

ML+ECO

key

- dimensionality reduction
- prediction
- discard model assumption

policy

- policy decision making/design (prediction)
 - health
 - law
 - finance
 - marketing
 - new data
 - poverty
 - safety
 - home value
 - economics indicators/trend
- optimal policy selection
 - bandit
- causal inference
 - unit selection (heterogeneous effect estimation)
 - policy evaluation(inference)

general usage

- Average Treatment Effects
 - targeted maximum likelihood
 - regularized regression(double lasso)
 - "double machine learning," (partial oad)
 - ipw, "residual balancing,"(adappt ipw)
- Heterogeneous Treatment Effects and Optimal Policies
 - causal forest
 - Targeted maximum likelihood, LASSO, metalearner
 - The goal is to select a policy function to minimize the loss from failing to use the (unfeasible) ideal policy, referred to as the "regret" of the policy.
- counterfactual
 - Factor Models and Matrix Completion
 - synthetic control
- Robustness and Supplementary Analysis (adapt different estimators)
- Factor Models and Structural Models
 - often Markov Chain Monte Carlo. Recently, the ML literature has developed a variety of techniques that allow similar types of Bayesian models to be estimated at larger scale.
 - tentat analysis and consumer choices of, for example, movies at Netflix.

causal inference tool box

- unconfoundedness(observed)
 - local methods, RD
 - non-parametric weighting
 - bandwidth selection
 - matching
 - generate propensity score (indirect)
 - neural network/ tree/other classifiers (direct)
 - panel data
 - synthetic control
 - generalized DID
 - iv lasso
- random assignment (experiment)
 - bandit

pros and cons

- advantage
 - model/variable selection
 - reduce omit variable bias,
 - increase credibility of model assumption
 - adapt high-dimension environment
 - deal with new data
- disadvantage
 - biased estimator
 - model uncertainty
 - interpretability

experiment

- experimental design
 - predict outcomes
- protocol
 - no need (data driven)
- assignment
 - bandit
- banlancing
 - whether covariates can predict assignment
- average treatment effect
- heterogeneous effect
- robustness check

behavioral

- combination (Colin F. Camerer)
 - Machine Learning to Find Behavioral Variables
 - Most propositions in behavioral economics add some variables to the list of features, such as reference dependence, context-dependence (menu eff. acts), anchoring, limited attention, social preference, and so forth.
 - include all of them in an ML approach
 - Human Prediction as Imperfect Machine Learning(机器学习模型一下接收人类决策模型(过度自信效应和过度反应, 过度使用变量等))
 - judgment and decisionmaking (JDM)
 - Sparsity Is Good for Y but Taste Is Bad
 - Hypothesis: Human Judgment Is Like Overfitted Machine Learning
 - AI Technology as a Blotchy Patch, or Malware, for Human Limits
 - comment (Daniel Kahneman)
 - the main characteristic of people is that they are very noisy
 - stochastic choice theory
 - human performance is not bias, it is just noise
 - You should replace humans by algorithms whenever possible. Even when the algorithm does not do very well, humans do so poorly and are so noisy that, just by removing the noise, you can do better than people.
 - how the robot would be diff. event from the people
 - the robot will be much better at statistical reasoning and less enamored with stories and narratives than people are
 - have a much higher emotional intelligence
 - the robot would be wiser
 - test theories
 - explore behavior pattern (each exp. as an obj)
- not much unsupervised/reinforcement learning

Summary

If behavioral economics is recast as open-mindedness about what variables might predict, then ML is an ideal way to do behavioral economics because it can make use of a wide set of variables and select which ones predict.

First, simple theories can be seen as bets that only a small number of features will predict well; that is, some off-acts (such as priors) are hypothesized to be 6, not order in magnitude.

Second, if longer lists of features predict better than a short list of theory-specified features, then what's going on is that a plausible upper bound on how much potential predictability is left to understand.

The results are also likely to create raw material for theory to figure out how to consolidate the additional predictive power into crystallized theory.