

Machine Learning: Basic Principles

Introduction

Salo, September 2018

Outline

① About Me

② Why Machine Learning?

③ What is Machine Learning?

About Me

- MSc (2008) and Phd (2012) in **electrical engineering/signal processing** at TU Vienna
- Assistant Professor TU Vienna 2013-2015
- since 2015, **Ass. Prof. for Machine Learning** at Aalto CS
- since 2018, **Advisor** for <https://www.miils.com/>

My Research Group

- heading the group “Machine Learning for Big Data”
- currently five Phd students, several MSc and BSc students
- we research **fundamental limits** and **efficient algorithms** for machine learning involving **massive datasets (big data)**
- guiding question(s):

When and How is Machine Learning Possible?

My Teaching

- since 2015, CS-E3210 “Machine Learning: Basic Principles”
(500+? students)
- since 2016, CS-E4020 “Convex Optimization for Big Data”
(50 students)
- from 2018, CS-E4800 “Artificial Intelligence” (500 students)

Outline

- ① About Me
- ② Why Machine Learning?
- ③ What is Machine Learning?

Data Is New Oil



Machine Learning Turns Raw Data Into Insights



ML refines raw data into important insights that can fuel your business!

ML for Finance



given the stock history, will it rise or fall?

ML for Cars



should i stay or should i go?

ML for Elderly People






















given all current available information (vital sensors, temperature, daytime,...), how likely will a person fall at home

ML for Coast Guards



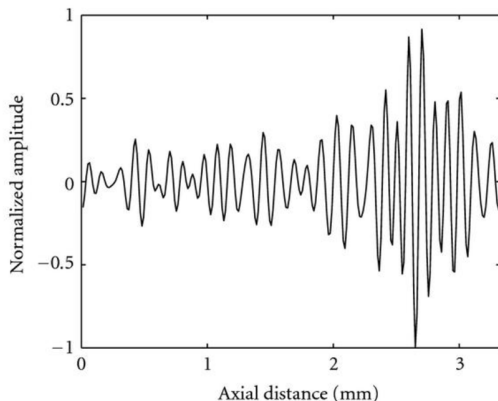
given this satellite image, when, where and how likely will a monster wave hit the coast?

ML for Snapshooters (Image Annotation)

	IMG-20170...0006.jpeg	✓	26 Feb 2017, 20:33	15
	IMG-20170...0008.jpeg	✓	26 Feb 2017, 20:33	24
	IMG-20170...A0010.jpg	✓	26 Feb 2017, 20:33	27
	IMG-20170...A0011.jpeg	✓	26 Feb 2017, 20:33	36
	IMG-20170...A0013.jpg	✓	26 Feb 2017, 20:33	17
	IMG-20170...A0014.jpeg	✓	26 Feb 2017, 20:33	32
	IMG-20170...A0007.jpg	✓	26 Feb 2017, 20:33	8
	IMG-20170...A0008.jpg	✓	26 Feb 2017, 20:33	26
	IMG-20170...A0009.jpg	✓	26 Feb 2017, 20:33	41
	IMG-20170...A0010.jpg	✓	26 Feb 2017, 20:33	18
	IMG-20170...A0011.jpg	✓	26 Feb 2017, 20:33	37
	IMG-20170...A0012.jpg	✓	26 Feb 2017, 20:33	28
	IMG-20170...A0013.jpeg	✓	26 Feb 2017, 20:33	19
	IMG-20170...A0015.jpeg	✓	26 Feb 2017, 20:33	16
	IMG-20170...A0017.jpeg	✓	26 Feb 2017, 20:33	18
	IMG-20170...A0019.jpg	✓	26 Feb 2017, 20:33	18
	IMG-20170...A0020.jpg	✓	26 Feb 2017, 20:33	18
	IMG-20170...0006.jpeg	✓	26 Feb 2017, 20:33	33
	IMG-20170...0008.jpeg	✓	26 Feb 2017, 20:33	34

how to automatically rename pile of snapshots with **meaningful** filenames?

ML for Signal Processing



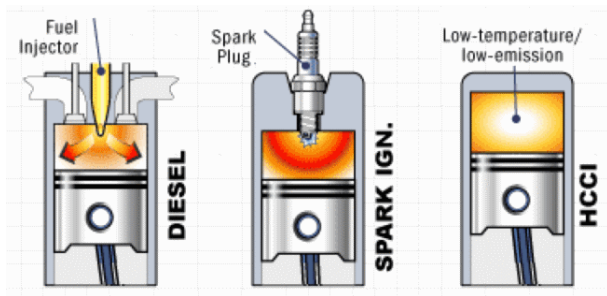
given this signal received by your smartphone antenna, did transmitter send “I LOVE YOU” or “F*** OFF” ?

