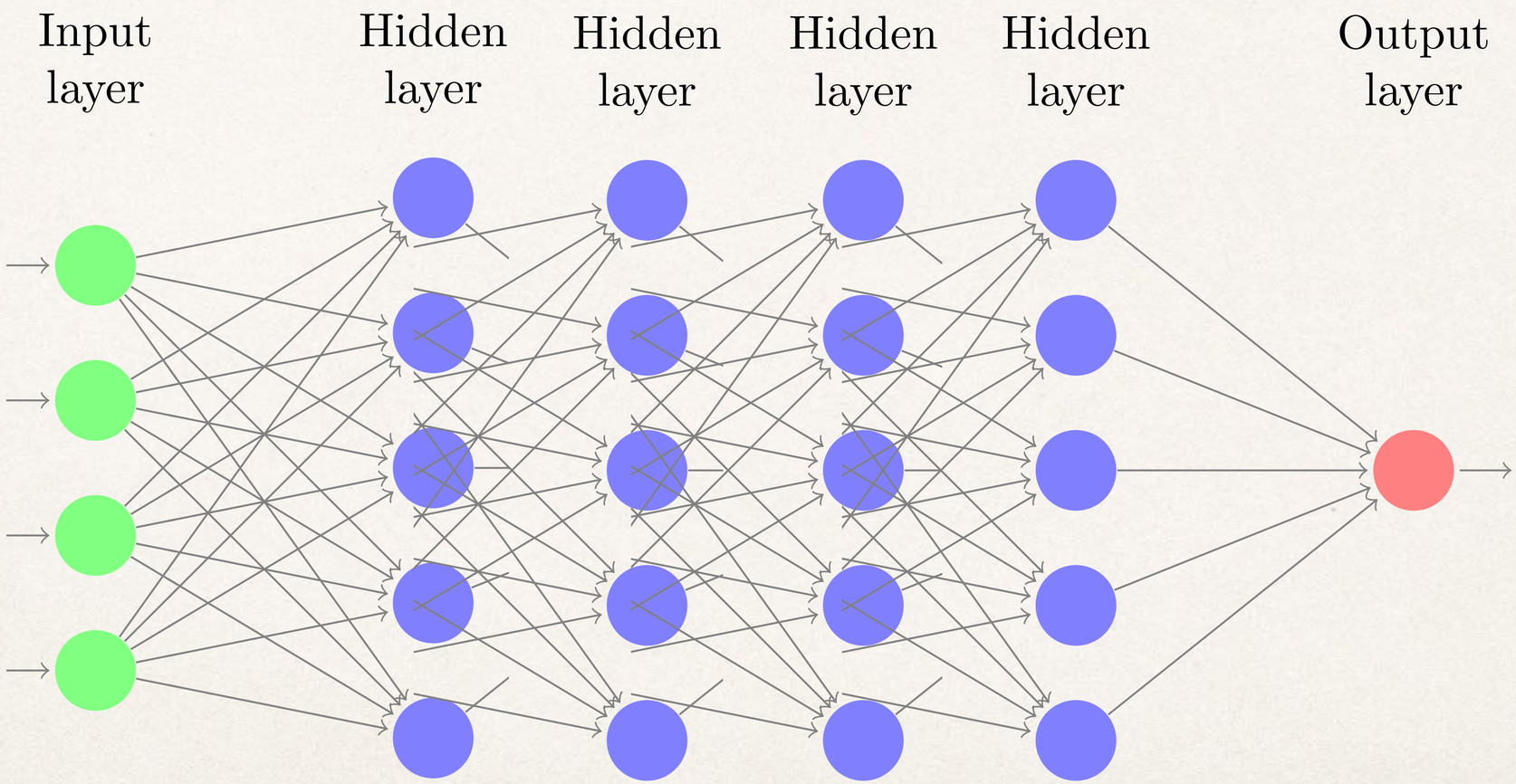
A Node in a Neural Network



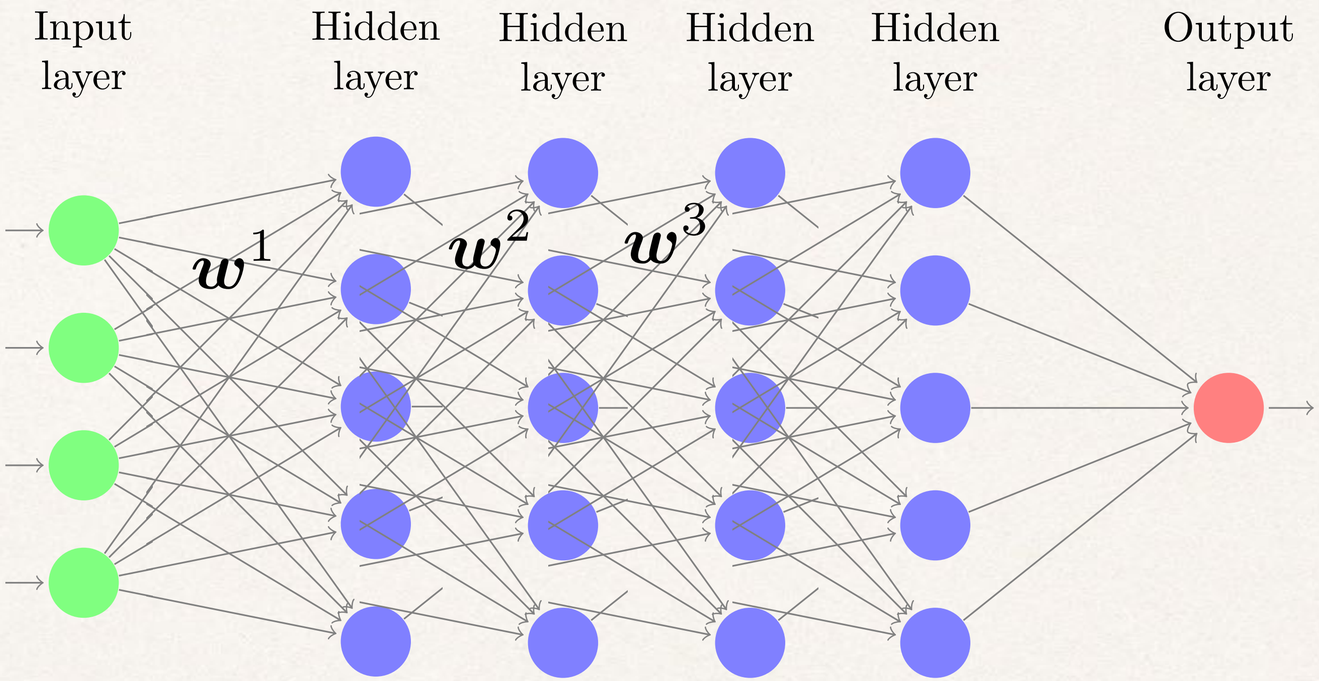
Neural Networks



