Artificial Intelligence: News and Questions

Michele Sebag
CNRS – INRIA – Univ. Paris-Saclay

ArenbergSymposium – Leuven – Nov. 27th, 2019

Credit for slides: Yoshua Bengio; Yann LeCun; Nando de Freitas; Léon Gatys; Max Welling; Victor Berger

Université Paris-Saclay in 1 slide

14 partners

- 3 Univ.
- 4 Grandes coles
- 7 Research Institutes

Wendelin Werner (mathematics), Valérie Masson-Delmotte (climate), Alain Aspect (physics), Patrick Couvreur (medicine), Albert Fert (nanoscience), Stanislas Dehaene (neuroscience)...

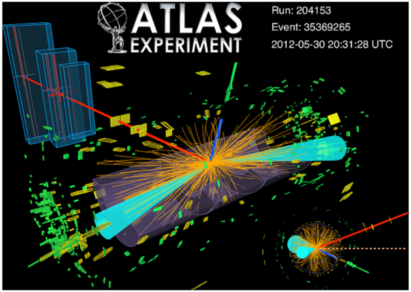
15% of French Research

5,500 PhD; 10,000 Faculty members

10 Field medals; 3 Nobel; 160 ERC.

Some AI projects

- Center for Data Science
  ML Higgs Boson Challenge (2015)
- Institute DataIA: AI for Society
- Big Data, Optimization and
  Energy (See4C challenge)



Two visions of AI – 1950 - 1960

Logical calculus can be achieved by machines !

- All men are mortal.
- Socrates is a man.
- Therefore, Socrates is mortal.

Primary operation: Deduction (reasoning) ? or Induction (learning)?

- Should we learn what we can reason with ?
- Should we reason with what we can learn ?

	Alan Turing	John McCarthy
HOW	Learning	Reasoning
VALIDATION	Human assessment	Pb Solving

Alan Turing (1912-1954)

Muggleton, 2014

1950: Computing Machinery and Intelligence

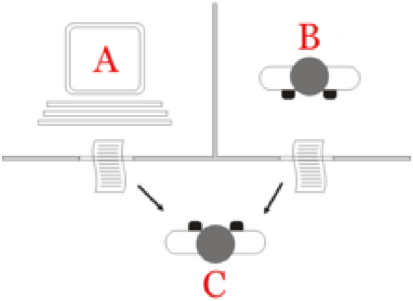
- Storage will be ok
- But programming needs *prohibitively large human resources*
- Hence, machine learning.

by (...) mimicking education, we should hope to modify the machine until it could be relied on to produce definite reactions to certain commands.

One could carry through the organization of an intelligent machine with only two interfering inputs, one for pleasure or reward, and the other for pain or punishment.

The imitation game

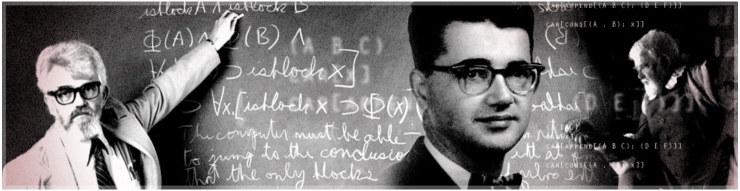
The Turing test



Issues

- Human assessment; no golden standard.

John McCarthy (1927-2011)



1956: The Dartmouth conference

- With Marvin Minsky, Claude Shannon, Nathaniel Rochester, Herbert Simon, et al.
- *The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.*

Computational Logic for AI

- Declarative languages
- Symbolic methods, deduction
- Focussed domains (expert systems)
- Games and problem solving

Issues

- How to ground symbols ?
- Where does knowledge come from ?

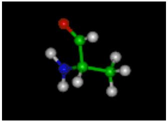
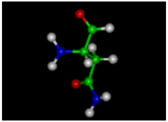
Automating Science using Robot Scientists

King et al, 04-19

From facts to hypotheses to experiments to new facts...

