# Relations Betweeen Machine Learning Problems

Robert Williamson
On behalf of the organisers

December 2011

THE AUSTRALIAN NATIONAL UNIVERSITY

NICTA

# Why this workshop?

- Machine Learning is engineering —
  goal is to create technologies.

# Why this workshop?

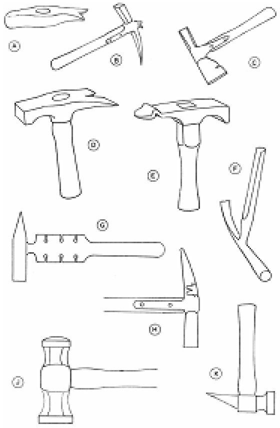
- Machine Learning is engineering —
  goal is to create technologies.
- *Synthetic* not *analytic*

# 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



# Analogy

*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

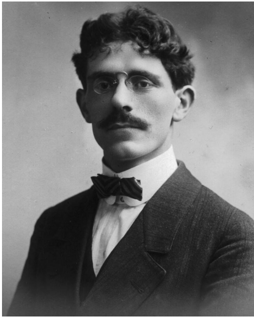
# Analogy

*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

# The 20th Century View...



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.

# (One possible) Goal

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

# Why “relations” rather than “attributes”?

- Of course not everything is a relation . . . things have properties and attributes that it is helpful to understand

# 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

# 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

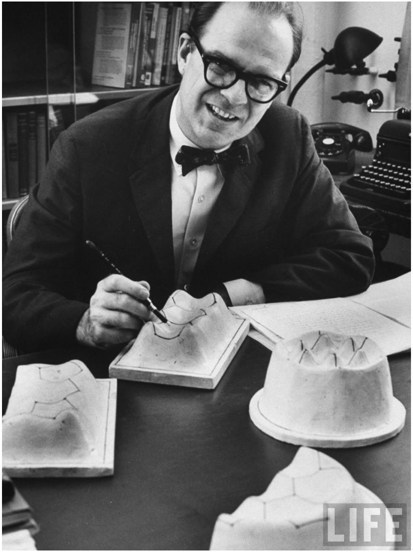
"…is a heuristic applied in certain abstract mathematical situations, with a rough meaning of taking for consideration families of 'objects' explicitly depending on parameters, as the basic field of study, rather than a single such object."

# Why “problems” and not “algorithms”

Attempting to relate *algorithms* is doomed. . .

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

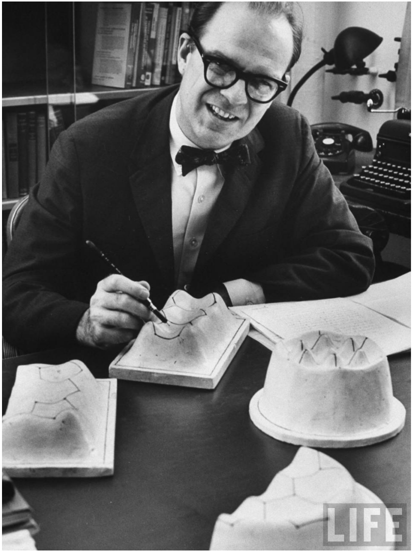
# Problem oriented versus method oriented




**John R. Platt:**

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

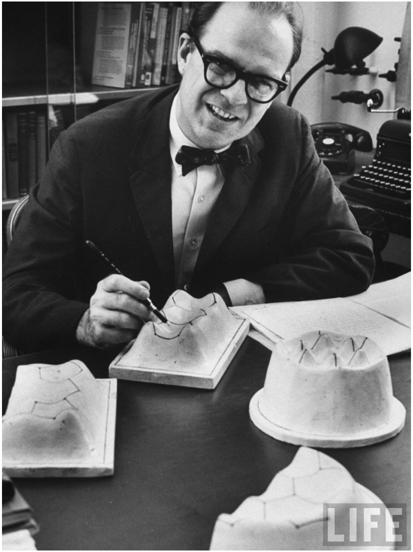
# Problem oriented versus method oriented




**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]

# A flood of problems

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

2011 Mark D. Reid and Robert C. Williamson, "Information Divergence and Risk in Binary Experiments," *Journal of Machine Learning Research*, 2011.

