Machine Learning
for Economists

Introduction

Anthony Strittmatter

Course Outline

1. Overview about Machine Learning in Economic Applications
2. Machine Learning for Prediction:
   - Regularized Regression Methods (Lasso and Ridge)
   - Non-parametric Machine Learning Methods (Trees and Forests)
3. Causal Machine Learning to Account for High-Dimensional Confounding:
   - Matching
   - Instrumental Variable Approach
   - Difference-in-Differences
4. Open (Flexible) Session:
   - Optimal Instruments
   - Effect Heterogeneity
   - Optimal Policy Learning

Schedule

Date	Time	
Monday 02.09.	14.00-15.30	Overview/Prediction
Tuesday 03.09.	10.45-12.15	Prediction
Tuesday 03.09.	14.00-15.30	Prediction
Wednesday 04.09.	10.30-12.00	Prediction/Causal
Wednesday 04.09.	14.00-15.30	Causal
Thursday 05.09.	10.45-12.15	Causal
Thursday 05.09.	14.00-15.30	Causal/Flexible
Friday 06.09.	10.45-12.15	Flexible

PC-Sessions

You need to bring a laptop to the PC-sessions. PC-sessions will be in R using Jupyter Notebooks. You can choose one of the following three options to participate in the PC-sessions:

1. Work online on MyBinder:
   - No need to install anything.
   - You work in Jupyter Notebooks, but have the option to download files in R.
   - You can access the PC-sessions on my Github repository: https://github.com/AStrittmatter/ZEW-Course
2. Install R:
   - Download and install R and Rstudio (open source).
   - You need to download the packages we are using during the sessions.
   - You will work in R on your own PC.

PC-Lab Sessions (cont.)

3. Install Anaconda:
   - Download and install Anaconda (open source). I recommend to save Anaconda in your private user folder to ensure proper access rights (but no guarantee).
   - Open Anaconda, go to Environments, and install “r-irkernel” and “r-essentials”.
   - No need to download additional packages.
   - You will work in Jupyter Notebooks on your own PC (you also have the option to download RStudio into your Anaconda navigator).

Main Literature

Recommended readings:

- Mullainathan and Spiess (2017): “Machine Learning: An Applied Econometric Approach”, Journal of Economic Perspectives, 31 (2), pp. 87-106, download.
- Athey (2019): “Beyond Prediction: Using Big Data for Policy Problems”, Science, 355 (6324), pp. 483-485, download.

Additional introductory literature:

- James, Witten, Hastie, and Tibshirani (2013): ”An Introduction to Statistical Learning”, Springer, download.
- Hastie, Tibshirani, and Friedman (2009): “Elements of Statistical Learning”, 2nd ed., Springer, download.
- Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, and Newey (2017): “Double/Debiased/Neyman Machine Learning of Treatment Effects”, American Economic Review, P&P, 107 (5), pp. 261-265, download.

Obligations of Course Participants

- Ask many questions.
- Give me feedback if the course is too slow or too fast.
- Meet me outside the lecture to talk about your research plans.
- Tell me your preferences for the flexible part of the course.
- I provide voluntary extra PC exercises. Feel free to talk with me outside the lecture about them.

Why Machine Learning (ML)?

- ML (or statistical learning) methods exist already since decades
- Currently ”Machine Learning” is a buzz word with no clear definition
- Probably most people think of ML as some computational intensive methods that make data-driven modelling decisions

- Consider the structural model $Y = f(X, U)$
  - ML methods can be very powerful to predict $\widehat{Y} = \widehat{f}(X)$ by balancing the bias-variance trade-off
  - ML methods can incorporate many (or even high-dimensional) covariates X in a convenient way
  - ML methods can model $\widehat{f}(\cdot)$ in a very flexible and data-driven way
  - **Main disadvantage:** ML is a black-box approach and we loose the interpretability of $\widehat{f}(X)$

LASSO Example: Predicting House Prices

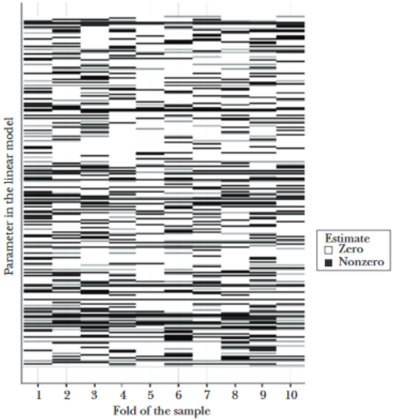
$$\arg\min_{\beta} \left\{ \sum_{i=1}^{N} \left(Y_i - \beta_0 - \sum_{j=1}^{p} X_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