- A proper representation and domain theory

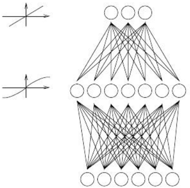
$$Benzene(A_1, A_2, A_3, A_4, A_5, A_6) :- Carbon(A_1), Carbon(A_2), \ldots$$
$$Bond(A_1, A_2), Bond(A_2, A_3), Bond(A_3, A_4) \ldots$$

- Active Learning — Design of Experiments
- Control of noise

The Wave of AI

Neural Nets, ups and downs in AI



(C) David McKay - Cambridge Univ. Press

History

1943	A neuron as a computable function $y = f(\mathbf{x})$ Intelligence $\rightarrow$ Reasoning $\rightarrow$ Boolean functions	Pitts, McCullough
1960	Connexionism + learning algorithms	Rosenblatt
1969	AI Winter	Minsky-Papert
1989	Back-propagation	Amari, Rumelhart & McClelland, LeCun
1992	NN Winter	Vapnik
2005	Deep Learning	Bengio, Hinton

The revolution of Deep Learning

- Ingredients were known for decades
- Neural nets were no longer scientifically exciting (except for a few people)
- Suddenly... 2006

Revival of Neural Nets

Bengio, Hinton 2006

1. Grand goal: AI

2. Requisites
   - Computational efficiency
   - Statistical efficiency
   - Prior efficiency: architecture relies on human labor
3. Compositionality principle: skills built on the top of simpler skills

Piaget 1936

Revival of Neural Nets

Bengio, Hinton 2006

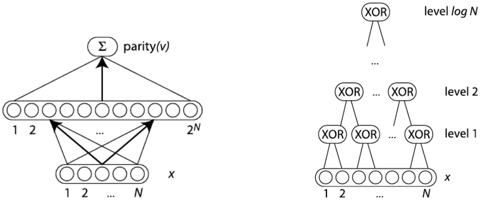
1. Grand goal: AI
2. Requisites
   - Computational efficiency
   - Statistical efficiency
   - Prior efficiency: architecture relies on **student** labor
3. Compositionality principle: skills built on the top of simpler skills

Piaget 1936

The importance of being deep

A toy example: *n*-bit parity

Hastad 1987



Pros: efficient representation

Deep neural nets are exponentially more compact

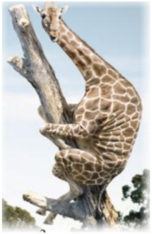
Cons: poor learning

- More layers → more difficult optimization problem
- Getting stuck in poor local optima.
- **Handled through smart initialization** Glorot et al. 10

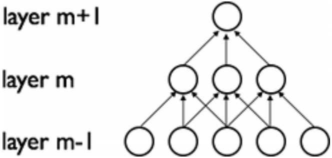
Convolutional NNs: Enforcing invariance

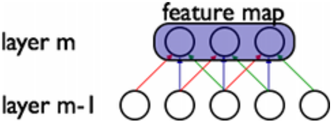
LeCun 98

Invariance matters



- Visual cortex of the cat — Hubel & Wiesel 68
  - cells arranged in such a way that
  - ... each cell observes a fraction of the visual field — receptive field
  - ... their union covers the whole field



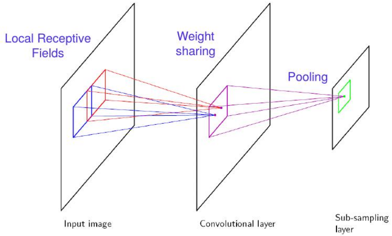


- Layer m: detection of local patterns (same weights)
- Layer $m+1$: non linear aggregation of output of layer m