Deep Neural Networks



Deep Neural Networks



Gradient Descent

Back propagation = chain rule

$$\frac{\partial f}{\partial w_{ij}^l}$$

Convolutional Neural Networks

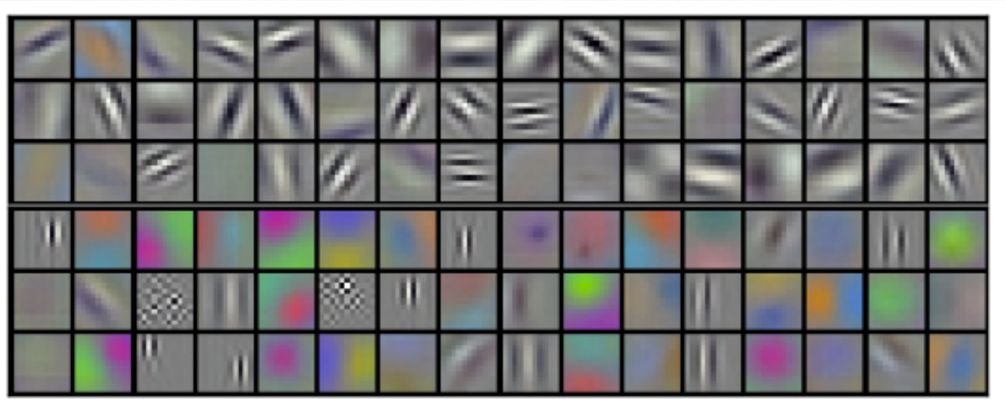
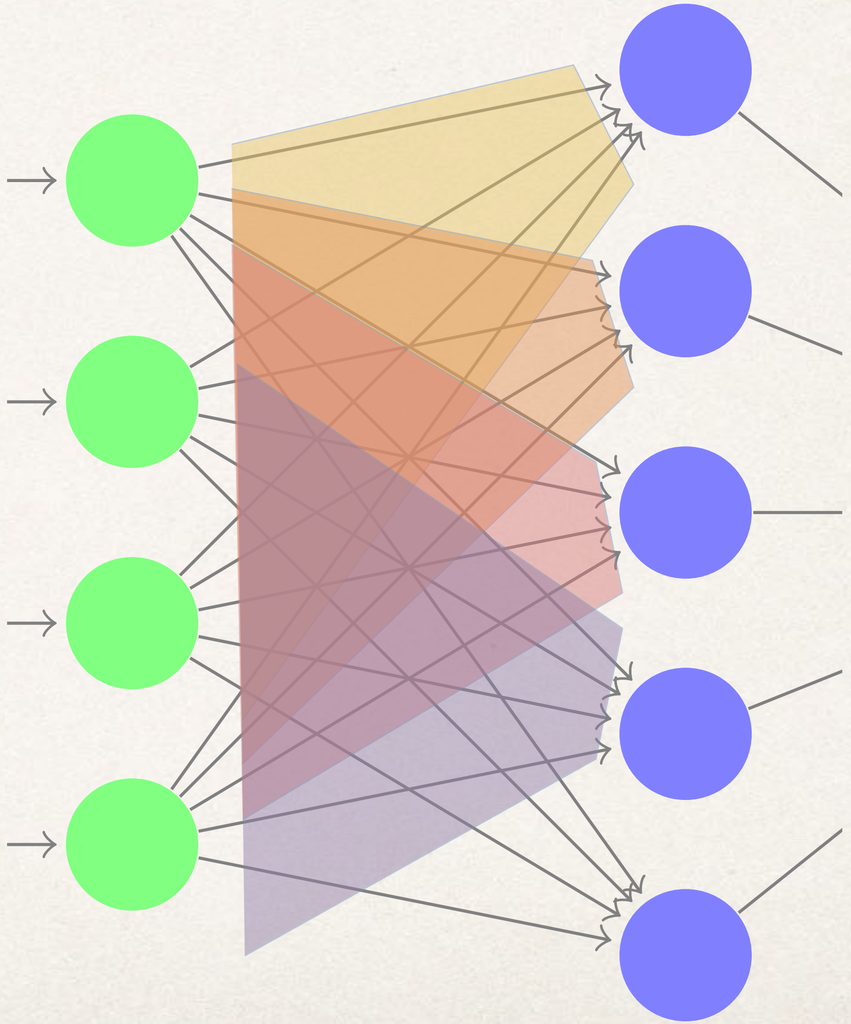


