Philosophy of Machine Learning
UC Irvine 2018

# Data-driven Science and Undercover Theory

⋆

Jan-Willem Romeijn
University of Groningen

# Bacon’s entomology

The theme of my talk is already present in Bacon’s Novum Organon:

> *The men of experiment are like the ant, they only collect and use; the reasoners resemble spiders, who make cobwebs out of their own substance. But the bee takes a middle course: it gathers its material from the flowers of the garden and of the field, but transforms and digests it by a power of its own.*
>
> Francis Bacon, The New Organon [Book One], 1620.

This talk is in support of bees.

## The anteater

I present machine learning as the work of ants... They focus on collecting data and "letting those data speak for themselves".

I argue against this idea of science, and outline a possible philosophical project about machine learning inspired on the philosophy of statistics.

# Plan of talk

1. Machine learning in science
2. The “new alchemy”
3. Learning from a fruit machine
4. Uncovering inductive assumptions
5. Philosophy of machine learning
6. Wrapping up

# 1 Machine learning in science

Examples of machine learning methods in the sciences, for prediction and automated model construction:

- Psychiatrists use machine learning, e.g., online targeted learning detects early-warning signals in smart phone data, and improve on models and classification.
- Linguists employ deep learning methods to replace statistical analyses of language use, e.g., statistical parsing, with the aim of automated translation.

Other such examples involve biomedical science, astronomy, and sociology.

**Historical and social dimensions**

The radically different nature of the new methods puts the continuity with existing theory under pressure.

That the new methods are “black-boxed” makes it hard to hold machine-learning research accountable and motivate policy with it.

**Transparency**

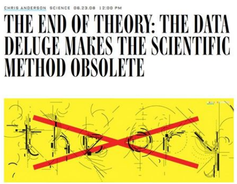
Continuity and accountability can be traced back to transparency: we need to get a handle on the implicit assumptions in machine learning.

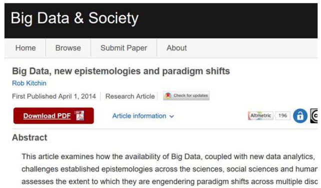
- If the assumptions implicit in the machine learning methods are uncovered, we can relate them to earlier models.
- A clear statement of the assumptions will allow us to criticize the methods and explain the results.

But the whole point of machine learning is that it is model free! Or is it?

# 2 The new alchemy

Several machine learning researchers have proclaimed the death of theory.

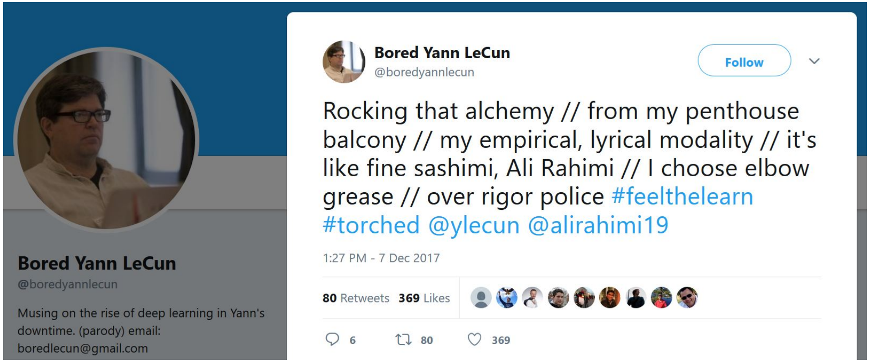



Breiman (2001) famously challenged the statistical community to take “algorithmic modeling” seriously.

**At NIPS 2017**

Rahimi sparked a fierce debate by deeming machine learning the "new alchemy" and calling for an active "rigor police".



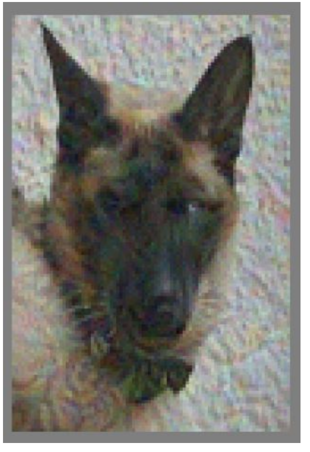

Theory-free methods invite concern over correctness and applicability.

