

DECEMBER 16, 2011, 07:30–10:30 AND 16:00–19:00

XXX WS8

## Relations between machine learning problems – an approach to unify the field

<http://rml.cecs.anu.edu.au/>

**Bob Williamson**

THE AUSTRALIAN NATIONAL UNIVERSITY AND NICTA

**John Langford**

YAHOO! RESEARCH

**Ulrike von Luxburg**

UNIVERSITY OF HAMBURG

**Mark Reid**

THE AUSTRALIAN NATIONAL UNIVERSITY AND NICTA

**Jennifer Wortman Vaughan**

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Bob.Williamson@anu.edu.au

jl@yahoo-inc.com

ulrike.luxburg@tuebingen.mpg.de

mark.reid@anu.edu.au

jenn@cs.ucla.edu

### Abstract

The workshop proposes to focus on relations between machine learning problems. We use “relation” quite generally to include (but not limit ourselves to) notions such as: one type of problem being viewed special case of another type (e.g., classification as thresholded probability estimation); reductions between learning problems (e.g., transforming ranking problems into classification problems); the use of surrogate losses (e.g., replacing misclassification loss with some other, convex loss); relations between sets of learning problems, such as those studied in the (old) theory of “comparison of experiments”; connections between machine learning problems and what could be construed as “economic learning problems” such as prediction markets and forecast elicitation.

The point of studying relations between machine learning problems is that it stands a reasonable chance of being a way to be able to understand the field of machine learning as a whole. It could serve to prevent re-invention, and rapidly facilitate the growth of new methods.

<b>7.30-8.00</b>	<b>Introductory Overview</b> WORKSHOP ORGANISERS	
<b>8.00-8.55</b>	<b>Machine Learning Markets: Putting your money where your mouth is</b> AMOS STORKEY	
<b>8.55-9.05</b>	Coffee break	
<b>9.05-10.00</b>	<b>Efficient Market Making via Convex Optimization, and a Connection to Online Learning</b> JENN WORTMAN VAUGHAN	
<del><b>10.00-10.30</b></del>	<del><b>Bounded Regret Sequential Learning using Prediction Markets</b></del> <del>SINDHU KUTTY AND RAHUL SAMI</del>	Author unable to attend.
<b>16.00-16.55</b>	<b>We need a BIT more GUTS (=Grand Unified Theory of Statistics)</b> PETER GRÜNWARD	
<b>16.55-17.25</b>	<b>Anatomy of a Learning Problem</b> MARK REID	

17.25-17.40	Coffee break
17.40-18.10	<b>Degrees of Supervision</b> DARÍO GARCÍA-GARCÍA
18.10-19.00	Panel Discussion

### **Machine Learning Markets**

**Amos Storkey**, UNIVERSITY OF EDINBURGH

Prediction markets show considerable promise for developing flexible mechanisms for machine learning. Here, machine learning markets for multivariate systems are defined, and a utility-based framework is established for their analysis. This differs from the usual approach of defining static betting functions. It is shown that such markets can implement model combination methods used in machine learning, such as product of expert and mixture of expert approaches as equilibrium pricing models, by varying agent utility functions. They can also implement models composed of local potentials, and message passing methods. Prediction markets also allow for more flexible combinations, by combining multiple different utility functions. Conversely, the market mechanisms implement inference in the relevant probabilistic models. This means that market mechanism can be utilized for implementing parallelized model building and inference for probabilistic modelling.

### **Efficient Market Making via Convex Optimization, and a Connection to Online Learning**

**Jenn Wortman Vaughan**, UNIVERSITY OF CALIFORNIA, LOS ANGELES

A prediction market is a financial market designed to aggregate information. To facilitate trades, prediction markets are often operated by automated market makers. The market maker trades a set of securities with payoffs that depend on the outcome of a future event. For example, the market maker might offer a security that will pay off \$1 if and only if a Democrat wins the 2012 US presidential election. A risk neutral trader who believes that the probability of a Democrat winning is  $p$  should be willing to purchase this security at any price below  $p$ , or sell it at any price above  $p$ . The current market price can then be viewed as the traders' collective estimate of how likely it is that a Democrat will win the election.

Market-based estimates have proved to be accurate in a variety of domains, including business, entertainment, and politics. However, when the number of outcomes is very large, it is generally infeasible to run a simple prediction market over the full outcome space. We propose a general framework for the design of securities markets over combinatorial or infinite state or outcome spaces. Our framework enables the design of computationally efficient markets tailored to an arbitrary, yet relatively small, space of securities with bounded payoff. We prove that any market satisfying a set of intuitive conditions must price securities via a convex cost function, which is constructed via conjugate duality. Rather than deal with an exponentially large or infinite outcome space directly, our framework only requires optimization over a convex hull. By reducing the problem of automated market making to convex optimization, where many efficient algorithms exist, we arrive at a range of new polynomial-time pricing mechanisms for various problems.

