Journal of Economic Behavior & Organization 133 (2017) 373–384

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity[☆]

Alexander Peysakhovich[a,*,1], Jeffrey Naecker[b]

[a] Yale University, United States
[b] Wesleyan University, United States

ARTICLE INFO

Article history:
Received 9 October 2015
Received in revised form 2 June 2016
Accepted 30 August 2016
Available online 26 October 2016

Keywords:
Behavioral economics
Machine learning
Risk
Ambiguity
Decision-making

ABSTRACT

How can behavioral science incorporate tools from machine learning (ML)? We propose that ML models can be used as upper bounds for the "explainable" variance in a given data set and thus serve as upper bounds for the potential power of a theory. We demonstrate this method in the domain of uncertainty. We ask over 600 individuals to make a total of 6000 choices with randomized parameters and compare standard economic models to ML models. In the domain of risk, a version of expected utility that allows for non-linear probability weighting (as in cumulative prospect theory) and individual-level parameters performs as well out-of-sample as ML techniques. By contrast, in the domain of ambiguity, two of the most widely studied models (a linear version of maximin preferences and second order expected utility) fail to compete with the ML methods. We open the "black boxes" of the ML methods and show that under risk we "rediscover" expected utility with probability weighting. However, in the case of ambiguity the form of ambiguity aversion implied by our ML models suggests that there is gain from theoretical work on a portable model of ambiguity aversion. Our results highlight ways in which behavioral scientists can incorporate ML techniques in their daily practice to gain genuinely new insights.

Decisions ranging from the mundane (e.g. choosing a restaurant) to the life-changing (e.g. choosing a job) include elements of uncertainty. For this reason, understanding how individuals evaluate uncertain prospects has been a key research area in the behavioral and social sciences for over two centuries (Bernoulli, 1738; Kreps, 1988). This has led to the creation of simple mathematical models that are characterized by parameters with intuitively understandable interpretations (e.g. the coefficient of risk aversion). There are many important recurring questions in this research program: How good are these models? What commonly used assumptions are the most restrictive? What domains of uncertainty appear to be potentially fruitful targets for theorists?

In this paper we complement traditional behavioral science techniques with machine learning (ML). We focus on two domains: risk (Camerer, 1995; Kahneman and Tversky, 2000; Kreps, 1988; Savage, 1972), where the probability of an uncertain outcome is perfectly known, and ambiguity (Knight, 1921; Ellsberg, 1961; Camerer and Weber, 1992; Trautmann and

☆ We thank Dean Eckles, Ido Erev, Drew Fudenberg, Alex Imas, Jon Levin, Muriel Niederle, David Rand, Alvin Roth, Andrei Schleifer, Sean Taylor, Erez Yoeli and participants at the IC2S2 Conference as well as the Stanford Experimental Economics course for valuable comments and suggestions. Errors remain our own. Peysakhovich thanks the John Templeton Foundation for financial support.

* Corresponding author.
E-mail addresses: alex.peys@gmail.com (A. Peysakhovich), jnaecker@wesleyan.edu (J. Naecker).

1 Current author affiliation: Facebook Inc., United States. These studies were performed and most of the analysis was done while the author was a Research Scientist at the Human Cooperation Lab at Yale University.

http://dx.doi.org/10.1016/j.jebo.2016.08.017
0167-2681/

Van De Kuilen, 2013), where decision-makers have partial, but not full, information to estimate the likelihood of an outcome. We recruit over 600 participants to indicate their willingness to pay for uncertain prospects whose features are randomly generated. As is common in the statistical learning literature (Friedman et al., 2009), we take a subset of these decisions as an out-of-sample "test set." We calibrate several economic models: expected utility (EU, von Neumann and Morgenstern, 1945) and expected utility with non-linear probability weighting (EUP, Tversky and Kahneman, 1992; Prelec, 1998) in the case of risk and second-order expected utility (SOEU, Grant et al., 2009) and maximin preferences (MM, Gilboa and Schmeidler, 1989; Levy et al., 2010; Tymula et al., 2012) in the case of ambiguity on the remaining decisions. We then ask: how well do these models predict the held out test set decisions?