image: AlexNet



Recurrent Neural Networks

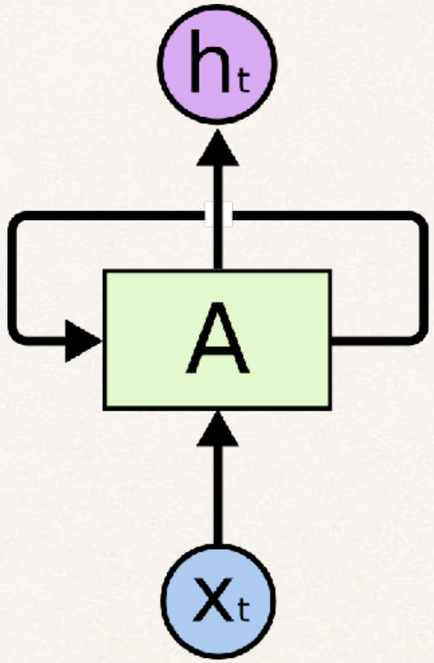


image: colah.github.io

Recurrent Neural Networks

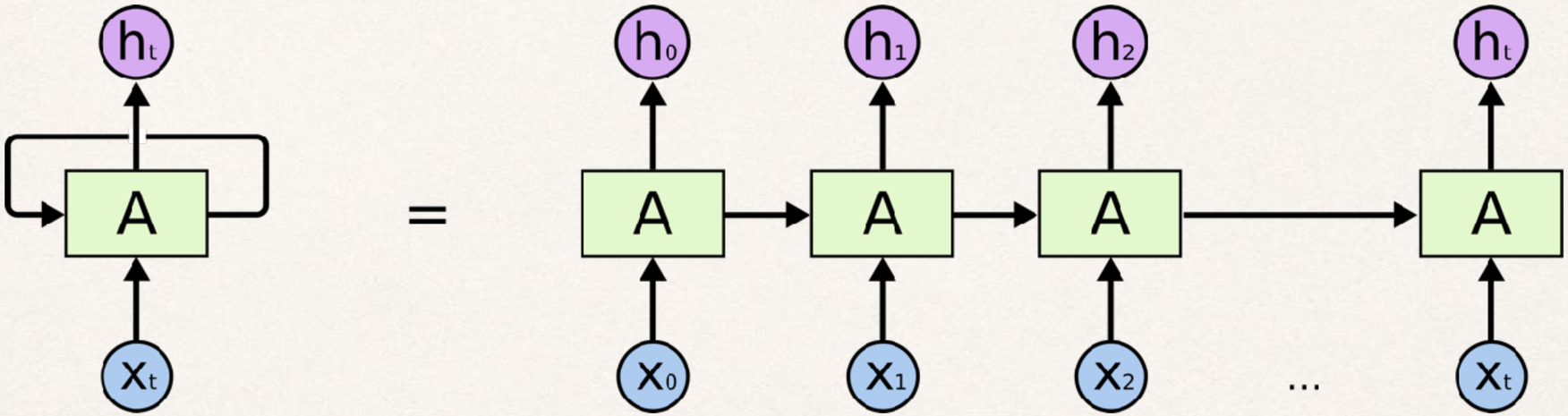
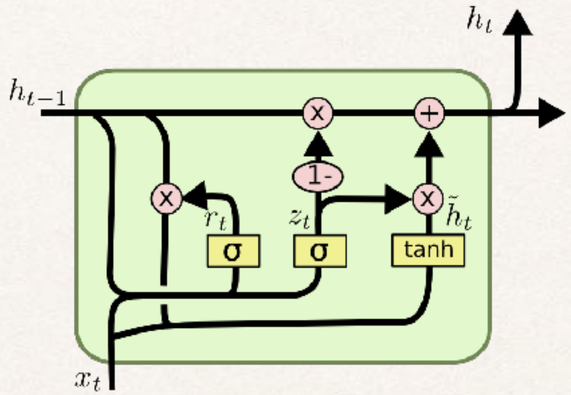


image: colah.github.io

Recurrent Neural Networks

Long Short Term Memory networks (LSTM)



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

image: colah.github.io

Deep Learning

Recommended Material

Lecture:

Andrew Ng Intro: https://youtu.be/n1ViNeWhC24
Nando de Freitas, Oxford
https://youtu.be/PlhFWT7vAEw?list=PLE6Wd9FR--EfW8dtjAuPoTuPcqmOV53Fu

Blogs:

http://karpathy.github.io/2015/10/25/selfie/
http://colah.github.io/

Papers / Reviews:

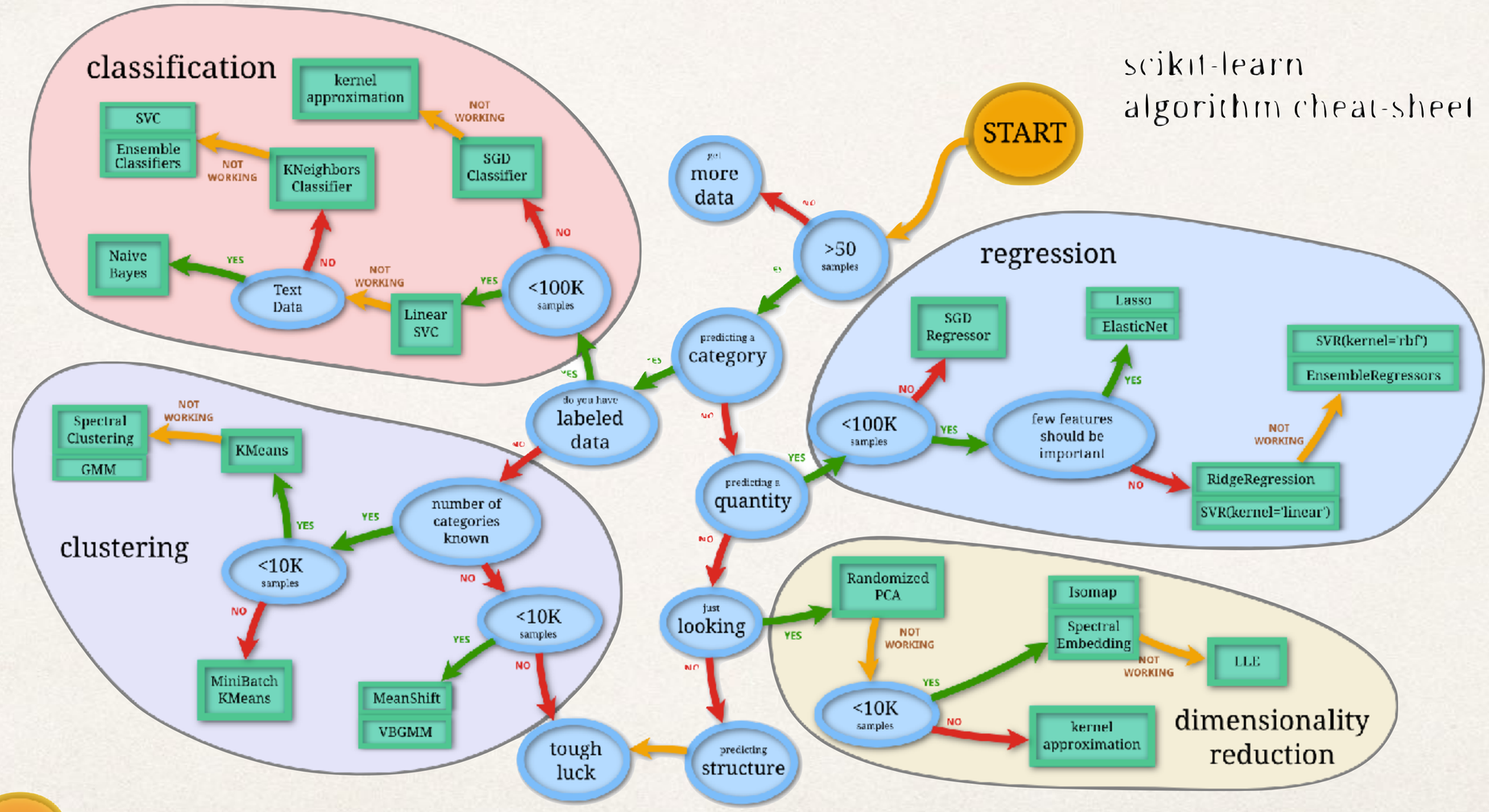
Bengio, Y., Courville, A., & Vincent, P. (2013).
Representation Learning: A Review and New Perspectives.
IEEE Transactions on PAMI. http://doi.org/10.1109/TPAMI.2013.50

Schmidhuber, J. (2015). **Deep learning in neural networks: An overview**.
Neural Networks, *61*, 85–117. http://doi.org/10.1016/j.neunet.2014.09.003

So Which ML Method to Use?

- Supervised Methods
  - *(Linear)* Classification & Regression
  - Kernel Methods, Decision Trees & Random Forests
  - Deep Learning
- Unsupervised Methods
  - Clustering, Dimensionality Reduction
  - Semantic Embeddings, Deep Learning

So Which ML Method to Use?



