Relations Between Machine Learning Problems

Robert Williamson
On behalf of the organisers

December 2011
Why this workshop?

Machine Learning is engineering —
goal is to create technologies.
Why this workshop?

- Machine Learning is engineering — goal is to create technologies.
- *Synthetic* not *analytic*
Why this workshop?

- Machine Learning is engineering — goal is to create technologies.
- *Synthetic* not *analytic*
- But ML lacks many attributes of mature engineering disciplines
Why this workshop?

- Machine Learning is engineering — goal is to create technologies.
- *Synthetic* not *analytic*
- But ML lacks many attributes of mature engineering disciplines
- Techniques, not problems
- Lots of reinvention
- Problems solved from scratch
- No standards
A function of a variable quantity is an analytic expression composed in any way whatsoever of the variable quantity and numbers or constant quantities. — Euler
A function of a variable quantity is an analytic expression composed in any way whatsoever of the variable quantity and numbers or constant quantities. — Euler

Mathematical Functions prior to the 20th Century

- Each studied individually — a bunch of tricks; no unifying theory
- No separation between the function and its formula
- Each new problem solved from scratch
Functional Analysis

- Functions as basic objects; not just formulas
- Development of *functionals* — mappings between functions
- Abstraction leading to simplicity, unification, and the ability to understand the whole
(One possible) Goal

Create the analog of functional analysis for machine learning.
Create the analog of functional analysis for machine learning.

- Machine learning **problems** as points
- Study transformations between points
- Build a map of machine learning problems
Of course not everything is a relation . . . things have properties and attributes that it is helpful to understand
Of course not everything is a relation . . . things have properties and attributes that it is helpful to understand.

Relations can provide a “multiplicative” impact — connect two things and you have multiplied knowledge — see Jenn’s talk.
Why “relations” rather than “attributes”?

- Of course not everything is a relation . . . things have properties and attributes that it is helpful to understand.
- Relations can provide a “multiplicative” impact — connect two things and you have multiplied knowledge — see Jenn’s talk.
- And sometimes in order to understand a single thing it is helpful to study how it relates to other things.
Grothendieck’s Relative Point of View

“... is a heuristic applied in certain abstract mathematical situations, with a rough meaning of taking for consideration families of 'objects' explicitly depending on parameters, as the basic field of study, rather than a single such object.”
Attempting to relate *algorithms* is doomed...

- There is no satisfactory definition of an algorithm.
- One cannot say whether two algorithms are the same.
- Could still be useful, but problems better.
- E.g. kernels as building blocks.
John R. Platt:

“Beware of the man of one method or one instrument, either experimental or theoretical.”

[Strong Inference, 1962]
Problem oriented versus method oriented

John R. Platt:

“Beware of the man of one method or one instrument, either experimental or theoretical. He tends to become method oriented rather than problem oriented."
Problem oriented versus method oriented

John R. Platt:

“Beware of the man of one method or one instrument, either experimental or theoretical. He tends to become method oriented rather than problem oriented. The method-oriented man is shackled: the problem-oriented man is at least reaching freely toward what is most important.”

[Strong Inference, 1962]
Consider the following (incomplete) list of words used to describe ML problems. For each word there are several papers in the literature already presenting methods.
Batch, online, transductive, off-training set, semi-supervised, noisy (label, attribute, constant noise / variable noise, data of variable quality, data of different costs, weighted loss functions, active, distributed, compressed sensing, classification (binary, weighted binary, multi-class, structured output), probabilistic concepts / scoring rules / class probability estimation, learning with statistical queries, Neyman-Pearson classification, ordinal regression, ranked regression, ranking, ranking the best, optimising the ROC curve, optimising the AUC, regression, selection, novelty detection, multi-instance learning, minimum volume sets, density level sets, regression level sets, sets of quantiles, quantile regression, density estimation, data segmentation, clustering, co-training, co-validation, learning with constraints, conditional estimators, estimated loss, confidence / hedging estimators, hypothesis testing, distributional distance estimation, learning relations, learning total orders, learning causal relationships, estimating performance (cross validation)
Studying relations between ML problems is not new...