This exercise allows us to tackle two issues. First, it allows us to consider the relative explanatory power of the economic models. Note that because EUP nests EU but has an additional parameter, it will always fit (weakly) better in-sample. However, this may simply be over-fitting and the more complicated model may actually do worse out-of-sample. Thus comparing the models' out-of-sample fit allows us to ask whether the additional model complexity adds value in terms of playing an important part in predicting variation in behavior in problems the model has not yet encountered.

Of course, a statement that a model explains X% of the variance in a particular domain begs the question: is that good or bad? A model that predicts 10% of the variance in a very clean data set might be considered to have quite poor explanatory power. However, if there is substantial noise (either due to sampling error, poor data construction, or other factors), explaining 10% of the variance may actually be quite good.

Thus, we are interested in (out-of-sample) explained variance as a proportion of *explainable variance*. To estimate explainable variance, we turn ML. These tools are designed specifically for prediction and so we use their accuracy on the test set as an estimate of explainable variance in our experiments.[3] As our ML benchmark model we use a cross-validated regularized regression. To allow linear regression to fit non-linear functions we take a basis expansion of all potential decision-relevant variables (probabilities and prizes for each outcome) as well as their interactions. In our most powerful model we also include interactions of each decision-relevant variable with subject-level dummies. This gives us 55,000+ parameters to estimate, so to prevent overfitting we cross-validate and regularize the model (i.e. penalize the model for complexity).

We find that the regularized regression outperforms expected utility models by a large margin under a representative agent assumption. We also find that attempting to fit representative agent models without allowing for individual-level heterogeneity makes the predictive power of any model quite poor. However, when individual level parameters are allowed EUP does as well as the machine learning algorithms. We interpret this as a victory for probability weighting: this parameter increases out of sample prediction considerably, so it is an important feature of models of uncertain choice. We also consider this a victory for the economic models: a ~600 parameter model (2 per person x ~300 subjects) that is interpretable (i.e. the coefficient of risk aversion has an economic meaning outside of the model) is able to predict choices as well as the ML algorithm which has two orders of magnitude more parameters (~55,000) and is optimized purely for prediction and not interpretability. Additionally, we show that the implied probability weighting curve generated by the best ML model is remarkably similar to the famous S-shaped weighting curve of the EUP model.

On the other hand, in the domain of ambiguity we find that neither second order expected utility nor maximin preferences are able to predict individual out-of-sample choices as well as the ML models. In an attempt to diagnose this failure, we show that the implied ambiguity penalty is convex in the amount of ambiguity, a feature that is not predicted by either model of ambiguity we consider in this paper. We interpret this as an opportunity for empirically-minded theorists: these results, combined with the success of the EUP in the domain of risk, suggest there is ample room for the development of a simple model for the domain of ambiguity that predicts well and yet is relatively parsimonious.

1. Choice under risk

1.1. Experimental design

Our first experiment focuses on the domain of risk. Participants were recruited from Amazon Mechanical Turk and were compensated for their time with rates standard in the literature. All research was approved by the Institutional Review Board of Harvard University.

All decisions made were hypothetical but participants were instructed to treat each decision as if it were real. While online experiments are much less controlled, faster and have smaller stakes than traditional lab sessions there is substantial evidence that standard behavioral economic effects replicate on Mechanical Turk (Peysakhovich and Rand, 2014; Imas, 2014; Fudenberg and Peysakhovich, 2014; Naecker, 2015), the pool is more representative (Paolacci and Chandler, 2014) and that the size of stakes (even the use of pure hypotheticals) matters little (Amir and Rand, 2012; Peysakhovich and Karmarkar, 2015; Horton et al., 2011). There are known issues with Mechanical Turk samples: for example, participants are well experienced with experimental paradigms, much more so than student populations (Rand et al., 2014). Though we acknowledge this

[3] While relatively established in computer science and industry, data mining and machine learning approaches are only recently beginning to appear in social science. This is happening in experimental work (Fudenberg and Peysakhovich, 2014; Naecker, 2015), finance (Moritz and Zimmerman, 2014), time series analysis (Varian, 2014), heterogenous treatment effect estimation (Athey and Imbens, 2015; Wager and Athey, 2015; Peysakhovich and Lada, 2016) and political science (Grimmer, 2015).