	OLS	LASSO	Post-LASSO
Intercept	11.59	11.30	11.26
Baths	0.288	0.251	0.289
Bedrooms	0.011		
Kitchen	-0.270		
Living rooms	-0.126		
Lot size	-0.00000002		
Rooms	0.131	0.108	0.124
Garage	-0.540	-0.377	-0.549

Note: The sample contains 51,808 houses from American Housing Survey. House prices are measured in log dollars.

Selected Covariates Across 10 LASSO Regressions

Barcode Graph:

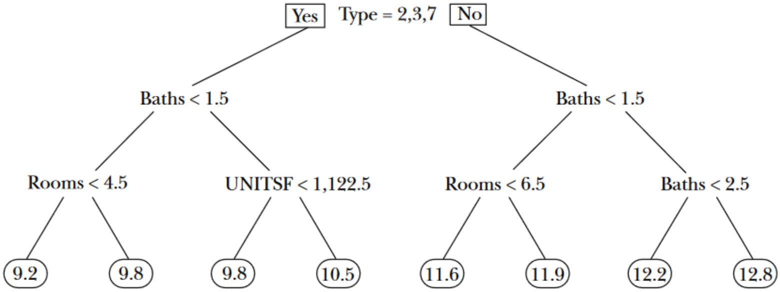


Source: Mullainathan and Spiess (2017)

$\rightarrow$ We do not learn the “true” structural model from ML

$\rightarrow$ ML is a black-box approach

Shallow Tree Example: Predicting House Prices



Note: Based on a sample from the 2011 American Housing Survey metropolitan survey. House-value predictions are in log dollars.

Source: Mullainathan and Spiess (2017)

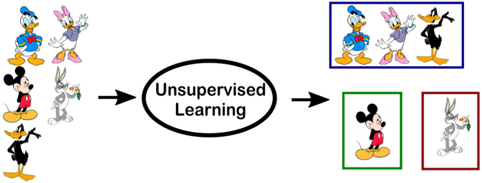
Supervised vs. Unsupervised Machine Learning

Supervised Machine Learning:

- We observe data on Y and X and want to learn the mapping $\widehat{Y} = \widehat{f}(X)$
- Classification when $\widehat{Y}$ is discrete, regression when $\widehat{Y}$ is continuous

Unsupervised Machine Learning:

- We observe only data on X and want to learn something about its structure
- Clustering: Partition data into homogeneous groups based on X



- Factor analysis (e.g., principal component analysis)

When could Predictions be Useful?

Tasks with a prediction purpose:

- Predict stock or commodity prices using Twitter data.
- Nowcasting unemployment rate or GDP using Google search queries.
- Pre-screening of job applications.
- Consumer demand (shipping before the order occurs).
- Movie recommendations on Netflix.
- Handwriting, image, face, or voice recognition.

Economic Studies Using ML for Prediction

- Glaeser, Kominers, Luca, and Naik (2016) use images from Google Street View to measure block-level income in New York City and Boston
- Jean et al. (2016) train a neural net to predict local economic outcomes from satellite data in African countries
- Chandler, Levitt, and List (2011) predict shootings among high-risk youth so that mentoring interventions can be appropriately targeted
- Kang, Kuznetsova, Luca, and Choi (2013) use restaurant reviews on Yelp.com to predict the outcome of hygiene inspections
- Huber and Imhof (2018) use machine learning to detect bid-rigging cartels in Switzerland

Predictions vs. Causal Inference

- Outcome (e.g., house price): Y
- Binary Treatment (e.g., incinerator in neighbourhood): $D \in \{0, 1\}$
- Potential Outcome:
  - $Y(1)$ house price when an incinerator is in the neighbourhood
  - $Y(0)$ house price when no incinerator is in the neighbourhood

  $\rightarrow$ Only one potential outcome can be observed
- Causal effect: $\delta = Y(1) - Y(0)$