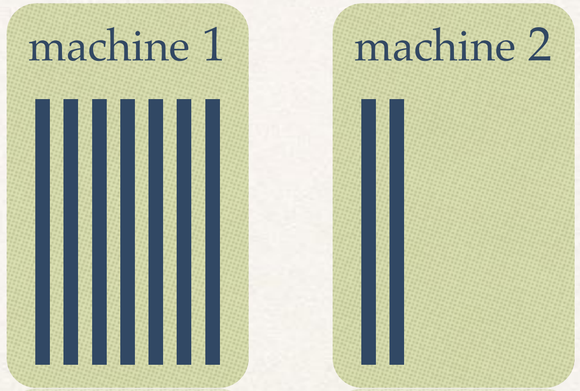
Part II

Big Data Infrastructure

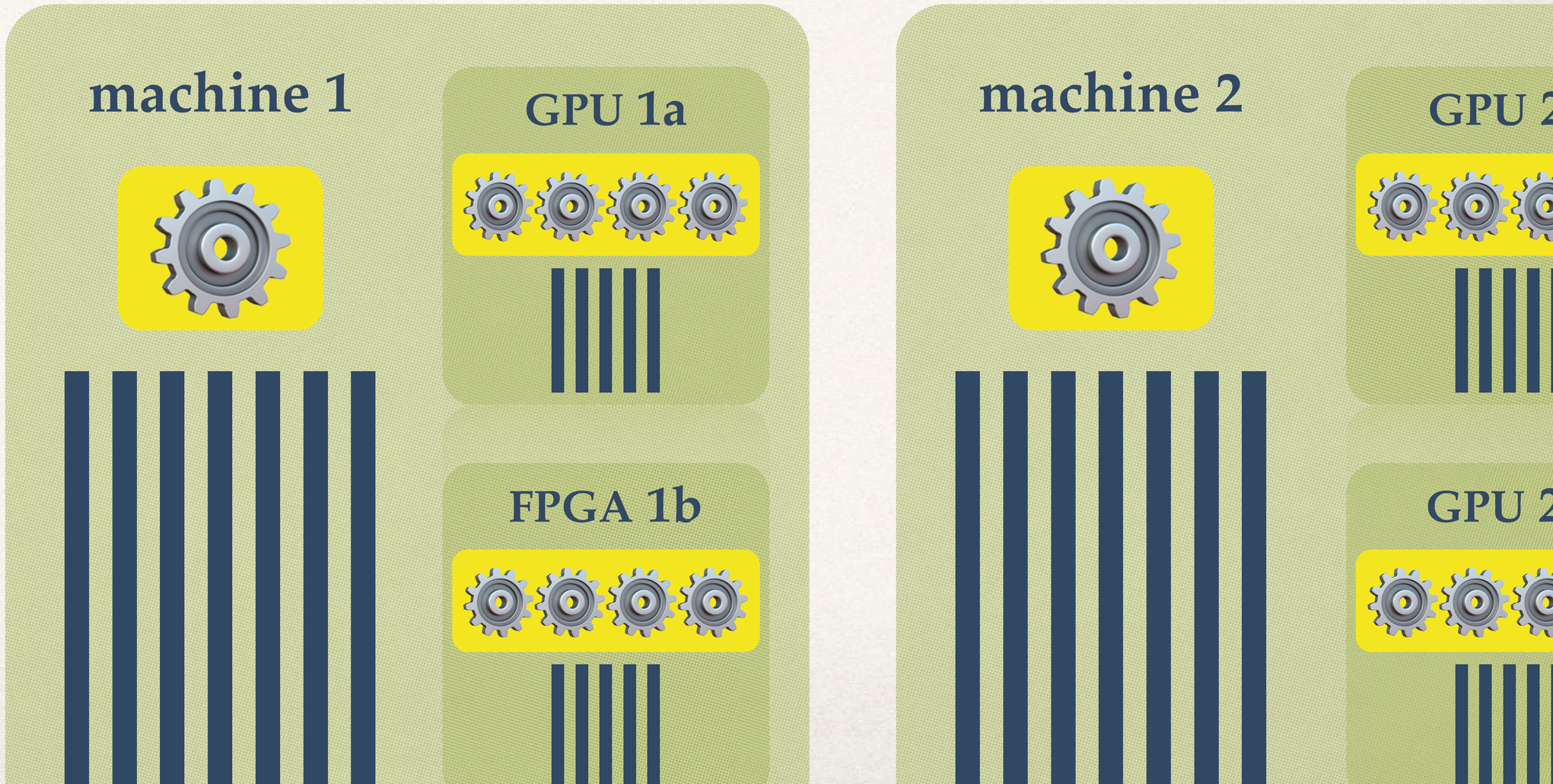
- **Distributed Training**
- **Map Reduce**

