DOTE 6635: Artificial Intelligence for Business Research

Double Machine Learning

Renyu (Philip) Zhang

1

Machine Learning for Causal Inference

- Using machine learning for causal inference is generally very challenging.
 - Cross-validation cannot be directly applied to hyper-parameter tuning for causal inference models (Athey and Imbens, 2016).
 - Good performance in predicting propensity scores or outcomes cannot be directly translated into good causal performance (Belloni et al., 2014).
 - Regularization in ML will introduce additional biases in causal inference (Belloni et al., 2016).
- · What are the most critical issues in causal inference at large?
 - · Confoundedness, non-overlapping, balance, etc.
 - AI/ML cannot magically solve these fundamental problems of causal inference.
- Double machine learning (DML) provides a framework to empower causal inference with ML.
 - Compared with other fields in business research, this is a very fast-evolving field of study.

Recursive partitioning for heterogeneous causal effects <u>S. Athey. G. Imbens.</u> Proceedings of the National Academy of Sciences, 2016 - pnas.org
... We refer to the estimators developed in this section as "causal tree" (CT) estimators.... for constructing trees for causal effects that allow us to do valid inference for the causal effects in

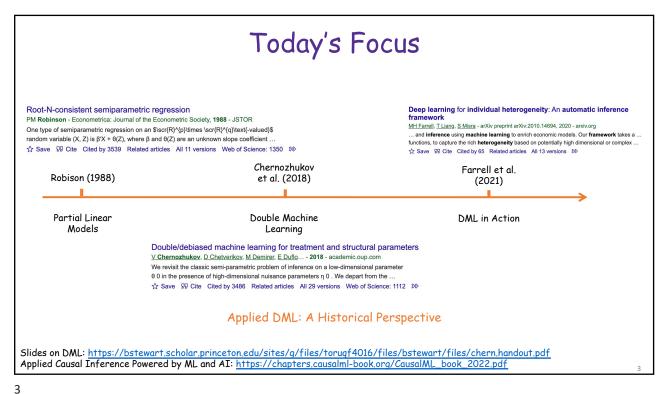
\$\times\$ Section Section 117 versions \$\times\$ In 117 versions \$\times\$

Post-selection inference for generalized linear models with many controls <u>ABelloni, V Chernozhukov, Y Wei</u> - Journal of Business & ..., 2016 - Taylor & Francis
This article considers generalized linear models in the presence of many controls. We lay out a general methodology to estimate an effect of interest based on the construction of an ...

☆ Save 99 Cite Cited by 233 Related articles All 11 versions Web of Science: 87 ≫

2

2



3

Agenda

- Partial Linear Models
- General Double Machine Learning Framework
- · Double Machine Learning in Action: Practical Recipe and Pitfalls

4

4

High-Level Takeaways of the DML Literature



Causal Inference: A Statistical Learning Approach https://web.stanford.edu/~swager/causal_inf_book.pdf
Applied Causal Inference Powered by ML and AI: https://chapters.causalml-book.org/CausalML_book_2022.pdf
DML Package: https://chapters.causalml-book.org/CausalML_book_2022.pdf

Slides on DML: https://bstewart.scholar.princeton.edu/sites/q/files/toruqf4016/files/bstewart/files/chern.handout.pdf

- Provides a general framework, by leveraging Neyman Orthogonality, for estimating treatment effects using ML methods.
 - Causal inference usually requires estimating the, expected outcomes (and propensity scores) conditioned
 on covariates or confounders
 - Standard econometrics methods make strong functional form assumptions (e.g., linear models), which require strong substantive justifications; if mis-specified, causal estimates will be significantly biased.
 - DML framework automatically learns the form of conditional expectation functions from data.
- DML framework requires:
 - · Some regularity conditions (recall the assumptions for AIPW)
 - ML estimators to converge, in RMSE/L2-norm at a rate of $o(n^{-1/4})$, slower than $o(n^{-1/2})$, the rate of most parametric models according to the delta method.
- DML framework outputs:
 - Root-n consistent estimators for treatment effects: Convergence to the ground-truth in probability at a rate $O(n^{-1/2})$, a property natural and common in a parametric world.
 - In frequentist perspective, root-n consistency typically means asymptotically normal, which means you can construct valid confidence intervals and do inference on your estimators.

5

Let's First Look at a Simple Model



Causal Inference: A Statistical Learning Approach https://web.stanford.edu/~swager/causal_inf_book.pdf
Applied Causal Inference Powered by ML and AI: https://chapters.causalml-book.org/CausalML_book_2022.pdf
DML Package: https://docs.doubleml.org/stable/index.html#

 ${\bf Slides\ on\ DML:\ } \underline{https://bstewart.scholar.princeton.edu/sites/g/files/toruqf4016/files/bstewart/files/chern.handout.pdf}$

- We use the partial linear model (PLM) to illustrate the DML framework:
 - Why ML is important and useful;
 - · How the statistical theory works.

Partially Linear Model Set-up

- Y: Outcome
- D: Treatment

$$Y = D\theta_0 + g_0(X) + U$$

X: Measured confounders

$$D=m_0(X)+V$$

- \bullet U and V are our error terms
- We assume zero conditional mean:

$$E[U | X, D] = 0$$
 $E[V | X] = 0$

6

6