Our framework also provides new insights into the relationship between market design and machine learning. In particular, we show that the tools that have been developed for online linear optimization are strikingly similar to those we have constructed for selecting pricing mechanisms. This is rather surprising, as the problem of learning in an online environment is semantically quite distinct from the problem of pricing securities in a prediction market: a learning algorithm receives losses and selects weights whereas a market maker manages trades and sets prices. We show that although the two frameworks have very different semantics, they have nearly identical syntax in a very strong sense.

### **Bounded-Regret Sequential Learning using Prediction Markets**

**Sindhu Kuty**, UNIVERSITY OF MICHIGAN, ANN ARBOR

**Rahul Sami**, UNIVERSITY OF MICHIGAN, ANN ARBOR

We demonstrate a relationship between prediction markets and online learning algorithms by using a prediction market metaphor to develop a new class of algorithms for learning exponential families with expert

advice. The specific problem we consider is that of prediction when data is distributed according to a particular member of an exponential family. In such a case, cost function based prediction markets provide a convenient analytical tool for evaluating performance. Prediction markets also provide a natural technique for learning in an environment where expert advice is not available simultaneously but sequentially; and experts are either honest and informative, or dishonest and adversarial. As in traditional models, we combine advice using weights on experts. However, to exploit these particular features, we use a form of Kelly gambling to relate the weight of an expert to her budget. We give a formal description of this new model along with relevant definitions and show an equivalence between learning maximum likelihood estimates of the natural parameters of an exponential family and combining advice in prediction markets. We provide an abstract architecture for learning in this model that uses this equivalence to simulate a prediction market to update budgets of experts based on their individual loss. We apply this technique to construct a concrete algorithm that achieves bounded-regret.

### **We need a BIT more GUTS (=Grand Unified Theory of Statistics)**

**Peter Grünwald**, CENTRUM VOOR WISKUNDE EN INFORMATICA

A remarkable variety of problems in machine learning and statistics can be recast as data compression under constraints: (1) sequential *prediction* with arbitrary loss functions can be transferred to equivalent log loss (data compression) problems. The worst-case optimal regret for the original loss is determined by Vovk's *mixability*, which in fact measures how many bits we lose if we are not allowed to use mixture codes in the compression formulation. (2) in *classification*, we can map each set of candidate classifiers  $C$  to a corresponding probability model  $M$ . *Tsybakov's condition* (which determines the optimal convergence rate) turns out to measure how much more we can compress data by coding it using the convex hull of  $M$  rather than just  $M$ . (3) *hypothesis testing* in the applied sciences is usually based on p-values, a brittle and much-criticized approach. Berger and Vovk independently proposed *calibrated p-values*, which are much more robust. Again we show these have a data compression interpretation. (4) *Bayesian nonparametric approaches* usually work well, but fail dramatically in Diaconis and Freedman's pathological cases. We show that in these cases (and only in these) the Bayesian predictive distribution does not compress the data. We speculate that all this points towards a general theory that goes beyond standard MDL and Bayes.

### **Anatomy of a Learning Problem**

**Mark Reid**, THE AUSTRALIAN NATIONAL UNIVERSITY AND NICTA

**James Montgomery**, THE AUSTRALIAN NATIONAL UNIVERSITY

**Mindika Premachandra**, THE AUSTRALIAN NATIONAL UNIVERSITY

In order to relate machine learning problems we argue that we need to be able to articulate what is meant by a single machine learning problem. By attempting to name the various aspects of a learning problem we hope to clarify ways in which learning problems might be related to each other. We tentatively put forward a proposal for an anatomy of learning problems that will serve as scaffolding for posing questions about relations. After surveying the way learning problems are discussed in a range of repositories and services. We then argue that the terms used to describe problems to better understand a range of viewpoints within machine learning ranging from the theoretical to the practical.

### **Degrees of Supervision**

**Darío García-García**, THE AUSTRALIAN NATIONAL UNIVERSITY

**Robert C. Williamson**, THE AUSTRALIAN NATIONAL UNIVERSITY AND NICTA

Many machine learning problems can be interpreted as differing just in the level of supervision provided to the learning process. In this work we provide a unifying way of dealing with these different degrees of supervision. We show how the framework developed to accommodate this vision can deal with the continuum between classification and clustering, while also naturally accommodating less standard settings such as learning from label proportions, multiple instance learning,...All this emanates from a simple common principle: when in doubt, assume the simplest possible classification problem on the data.