Systems & Distributed Machine Learning

What if the data does not fit onto one computer anymore?



ML and Systems



Project:

Distributed Machine Learning Benchmark

Apache

Google

Apache

HPC

Goal:
Public and Reproducible
Comparison of Distributed Solvers

github.com/mlbench/mlbench

Problem 1

The Cost of Communication

$$\boldsymbol{v} \in \mathbb{R}^{100}$$

- Reading $\boldsymbol{v}$ from Memory (RAM)

  100 ns

- Sending $\boldsymbol{v}$ to another Machine

  500'000 ns

- One Typical Map-Reduce Iteration *(Hadoop)*

  10'000'000'000 ns

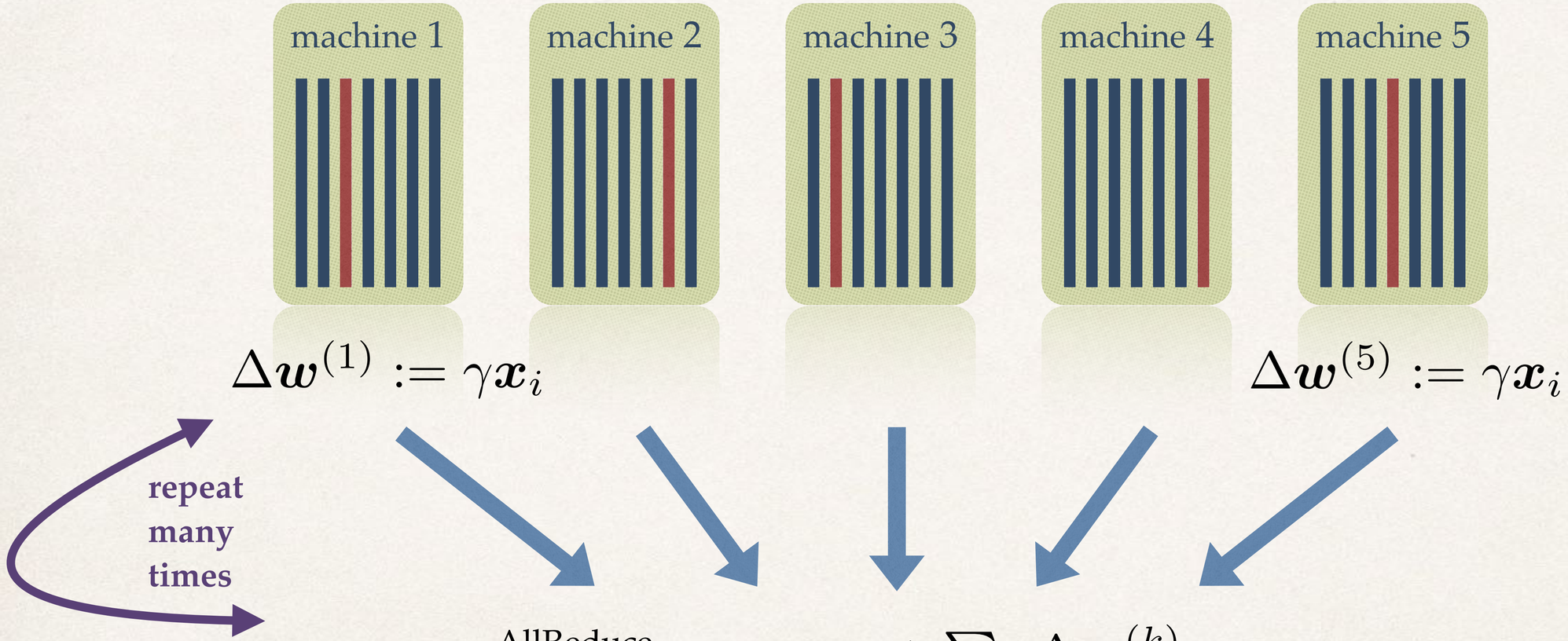
Problem 2

- Parallel Algorithms are Hard
- Single Machine Solvers are Fast

- no **reusability** of good single machine algorithms

Distributed Optimization

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