$\rightarrow$ Predictions have the observable estimation target $\widehat{Y}$

$\rightarrow$ Causal inference has the (partly) unobservable estimation target $\widehat{\delta}$

Training of ML Algorithms

Out-of-Sample Mean-Squared-Error (MSE):

- Observed outcome Y:

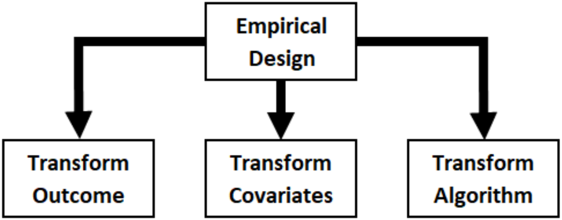
$$MSE_{\widehat{Y}} = E\left[(\widehat{Y} - Y)^2\right] = \underbrace{E\left[(\widehat{Y} - E[\widehat{Y}])^2\right]}_{\text{Variance}} + \underbrace{(E[\widehat{Y} - Y])^2}_{\text{Squared Bias}}$$

- Causal Effect δ:

$$MSE_{\widehat{\delta}} = E\left[(\widehat{\delta} - \delta)^2\right] = \underbrace{E\left[(\widehat{\delta} - E[\widehat{\delta}])^2\right]}_{\text{Variance}} + \underbrace{(E[\widehat{\delta} - \delta])^2}_{\text{Squared Bias}}$$

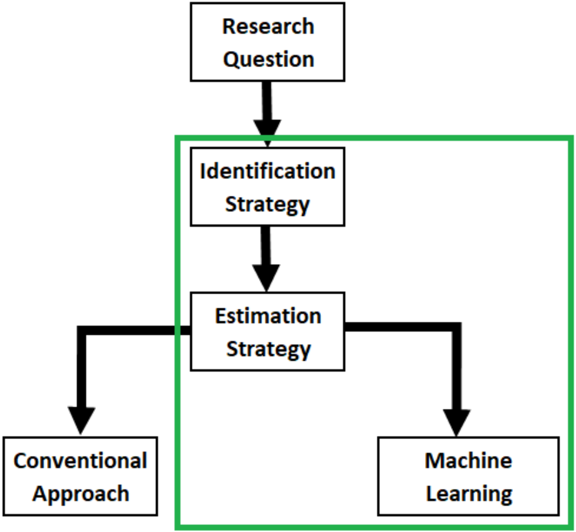
$\rightarrow$ δ is unobservable

Causal Machine Learning (CML) Designs



⇒ Knaus, Lechner, Strittmatter (2018) provide a comparison of all designs.

Causal Machine Learning (CML)



When could CML be Useful?

- Account for (very) many instruments in IV or Heckit approach (prediction problem, issues with inference).
- Account for confounders, e.g., in matching, IV, or difference-in-difference approaches:
  - ML enables the incorporation of (very) many covariates which can make the exclusion restriction more credible.
  - Some ML approaches make little functional form assumptions.
- Heterogeneous effects:
  - Principled approach makes it less likely to overlook important heterogeneity.
  - Problem: Issues with interpretability and valid inference (functionals).
- Optimal policy rules:
  - Focus on the (discrete) treatment decision instead on the effect size.

Limitations of Causal Machine Learning (CML)

- ML algorithms cannot distinguish between causation and correlation.
  - $\rightarrow$ CML will not select the relevant causal parameters automatically.
  - $\rightarrow$ We have to provide some structure to the CML algorithm.
- CML can estimate causal effects only for a few (usually only one) endogenous variables.
  - $\rightarrow$ We will not obtain the (complete) structural model.
- Identifying assumptions do not change, no matter if we use ML methods or not.
- We should resist the temptation to interpret prediction models in a causal way.

Applications of CML Methods

- Bertrand, Crépon, Marguerie, and Premand (2017) and Knaus, Lechner, and Strittmatter (2017) estimate the employment effects of training programmes for unemployed persons.
- Knaus (2018) estimates the effects of musical practice on student's skills and selects confounders with ML methods.
- Davis and Heller (2017) investigate the effects of summer jobs on the probability to commit a violent crime.
- Ascarza (2018) target marketing campaigns.
- Andini et al. (2018) target tax rebates.