# The Impact of Machine Learning on Economics

Susan Athey, Stanford University

# Machine Learning

## Supervised ML

- Outcomes Y
- Features X
- Independent obs.
- Goal: Use X to predict Y on an independent test set

$$\hat{\mu}(x) = E[Y|X = x]$$

## Unsupervised ML

- Features X
- Goals:
  - Clustering
  - Dimensionality Reduction
- "I discovered cats!"

# I discovered Town and Country!

| cluster | location | city |
|---|---|---|
| 17 | palo alto high school | palo alto |
| 17 | trader joe's (el camino real) | palo alto |
| 17 | calafia cafe | palo alto |
| 17 | mayfield bakery & cafe | palo alto |
| 17 | gotts roadside | palo alto |
| 17 | whole foods market (emerson street) | palo alto |
| 17 | philz coffee (forest ave.) | palo alto |
| 17 | bloomingdale's | palo alto |
| 17 | apple retail store (university avenue) | palo alto |
| 17 | foothills tennis & swimming club | palo alto |
| 17 | lytton gardens community housing | palo alto |
| 17 | nest labs | palo alto |
| 17 | tin pot creamery (el camino real) | palo alto |
| 17 | palo alto clay and glass festival | palo alto |
| 17 | blue bottle coffee | palo alto |
| 17 | kirks steakburgers | palo alto |
| 17 | orens hummus shop (university ave) | palo alto |
| 17 | pediatric dentistry of palo alto | palo alto |
| 17 | douce france cafe & bakery | palo alto |
| 17 | cvs (el camino real) | palo alto |
| 17 | nob hill foods (grant road) | mountain view |
| 17 | peets coffee (el camino real) | palo alto |

# Predictions for Economics

- Adoption of off-the-shelf ML methods for their intended tasks (prediction, classification, and clustering, e.g. for textual analysis)
- Extensions and modifications of prediction methods to account for considerations such as fairness, manipulability, and interpretability
- Development of new econometric methods based on machine learning designed to solve traditional social science estimation tasks, e.g. causal inference
- Increased emphasis on model robustness and other supplementary analysis to assess credibility of studies
- Adoption of new methods by empiricists at large scale
- Revival and new lines of research in productivity and measurement
- New methods for the design and analysis of large administrative data, including merging these sources
- Increase in interdisciplinary research
- Changes in organization, dissemination, and funding of economic research
- "Economist as engineer" engages with firms, government to design and implement policies in digital environment
- Design and implementation of digital experimentation, both one-time and as an ongoing process, in collaboration with firms and government
- Increased use of data analysis in all levels of economics teaching; increase in interdisciplinary data science programs
- Research on the impact of AI and ML on economy

# What Are Unique Features of Cross-Sectional Econometrics v. Other Branches of Statistics?

- Framework and language for causality
- Causal inference from observational data
  - Theory and PRACTICE
- Structural models to do counterfactuals for environments that have never been observed
- Emphasis on interpretable (~causal) models
- Relatively little emphasis on systematic model selection in applied micro-econometrics
  - Even in environments where theory does not motivate functional forms
- Emphasis on standard errors for a pre-specified models
  - Estimators must have established properties

# What We Say v. What We Do (Econometrics)

- What We Say
  - Causal inference and counterfactuals
  - God gave us the model
  - We report estimated causal effects and appropriate standard errors
  - Plus a few additional specifications for robustness
- What we do
  - Run OLS or IV regressions
    - Try a lot of functional forms
    - Report standard errors as if we ran only one model
    - Have research assistants run hundreds of regressions and pick a few "representative" ones
  - Use complex structural models
    - Make a lot of assumptions without a great way to test them

# Some Broad Generalizations About ML Versus Cross-Sectional Econometrics

- Guiding principle: prediction
  - Training, testing
  - Big concern: overfitting with small data
  - Also: underfitting with large data
- Counterfactuals: within current "regime"
  - If joint distribution among variables changes, just retrain your model
  - Many argue that predicting for a new stochastic process not justified
- Some key features
  - Quality of a predictive algorithm can be summarized in a single number per observation
  - Can assess performance in a model-free way
- Relatively small ML literature on causality
  - "graphical" representations of causal relationships (Judea Pearl)
  - Reinforcement learning & bandit problems
  - Little empirical work outside of randomized experiments, no IV or IV analog
  - If model predicts well in current regime, what more do you need?
- Relatively little emphasis on statistical properties of estimators or interpretability of models
- Not historically an empirical field—not about measurement/estimation or about the numbers

# What We Say v. What We Do (ML)

- What we say
  - ML = Data Science, statistics
    - Is there anything else?
  - Use language of answering questions or solving problems, e.g. advertising allocation, salesperson prioritization
  - Aesthetic: human analyst does not have to make any choices
  - All that matters is prediction
- What we do
  - Use predictive models and ignore other considerations, e.g. causality
  - Wonder/worry about interpretability/reliability/robustness/adaptability, but have little way to conceptualize or ask algos to optimize for it