Naive Distributed SGD

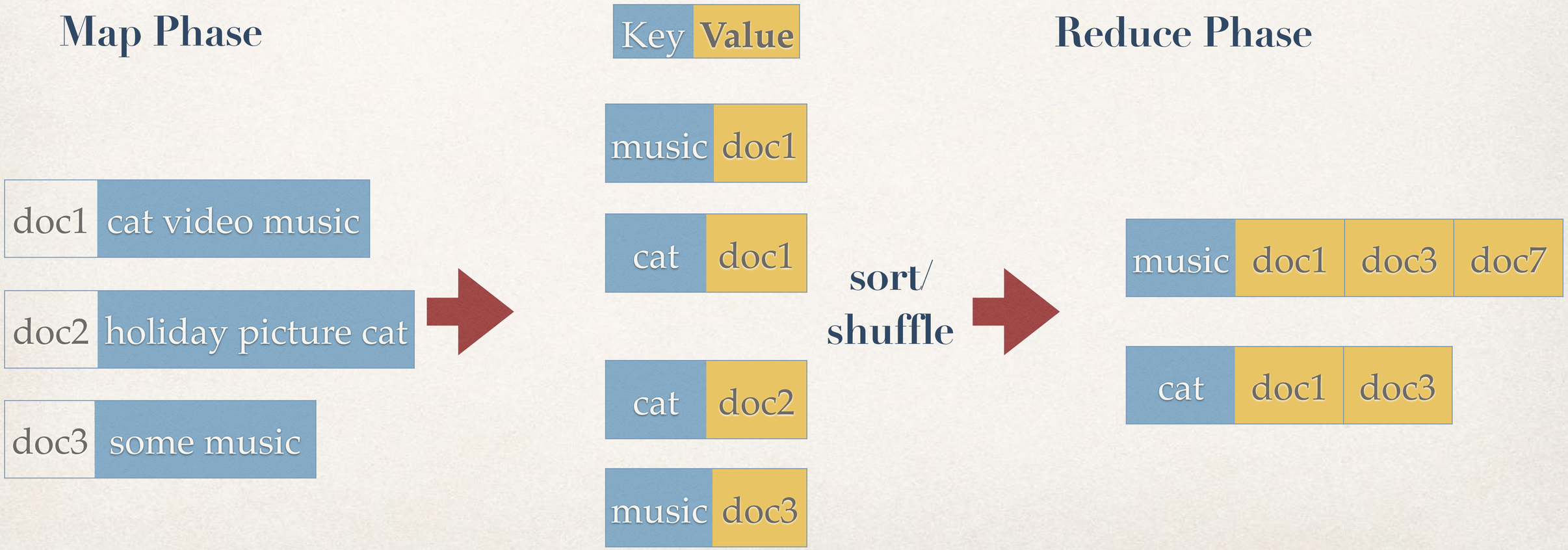
Part II

Big Data Infrastructure

- almost finished for today…

MapReduce

A general framework to easily distribute (certain) tasks across many computers



MapReduce

A general framework to easily distribute (certain) tasks across many computers

- **Highly Scalable**
- **Relatively Slow**

- **Many Applications:**
  - **Build an index** (lookup table) **for search engine**
  - **Data preprocessing**
  - **PageRank** (Matrix x Vector computation)

End

Your turn!
Recommended Practical Pointers

- scikit learn

(python)
- real data applications: kaggle.com
- PyTorch & TensorFlow

LausanneML
ZüriML

Questions?

Martin Jaggi
martin.jaggi@epfl.ch

EPFL Machine Learning and Optimization Laboratory
mlo.epfl.ch