2006 Peter L. Bartlett, Michael I. Jordan and Jon D. McAuliffe, "Convexity Classification and Risk Bounds," *Journal of the American Statistical Association*, 101(473),138-156, March 2006.

2006 J. Langford, R. Oliveira, and B. Zadrozny. "Predicting Conditional Quantiles via Reduction to Classification." In *Proceedings of the 22nd Conference in Uncertainty in Artificial Intelligence*. AUAI Press, 2006

2006 Ingo Steinwart. "How to compare different loss functions and their risks." Preprint, Modeling, Algorithms and Informatics Group, CCS-3, Los Alamos National Laboratory, September 2006.

2006 F. Liese and I. Vajda. “On divergences and informations in statistics and information theory.” *IEEE Transactions on Information Theory,* 52(10):43944412, 2006.

2005 A. Beygelzimer, V. Dani, T. Hayes, J. Langford, and B. Zadrozny. “Error limiting reductions between classification tasks.” In *Proceedings of the 22nd International Conference on Machine Learning*, Bonn, 2005.

2005 John Langford and Bianca Zadrozny. “Estimating class membership probabilities using classifier learners.” In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTAT05)*, 2005.

2002 Alison L. Gibbs and Francis Edward Su. “On choosing and bounding probability metrics.” *International Statistical Review,* 70:419435, 2002.

1998 M. Kearns. “Efficient noise-tolerant learning from statistical queries.” *Journal of the ACM*, 45(6):9831006, November 1998.

1996 Peter L. Bartlett, Philip M. Long, and Robert C. Williamson. Fat-shattering and the learnability of real-valued functions. *J. Comput. Syst. Sci.*, 52(3):434452, 1996.

1995 N. Klasner and H.U. Simon. “From noise-free to noise-tolerant and from on-line to batch learning.” In *Proceedings of the 8th Annual Conference on Computational Learning Theory*, pages 250257. ACM Press New York, NY, USA, 1995.

1989 S.A. Goldman, R.L. Rivest, and R.E. Schapire. “Learning binary relations and total orders,” In *Foundations of Computer Science, 1989., 30th Annual Symposium on* 4651, 1989.

1982 Nicolai Nikolaevich Cencov. *Statistical Decision Rules and Optimal Inference,* volume 53 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, Rhode Island, 1982.

1981 W.J. Conover and R.L. Iman, “Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics”, *The American Statistician*, 35(3), 124-129, 1981

1966 N. Morse and R. Sacksteder. “Statistical Isomorphism.” *The Annals of Mathematical Statistics*, 37(1):203 214, 1966.

1964 L. LeCam. “Sufficiency and Approximate Sufficiency.” *The Annals of Mathematical Statistics*, 35(4):1419 1455, 1964.

1961 Allan Birnbaum, “On the foundations of statistics: Binary Experiments,” *The Annals of Mathematical Statisitics*, 32(2), 414-435, June 1961.

1956 D.V. Lindley. “On a Measure of the Information Provided by an Experiment.” *The Annals of Mathematical Statistics*, 27(4):9861005, 1956

1953 D. Blackwell. “Equivalent Comparisons of Experiments.” *The Annals of Mathematical Statistics*, 24(2):265272, 1953

1951 D. Blackwell. “Comparison of Experiments.” In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1950, pages 93102, Berkeley and Los Angeles, 31 July 12 August 1951. University of California Press.

# Not just “reductions ”

- 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

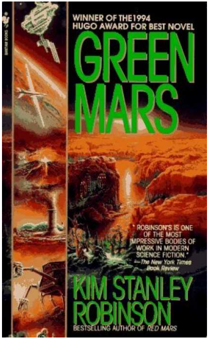
**Domination Relationships** Comparison of experiments. A variant: when is one loss better than another. Or what is the "best" convex surrogate loss?

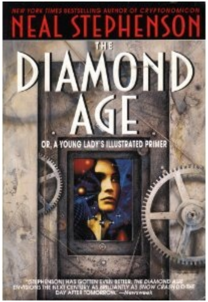
# Monocausotaxophilia considered harmful

# 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”



# 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

# Desired Outcomes

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 conceiveable 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.