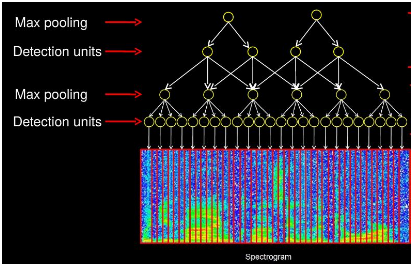
Convolutional architectures

LeCun 1998

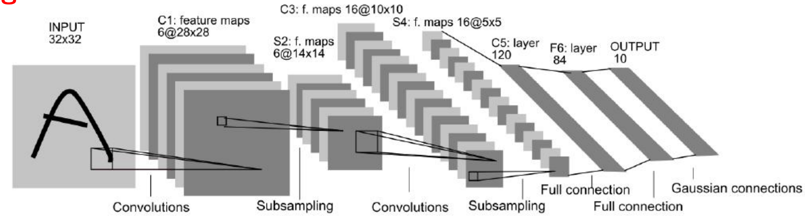
For images



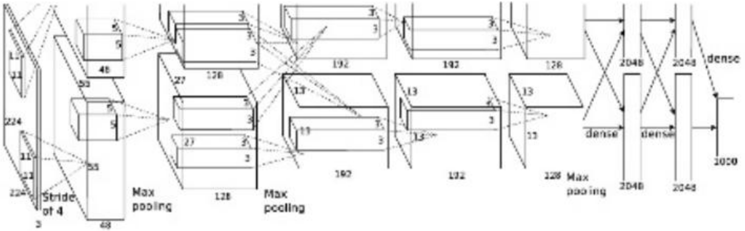
For signals



Gigantic architectures



LeCun 1998



Kryzhevsky et al. 2012

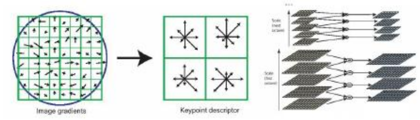
Properties

- Invariance to small transformations (over the region)
- Reducing the number of weights by several orders of magnitude

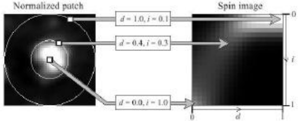
What is new ?

Former state of the art

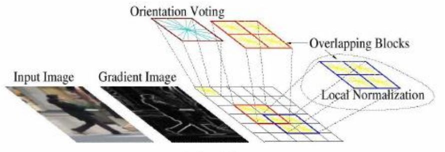
e.g. in computer vision



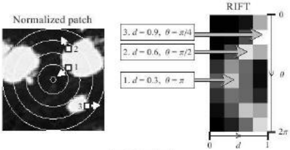
SIFT



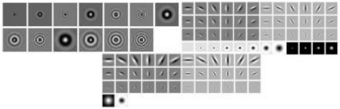
Spin image



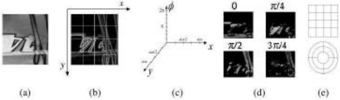
HoG



RIFT



Textons



GLOH

SIFT: scale invariant feature transform
HOG: histogram of oriented gradients
Textons: “vector quantized responses of a linear filter bank”

What is new, 2

Traditional approach

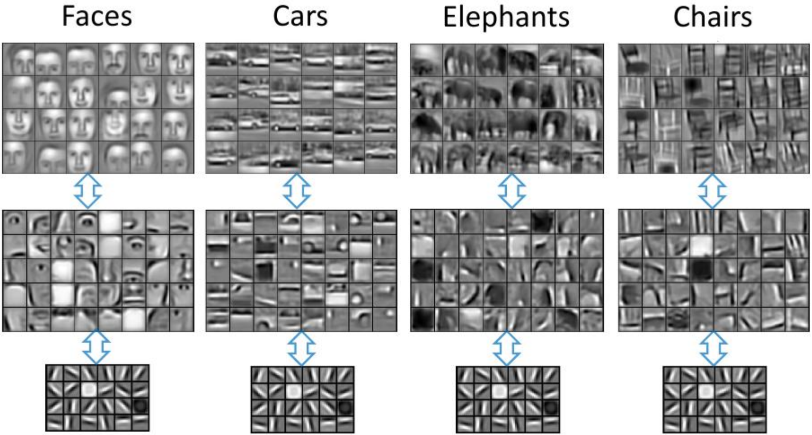
→ Manually crafted features → Trainable classifier

Deep learning

→ Trainable feature extractor → Trainable classifier

A new representation is learned

Bengio et al. 2006



Why Deep Learning now ?