Everything of importance has been said before by somebody who did not discover it
— Alfred North Whitehead
Lots of Existing Relations


Relations are not just reductions.
But reductions are great relations — theoretical and practical
Confer probing reduction and integral representations
Types of Relations

Unary representations  Take a particular object and look at different representations of it. Or characterise certain properties in terms of representations (convexity of proper losses in terms of “weight functions”).

Binary Relations between objects  $f$-divergences and Bayes risks for binary experiments with proper losses.

Binary relations between representations of objects  Integral representations of $f$-divergences and losses

Primitives  Elementary divergences and losses

Domination Relationships  Comparison of experiments. A variant: when is one loss better than another. Or what is the “best” convex surrogate loss?
Monocausotaxophilia considered harmful

Taxophilia — the urge to find memorable pattern and harmony in the environment.

We are not advocating single explanations for everything. "The real essence of learning is X" does not help. No "unified frameworks" other than the idea of a relation. Analogy: Multicultural societies. Confer the "Common economic protocol".
Taxophilia — the urge to find memorable pattern and harmony in the environment

- We are not advocating single explanations for everything
- “The real essence of learning is X” does not help
- No “unified frameworks” other than the idea of a relation
- Analogy: Multicultural societies. Confer the “Common economic protocol”
Standards and Modularisation

- Mature disciplines have standards and modularisation
- It helps tremendously (scientifically, technologically and economically)
  - Atomic theory of matter
  - GSM
  - Grid electricity
  - Bolts
  - html/http/pdf etc
  - NP-completeness catalogue
  - Linnean taxonomy
  - International Classification of Diseases
- To build big systems control the interfaces
- Not boring — they free the designer for more creative tasks. Enable pervasiveness.
- Standards define what becomes *infrastructure*
Not just a theoretical exercise!

- How is ML software written?
Not just a theoretical exercise!

- How is ML software written?
- How is ML software used?
Not just a theoretical exercise!

- How is ML software written?
- How is ML software used?
- Developers versus users
Not just a theoretical exercise!

- How is ML software written?
- How is ML software used?
- Developers versus users
- How to make ML a service
Not just a theoretical exercise!

- How is ML software written?
- How is ML software used?
- Developers versus users
- How to make ML a service
- Prediction markets, ML markets, Data markets
Not just a theoretical exercise!

- How is ML software written?
- How is ML software used?
- Developers versus users
- How to make ML a service
- Prediction markets, ML markets, Data markets
- Jenn’s talk, Amos’ talk, Mark’s talk
The Workshop itself

- Understand some new relations
- Identify researchers interested in this line of work
- Identify open problems that should be attacked next
- Discussion on how to discover new relations, how to organise and how to communicate them
Desired Outcomes

Community building

**Map of ML**  Is it possible, is it desirable, how should it be done, and what should be done with it?
Desired Outcomes

Community building

Map of ML  Is it possible, is it desirable, how should it be done, and what should be done with it?

Communication  What mechanisms do we need as a community?
Desired Outcomes

Community building

**Map of ML**  Is it possible, is it desirable, how should it be done, and what should be done with it?

**Communication**  What mechanisms do we need as a community?

**Standards**  Problems and Software. Do we need them or want them? How to construct?
Desired Outcomes

Community building

Map of ML  Is it possible, is it desirable, how should it be done, and what should be done with it?

Communication  What mechanisms do we need as a community?

Standards  Problems and Software. Do we need them or want them? How to construct?

Network Effects  How to scale up. The bigger the catalogue, the more helpful it is
Desired Outcomes

Community building

Map of ML  Is it possible, is it desirable, how should it be done, and what should be done with it?

Communication  What mechanisms do we need as a community?

Standards  Problems and Software. Do we need them or want them? How to construct?

Network Effects  How to scale up. The bigger the catalogue, the more helpful it is

Culture  is it conceivable to shift the ML research culture? First try and relate you problem at hand to a known problem; be creatively lazy rather than creatively busy.