ML for Coffee Connoisseurs



has this coffee been produced under **fair working conditions**?

ML for Reducing Emissions

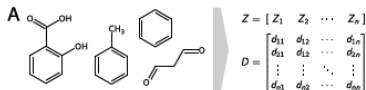


what is the optimum timing of fuel injection?

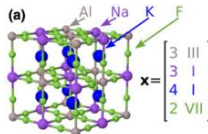
<https://youtu.be/qQG7ocnE3EA>

ML for Chemistry

ML for Chemists



Machine Learning Energies of
2 Million Elpasolite (ABC_2D_6)
Crystals



what are the chemical properties of a given molecule? can we design new (unseen) structures with desired physicochemical properties?

figure donated by Prof. Mikko O. Hakala

ML for Forest Fire Management



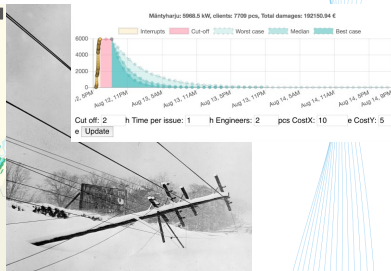
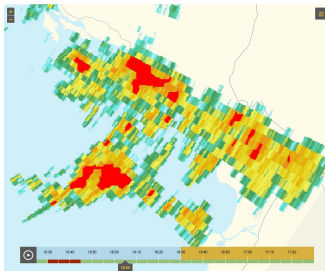
where is a fire most likely to start?

ML for Finnish Meteorological Institute



FINNISH METEOROLOGICAL INSTITUTE

ML for power grid



will this storm cause damage? where? how bad?

figure donated by Roope Tervo

ML for Farmers



given this snapshot and weather forecast, when is optimum time to harvest?

ML for Cyclists



given this snapshot, will it rain in the next 2 hours? how **reliable** is the prediction?

ML for Hedonists



given this snapshot, where can i find the next Alko?

ML for Robots



given a video of human motion, how to imitate it?

ML for Mathematicians

Theorem 4. *Consider a clustered graph signal $\mathbf{x}_c \in \mathcal{X}$ (cf. (5)) which is observed only at the sampling set $\mathcal{M} \subseteq \mathcal{V}$. If $\text{NNSP}(\mathcal{M}, \mathcal{F})$ holds, then the solution of (4) is unique and coincides with \mathbf{x}_c .*

given the theorem formulation, how would a solid proof look like?

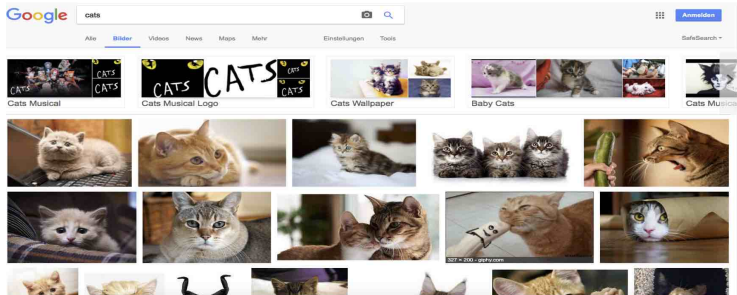
Outline

- ① About Me
- ② Why Machine Learning?
- ③ What is Machine Learning?

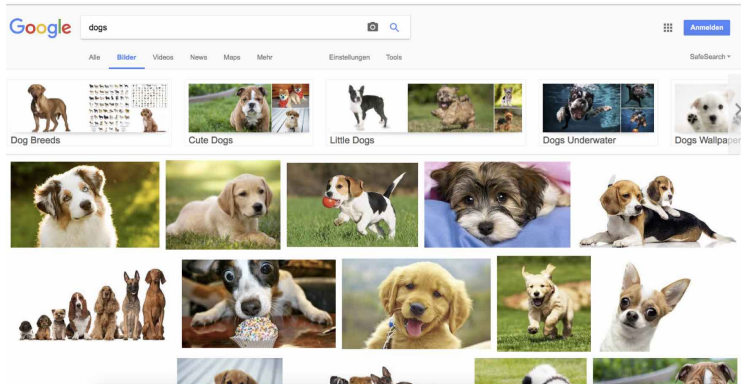
What is Machine Learning?