CONS

- a non-convex optimization problem
- no theorems
- delivers a black box model

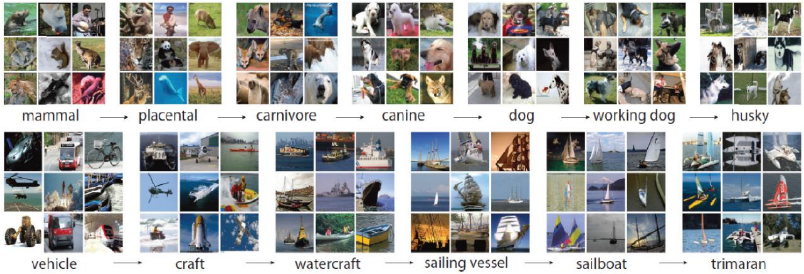
PROS

- Linear complexity w.r.t. #data
- Performance leaps if **enough data and enough computational power**.

A leap in the state of the art: ImageNet

Deng et al. 12

15 million labeled high-resolution images; 22,000 classes.



Large-Scale Visual Recognition Challenge

- 1000 categories.
- 1.2 million training images,
- 50,000 validation images,
- 150,000 testing images.

A leap in the state of the art, 2

2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

2013 Teams	%error
Clarifai (NYU spinoff)	11.7
NUS (singapore)	12.9
Zeiler-Fergus (NYU)	13.5
A. Howard	13.5
OverFeat (NYU)	14.1
UvA (Amsterdam)	14.2
Adobe	15.2
VGG (Oxford)	15.2
VGG (Oxford)	23.0

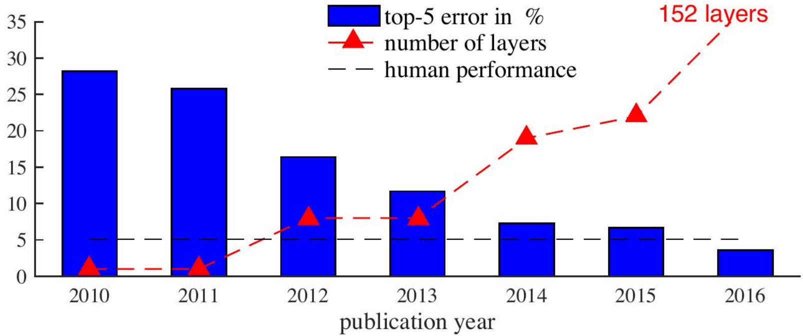
2014 Teams	%error
GoogLeNet	6.6
VGG (Oxford)	7.3
MSRA	8.0
A. Howard	8.1
DeeperVision	9.5
NUS-BST	9.7
TTIC-ECP	10.2
XYZ	11.2
UvA	12.1