**And then there are adversarials. . .**

What animal is this? Computer says “cat”.

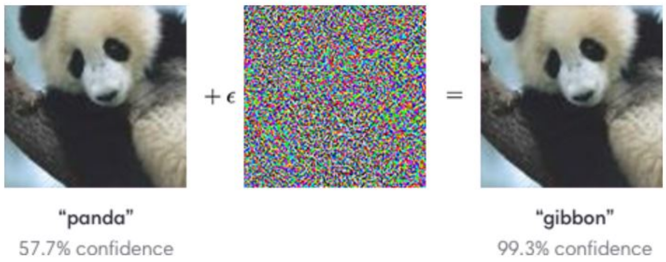
**Adding a layer of noise**

So what animal is this? Computer says “dog”.



**Adversarials**

Machine learning methods are vulnerable to highly unexpected error.




Adverserials can perhaps be counteracted. Still, to better control the reliability of the methods we have to gain insight into their assumptions.

**Wish list**

In sum, despite the attractiveness of theory-free methods, we want methods to. . .

- allow continuity in research,
- facilitate accountability,
- be understandable and communicable,
- have clear application criteria,
- avoid erratic mistakes.

For this we need clarity on the assumptions. How to reconstruct those?

# 3 Learning from a fruit machine

Inductive logic is arguably a precursor of machine learning. Consider sampling pieces of fruit $Q_i$:

Carnapian predictions are made on the basis of data alone:

$$P(Q_{n+1} = a | Q_1 \ldots Q_n) = \frac{n_a + \lambda/k}{n + \lambda},$$

where the number of possible results $k = 4$ and we might choose $\lambda = k$.

**Analogy effects**

Carnap gradually admitted more flexibility in the prediction rules. A good example is analogical prediction, e.g.,

$$P(Q_{n+1} = a | Q_1 \dots Q_n) \quad = \quad \frac{n_{\{a,c\}} + \mu/2}{n + \mu} \times \frac{n_a + \lambda/2}{n_{\{a,c\}} + \lambda}.$$

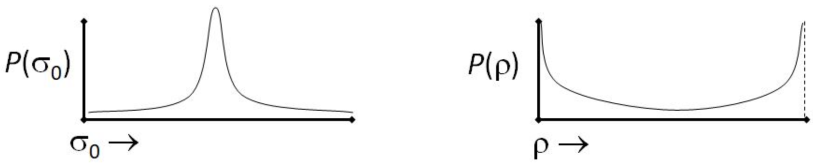
If $\mu < \lambda$, apples and bananas affect our expectation of cherries differently:

$$P(Q_{n+2} = c | Q_1 \dots Q_n \wedge Q_{n+1} = a) \quad > \quad P(Q_{n+2} = c | Q_1 \dots Q_n \wedge Q_{n+1} = b).$$

The literature offers numerous other systems for analogy effects in the predictions.

## The use of models

It is helpful to redefine analogical prediction in Bayesian terms, by a prior over multinomial distributions: $P(H_\theta)$ with $\theta \in \langle \rho, \sigma_0, \sigma_1 \rangle$.




Translating prediction rules into Bayesian models is illuminating. Can we translate machine learning methods in the same way?

**Putnam's curse**

As an aside: there may be a parallel between adversarials and so-called unlearnable sequences.

- Putnam (1963) challenged Carnap's project by constructing a sequence that, relative to a set of prediction rules, is not predictable.
- If some rule assigns a high probability to an observation, the sequence wille have some other observation as its continuation.
- The literature following Putnam's curse might shed light on the actively researched issue of adversarials in machine learning.

# 4 Uncovering inductive assumptions

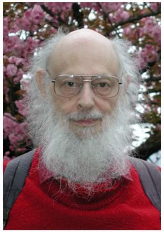
Philosophy and statistics have seen many more unsuccessful attempts to rid inductive inference from its theoretical starting points.

