The Rise of the Machines

[Based on Larry Wasserman's paper]

Maria D. Lomeli-Garcia

Gatsby Computational Neuroscience Unit

March 26, 2013

Table of Contents I

- Leo Breiman's paper "Statistical Modelling: The Two Cultures" (2001) reduces Machine Learning to a set of primitive methods, random forests and neural networks mostly.
- It's been more than 10 years between Wasserman's and Breiman's papers and ML has become a much more mature field since then.
- This paper is more of conciliatory nature between the two fields. It tells statisticians that ML is worth exploring: new research areas, new appplications, new colleagues to work with.
- It also gives the false idea that there is a unified view in ML rather than various schools of thought.

Quoting Larry I: Machine Learning and Statistics

Statistics is the science of learning from data. Machine Learning is the science of learning from data. These fields are identical in intent although they differ in their history, conventions, emphasis and culture.

Quoting Larry II

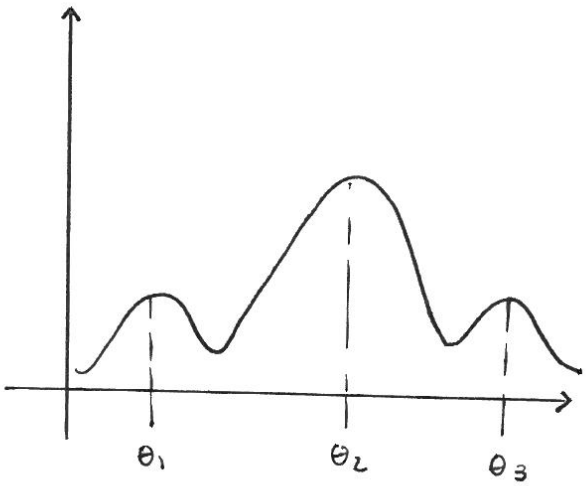
During my 25 year career I have seen Machine Learning evolve from being a collection of rather primitive (yet clever) set of methods to do classification to a sophisticated science that is rich in theory and applications.

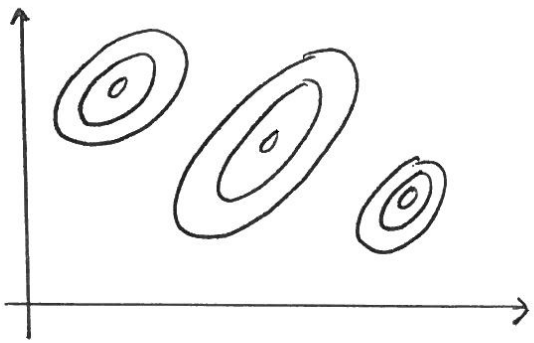
- The fast pace in which ML evolves is due to, partly, the conference culture. The main venue for research is refereed conference proceedings rather than journals.
  Having constant dealines is good.

Quoting Larry III

Consider the problem of estimating a mixture of Gaussians. In Statistics we think of this as a solved problem. You could use for example, the EM algorithm. But the EM algorithm does not solve the problem. There is no guarantee that the EM will actually find the MLE.

Mixture of Gaussians Example





All optimization methods we know about are doing a local search so we cannot guarantee they will find the global optimum.

Quoting Larry IV:What does it mean to solve a problem?

In ML, when you say you've solved the problem, you mean that there is a polynomial time algorithm with provable guarantees.

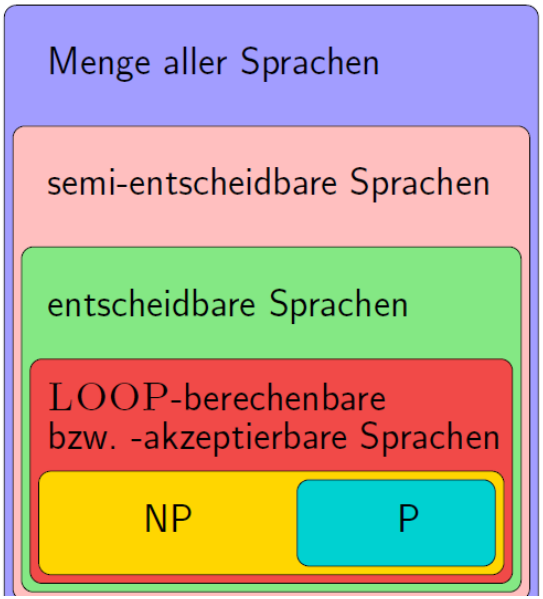
Finite sample (non-asymptotic) guarantees.

Quoting Larry V: Most controversial statement

ML puts heavier emphasis on computational thinking. Consider, for example, the difference between P and NP... Running an MCMC on an NP hard problem is often meaningless. Instead, it is usually better to approximate the NP problem with a simpler problem.

Komplexitätsklassen

Die Komplexitätsklassen P und NP können in die Chomsky-Hierarchie folgendermaßen eingeordnet werden:



Some Definitions

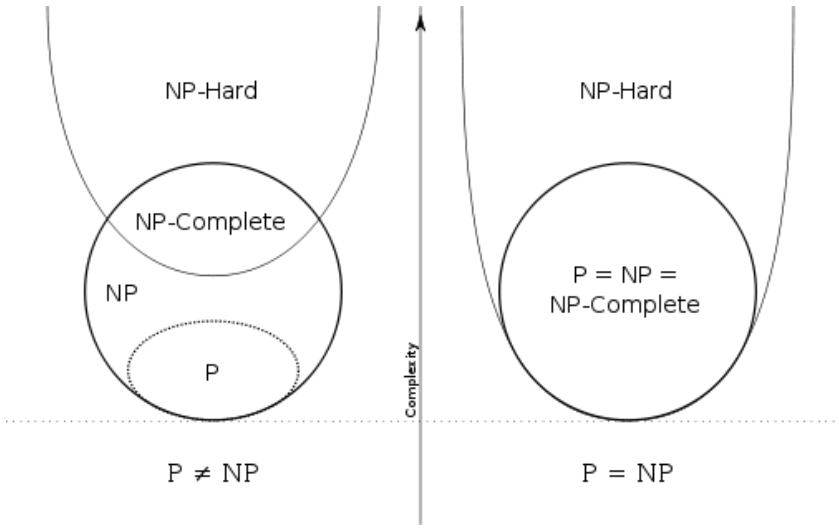
1. Problem A is NP-hard iff

$$\forall L \in NP; L \leq_p A$$

2. Problem A is NP-complete iff
   - A is NP-hard.
   - $A \in NP$.
3. Let $A \leq_p B$
   - If $B \in P$ then $A \in P$.
   - If $B \in NP$ then $A \in NP$.

Wikipedia Definitions

NP-hard (non-deterministic polynomial-time hard), in computational complexity theory, is a class of problems that are, informally, "at least as hard as the hardest problems in NP".



Examples:

- Travelling salesman problem. Optimization problem of finding the least-cost cyclic route through all nodes of a weighted graph (NP-hard)
- Subset sum problem.Given a set of integers, does any non-empty subset of them add up to zero? (NP-Complete)
- For more: Google list of NP-hard problems.

Approximate-Exact v.s. Exact-Approximate

- How about doing approximate inference on an exact model?
- Does this means we should give up on improving MCMC methodologies and research?
- Do we always need frequentist type guarantees?
- Any other interpretations are welcomed.

Thanks to Heikko, Charles and Laurence.

Larry Wasserman's blog:
http://normaldeviate.wordpress.com/