shallow approaches

deep learning

Y. LeCun StatLearn tutorial

Super-human performances



2012 Alex Net

2013 ZFNet

2014 VGG

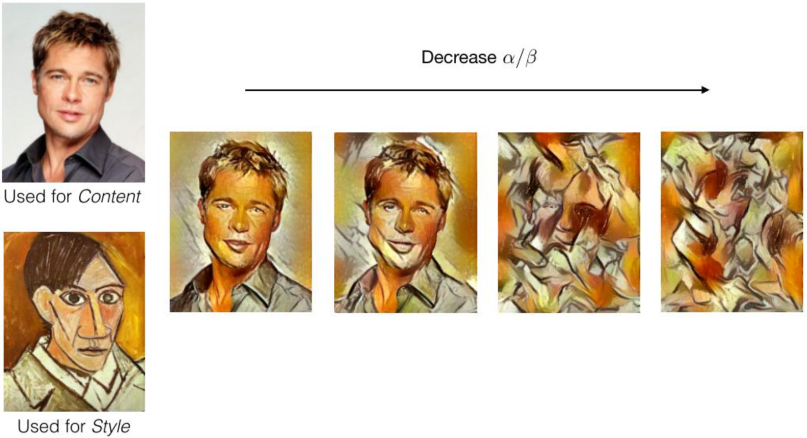
2015 GoogLeNet / Inception

2016 Residual Network

karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/

Playing with representations

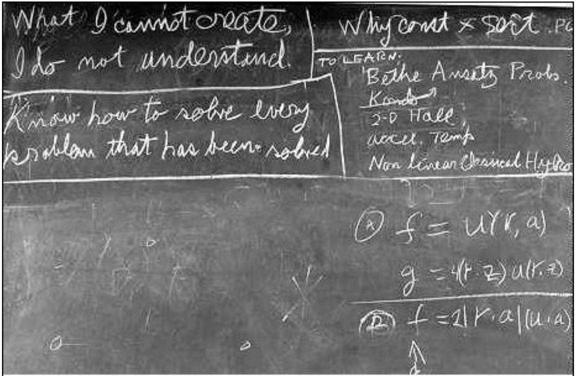
Gatys et al. 15, 16



- Style and contents in a convolutional NN are separable
- Use a trained VGG-19 Net:
  - applied on image 1 (content)
  - applied on image 2 (style)
  - find input matching hidden representation of image 1 (weight α) and hidden representation of image 2 (weight β)

Beyond classifying, modifying data: generating data

"What I cannot create I do not understand" Feynman 88



Generative Adversarial Networks

Goodfellow et al., 14

Goal: Find a generative model

- Classical: learn a distribution hard
- Idea: replace a distribution evaluation by a 2-sample test

Principle

- Find a good generative model, s.t. generated samples **cannot be discriminated** from real samples

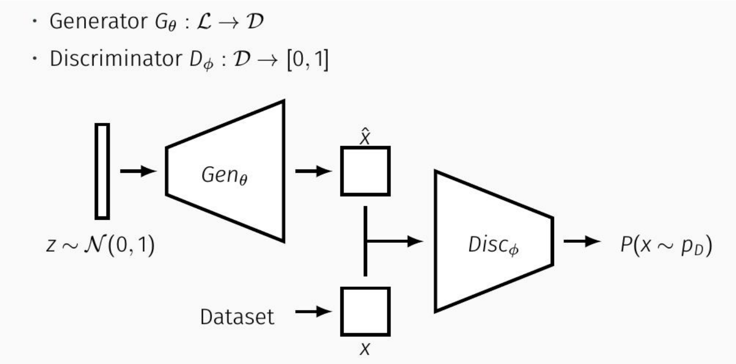
(not easy)

Generative Adversarial Networks, 2

Goodfellow, 2017

Elements

- Dataset, true samples $\mathbf{x}$ (**real**)
- Generator G, generated samples (**fake**)
- Discriminator D: discriminates *fake* from *real*



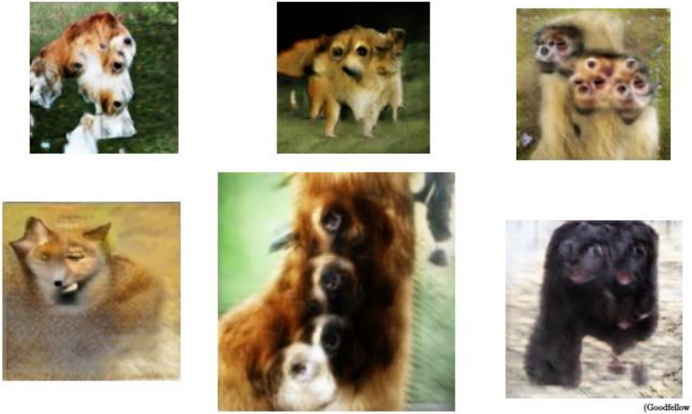
A min-max game **Embedding a Turing Test !**

$$Min_G \ Max_D \mathbb{E}_{x \in data}[\log(D(\mathbf{x}))] + \mathbb{E}_{z \sim p_x(z)}[\log(1 - D(z))]$$

Generative Adversarial Networks: successes

Mescheder, Geiger and Nowozin, 2018

and monsters



(Goodfellow 2016)

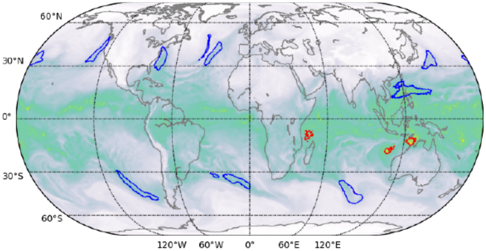
From Deep Learning to Differentiable Programming