We can learn from these examples of data-driven science. Where did the implicit theoretical assumptions go to hide?

**Universal prediction**

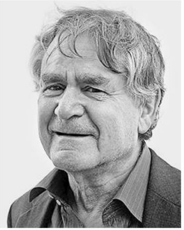
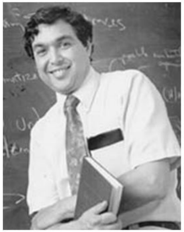
Sterkenburg (2017) offers an in-depth analysis of Solomonoff's idea of universal prediction, i.e., of considering all possible data patterns in prediction.

The predictions rest on the assumption of a a highly skewed prior over all semi-computable measures. And in the end they fall prey to Putnam's curse.

**Fiducial argument**

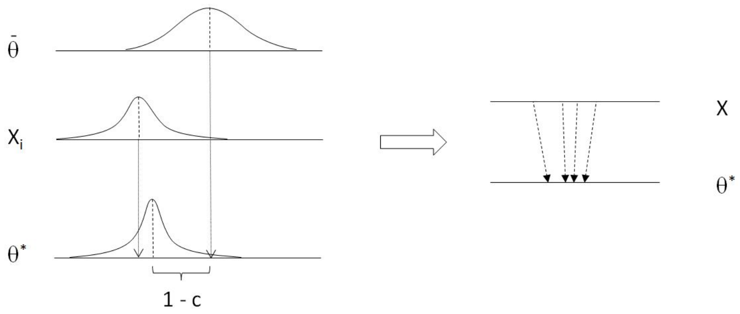
Fisher attempted to generate probabilistic conclusions about statistical hypotheses on the basis of data only.

But... his argument rests on the assumption of an improper implicit prior, projected onto the hypotheses via a functional model.

**Shrinkage estimators**

James and Stein (1957) derive that maximum likelihood estimators can be improved if we consider a collection of estimation problems.




As Efron and Morris (1977) already show, the improvement rests on an implicit empirical prior.

**Bayesian statistics**

Modeling assumptions and prior opinion are explicitly adopted. Objective priors always rest on some principle of indifference.

Through the notion of exchangeability, even De Finetti's version of Bayesian inference rests on an assumed structure in the data.

# 5 Philosophy of machine learning

The foregoing suggests how we can uncover inductive assumptions inherent in the new machine learning methods. The rough ideas are:

- Identify modeling assumptions by translating between predictive systems and Bayesian statistics.
- Detect an implicit prior by framing the methods as a probabilistic inference.
- Look for applications of the principle of indifference to the conceptual structure posited by the method.
- And consider the assumptions inherent in how the observations are framed.

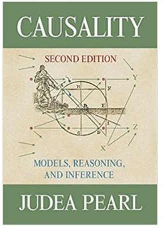
**Why again?**

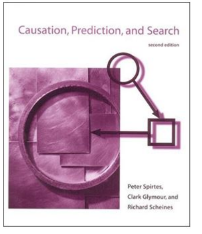
Uncovering these assumptions is an important task for the philosophy of science.

- It will help to integrate the new methods into existing and more theoretical approaches.
- Similarly it will improve on the communicability and public acceptance of machine learning results.
- And it will make it easier to hold researchers accountable and criticize their conclusions.
- Finally, it will help us guard against unreliable inferences.

**Improving the methods**

Apart from these benefits, attention for the foundations will help to improve machine learning itself.




Automated causal search is a good example. The identification of assumptions often invites the development of methods in which these assumptions are dropped.

# 6 Wrapping up

Philosophy of science can be a valuable contributor to making machine learning a success.

- Machine learning will very likely transform our sciences.
- To improve that process, our primary goal should be to identify the assumptions inherent in machine learning.
- To this aim we can take inspiration from the philosophy of induction and statistics.

# Thanks for your attention

This talk will be available at http://www.philos.rug.nl/~romeyn. For comments and questions, email j.w.romeijn@rug.nl.