# Some Lessons for Econometrics: More Emphasis on Validation

- Model "validation" essential in ML but often neglected in econometrics
  - To be fair, we are asking harder counterfactual questions
  - We are using models less prone to "overfitting"
- Examples in econometrics
  - Fitting moments that weren't used for estimation
  - Testing assumptions of structural models
  - Meta-studies of merger predictions v. outcomes
  - Athey/Levin/Seira (QJE), Athey-Coey-Levin (AEJ:Micro) on timber where we estimate on sealed-bid, unrestricted sales and predict to open ascending or small business

# Some Lessons for Econometrics: More Emphasis on Model Selection

- We don't *really* pick specifications in advance, but we don't emphasize our selection procedures
  - For larger datasets, really need systematic model selection
    - Regularized regression, etc.
  - Robustness
    - Athey and Imbens, 2015—standard deviation of estimates across models
  - Supplementary Analysis
    - See Athey and Imbens 2017 (JEP) for a review
    - Athey, Imbens, Pham and Wager (2017), etc.

- Need methods palatable and interpretable for applied research, valid standard errors

# ML and Causal Inference: Average Treatment Effects Under Unconfoundedness

- Focusing on prediction only leads to bias
  - To remove bias, control for confounders
  - Focusing on prediction "zero's out" confounders with weak effect on outcomes
  - Belloni, Chernozukov, and Hansen (series of papers)
  - Use LASSO as a variable selection method
    - Y on X
    - W on X
    - OLS of Y on W, selected X
    - Early example to show that Prediction and ML should have different objectives!

- Estimating propensity scores/assignment model neither necessary or a good idea
  - Assignment models often complex
  - Hard to estimate accurately in high dimensions
  - Focus directly on covariate balance
  - Athey, Imbens and Wager (2016) method does not rely on estimable propensity score
- Orthogonalization helps
  - Both BCH and AIW approaches rely on residualization
  - Hard to estimate high-dimensional models accurately
- Residual on Residual regression using ML – Chernozhukov et al (2017)

# Difference in Difference, Panel Data

- Doudchenko and Imbens (2017)
  - Regularized regression for Synthetic Control
- Athey, Bayati, Doudchenko, Imbens, Khosravi (2017)
  - Fit a matrix to panel data with penalization for "complexity"
  - Find general cross-sectional and time series patterns
  - Works with "wide" or "narrow" data
  - Observation: estimating what would have happened in the absence of the treatment is a prediction problem

# Heterogeneous Treatment Effects: Experiments, Unconfoundedness, IV, GMM

- Estimating heterogeneity with limited complexity
  - Causal Tree (Athey and Imbens, PNAS 2016)
    - Tailored objective, std errors
    - Sample splitting
    - Many applications from health to field experiments
  - Trees with GMM/ML Models
    - Zeiles (2008)
    - Asher, Nekipelov, Novosad, Ryan (2016)
    - Athey, Tibshirani, and Wager (2016)
  - LASSO
    - "Interpretability"? Arguably harder than trees when omitted variables.
    - E.g. Imai and Ratkovic, 2013
  - "Deep IV"
    - Matt Taddy, Greg Lewis et al (2017)

- Non-parametric estimation
  - $\hat{\tau}(x) = E[\tau_i | X_i = x]$
  - This is a hard problem!
  - Forest-based methods
    - Wager and Athey (2015) provide first asymptotic normality results, confidence intervals
    - Athey, Tibshirani, and Wager (2016) – any GMM model, e.g. IV, with confidence intervals
    - Use forests to generate weights
    - Forests replace kernels wherever they are used
  - "Deep IV"
    - Matt Taddy, Greg Lewis et al (2017)

# Optimal Policy Estimation

- E.g. personalized medicine
- A variety of approaches from ML literature
  - Imports ideas from causal inference literature such as propensity score weighting
  - Little attention to econometric efficiency
- Kitagawa and Tetenov (forthcoming, EMA)
- Athey and Wager (2017)
  - Improve the performance bringing in orthogonalization and ideas from econometric efficiency
- Bandits & Contextual Bandits
  - Steve Scott (Google)
  - John Langford team (MSR)
  - Eytan Bakshy team (Facebook)
  - Athey et al (methods & applications in progress… stay tuned)

# Some Lessons for Econometrics: Large Scale Bayesian Models

- ML & Econometrics closest when we do Bayesian statistics
- ML has well-developed literature on large scale
- Athey-Nekipelov (2014) – advertisers with heterogeneous preferences in search
- David Blei et al techniques
- Use matrix factorization for consumer demand systems with aggregated (Taddy et al 2017) or individual discrete choice (Athey et al (2017))