Principle

- Most programs can be coded as a neural net.
- Define a performance criterion
- Let the program interact with the world *and train itself*:
  programs that can learn

Revisiting Partial Differential Equations

1. Combining with simulations: recognizing Tropical Cyclones Kurth et al. 18



1152x768 spatial grid, 3 hours time step
3 classes: TCs (0.1%), Atmospheric Rivers (1.7%) and Background

Physics Informed Deep Learning

Data driven solutions of PDE

Raissi 19

Equation:

$$u_t = \mathcal{N}(t, u, x, u_x, u_{xx}, \ldots)$$

residual:

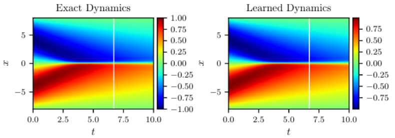
$$f := u_t - \mathcal{N}(t, u, x, u_x, u_{xx}, \ldots)$$

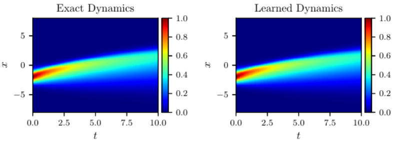
Initial and boundary conditions: $(t_u^i, x_u^i, u^i), i = 1 \ldots N$

Training points $(t_f^j, x_f^j), j = 1 \ldots N'$

Train $u(t, s)$ minimizing

$$\frac{1}{N}\sum_{i=1}^{N} |u(x_u^i, t_u^i) - u^i|^2 +$$





train

$u(0, x) = sin(\pi x/8)$

test

$u(0, x) = -exp(-(x+2)^2)$

Physics Informed Deep Learning

Data driven solutions of PDE

Raissi 19

Equation:

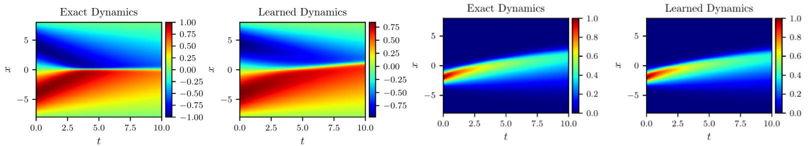
$$u_t = \mathcal{N}(t, u, x, u_x, u_{xx}, \dots)$$

residual:

$$f := u_t - \mathcal{N}(t, u, x, u_x, u_{xx}, \dots)$$

Initial and boundary conditions: $(t_u^i, x_u^i, u^i), i = 1 \dots N$
Training points $(t_f^j, x_f^j), j = 1 \dots N'$
Train $u(t, s)$ minimizing