- program computers to perform a particular task (find oil, find next Alko,...) by generalizing from examples
- program computers to extract patterns from raw data...
- ML based on “trial and error”
- e.g., program computer to distinguish between dog and cat by looking at tons of of dog and cat pictures

Database of Cat Pictures

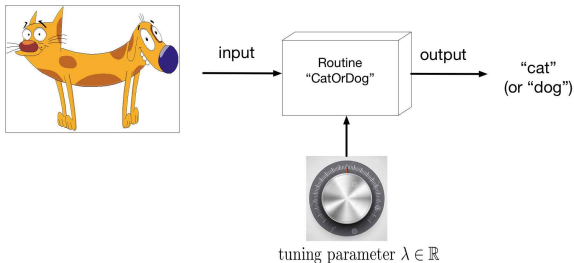


Database of Dog Pictures



The Classifier

a “classifier” is a (Python,..) routine which takes an image and parameter λ as input and outputs either “Dog” or “Cat”



ML is about finding the best value for tuning parameter λ

Optimize using Training Data

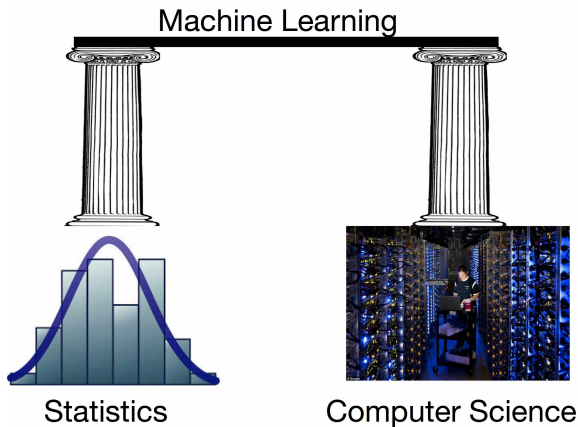
- consider routine $CatOrDog(image\ input, scalar\ \lambda)$
- for particular choice of λ count error on database
$$f(\lambda) := \text{nr. of pictures which are classified wrong}$$
- vary λ to minimize $f(\lambda)$
- “train” classifier using image database (e.g., ImageNet)
- database serves as “training data”

Three Main Components

a ML method is typically based on **three components**

- data model **including representation (features)**
- **performance measure/loss/risk**
- **optimization** to maximize performance (minimize error)

Two Main Pillars of Machine Learning



Statisticians Viewpoint on Data

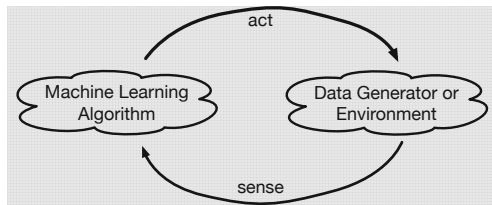
- consider list of numbers $\mathbf{x}^{(1)}=5, \mathbf{x}^{(2)}=50, \dots, \mathbf{x}^{(N)}=30$
- statistician thinks about underlying **statistical model**
- e.g., each $\mathbf{x}^{(i)}$ is i.i.d. copy of $\mathbf{x} \sim \mathcal{N}(\mu, \sigma^2 \mathbf{I})$
- how to **fit** mean μ and variance σ^2 to $\mathbf{x}^{(1)}, \dots$?
- what can we **predict** from $\mathbf{x}^{(1)}, \dots$?

Computational Viewpoint on Data

- consider list of numbers $\mathbf{x}^{(1)}=5, \mathbf{x}^{(2)}=50, \dots, \mathbf{x}^{(N)}=30$
- determining $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}^{(i)}$ is a **computational problem**
- how to compute $\hat{\mu}$ for $N = 10^{10}$?
- how to compute $\hat{\mu}$ if numbers are stored on **different servers**?
- is there an **optimum way** to compute $\hat{\mu}$?

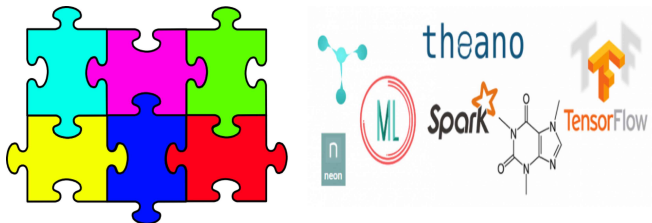
Active vs. Passive Learning

- in this course we assume dataset given (passive learning)
- sometimes we can influence data generation (active learning)
- interaction between learning algorithm and data generator
- artificial intelligence \approx active machine learning



Software

- all methods considered in this course **already implemented**
- powerful libraries (Theano, Tensorflow,...) available
- the challenge is to **combine pieces in the right way**



ML Engineering is about Choosing Right Tools