$$\frac{1}{N} \sum_{i=1}^{N} |u(x_u^i, t_u^i) - u^i|^2 +$$



test $u(0, x) = sin(\pi x/8)$

train $u(0, x) = -exp(-(x + 2)^2)$

Issues with black-box models

Good performances $\not\Rightarrow$ Accurate model

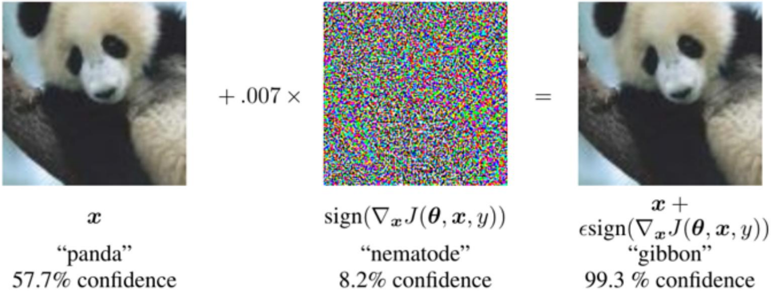




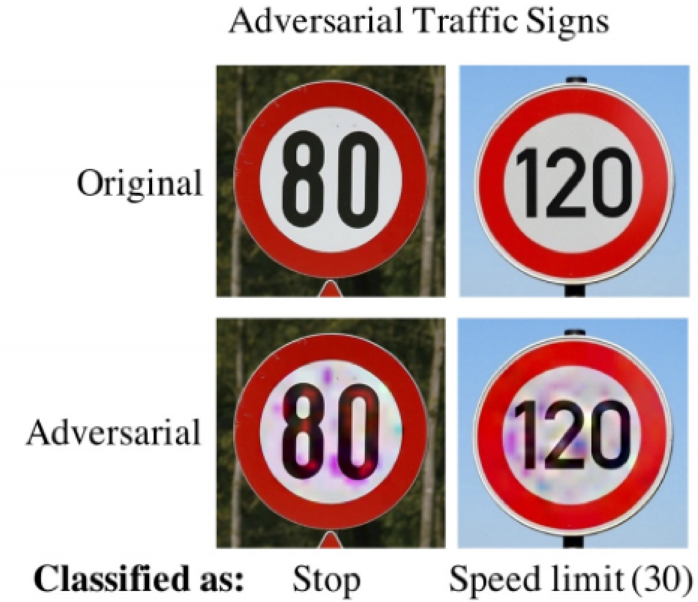
Robustness wrt perturbations

What can happen when perturbing an example ? Anything !

Malicious perturbations Goodfellow et al. 15



Adversarial examples



Artificial Intelligence / Machine Learning / Data Science

A Case of Irrational Scientific Exuberance

- Underspecified goals — Big Data cures everything
- Underspecified limitations — Big Data can do anything (if big enough)
- Underspecified caveats — Big Data and Big Brother

Wanted: An AI with common decency

- Fair — no biases
- Accountable — models can be explained
- Transparent — decisions can be explained
- Robust — w.r.t. malicious examples

ML & AI, 2

In practice

- Data are ridden with biases
- Learned models are biased (prejudices are transmissible to AI agents)
- Issues with robustness
- Models are used out of their scope

More

- C. O'Neill, *Weapons of Math Destruction*, 2016
- Zeynep Tufekci, *We're building a dystopia just to make people click on ads*, Ted Talks, Oct 2017.

Machine Learning: discriminative or generative modelling

Given a training set iid samples $\sim P(X, Y)$

$$\mathcal{E} = \{(\mathbf{x}_i, y_i), \mathbf{x}_i \in \mathbb{R}^d, i \in [[1, n]]\}$$

Find

- Supervised learning: $\hat{h} : X \mapsto Y$ or $\widehat{P}(Y|X)$
- Generative model $\widehat{P}(X, Y)$

Predictive modelling might be based on correlations

If umbrellas in the street, Then it rains

The implicit big data promise:

If you can predict what will happen,

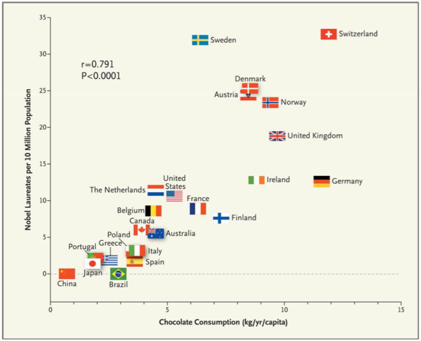
then how to make it happen what you want ?

Knowledge → Prediction → Control

ML models will be expected to support *interventions*:

- health and nutrition
- education
- economics/management
- climate

Correlations do not support interventions



F. H. Messerli: *Chocolate Consumption, Cognitive Function, and Nobel Laureates*, N Engl J Med 2012

Causal models are needed to support interventions

Consumption of chocolate enables to predict # of Nobel prizes
but eating more chocolates does not increase # of Nobel prizes

Discussion

Scientific Caveat

- Robustness of results
- Reproducibility of results (gigantic resources)

Economic Caveat

- Winner take all:

Value $\rightarrow$ **Data** $\rightarrow$ **More Value**

Societal Caveat

- Learning from biased data $\rightarrow$ carving prejudices in stone
- Accurate prediction of individual risks $\rightarrow$ ruins insurance mechanisms.