

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

[S. Athey](#), [G. Imbens](#) - 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 ...
 ☆ Save ⓘ Cite Cited by 2213 Related articles All 17 versions ⓘ

Inference on treatment effects after selection among high-dimensional controls

[A. Belloni](#), [V. Chernozhukov](#)... - Review of Economic Studies, 2014 - academic.oup.com
 We propose robust methods for inference about the effect of a treatment variable on a scalar outcome in the presence of very many regressors in a model with possibly non-Gaussian ...
 ☆ Save ⓘ Cite Cited by 2040 Related articles All 32 versions ⓘ

Post-selection inference for generalized linear models with many controls

[A. Belloni](#), [V. Chernozhukov](#), Y. Wei - Journal of Business & Economic Statistics, 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 ⓘ Cite Cited by 233 Related articles All 11 versions Web of Science: 87 ⓘ

2

2

Today's Focus

Root-N-consistent semiparametric regression

PM Robinson - Econometrica: Journal of the Econometric Society, 1988 - JSTOR

One type of semiparametric regression on an $\text{scalar}(R)^p \times \text{scalar}(R)^q \text{vector}(-\text{valued})$ random variable (X, Z) is $\beta'X + \theta(Z)$, where β and $\theta(Z)$ are an unknown slope coefficient ...

☆ Save Cite Cited by 3539 Related articles All 11 versions Web of Science: 1350

Deep learning for individual heterogeneity: An automatic inference framework

MH Farrell, T Liang, S Misra - arXiv preprint arXiv:2010.14694, 2020 - arxiv.org

... and inference using machine learning to enrich economic models. Our framework takes a ... functions, to capture the rich heterogeneity based on potentially high dimensional or complex ...

☆ Save Cite Cited by 65 Related articles All 13 versions

Robinson (1988)

Chernozhukov
et al. (2018)

Farrell et al.
(2021)

Partial Linear
Models

Double Machine
Learning

DML in Action

Double/debiased machine learning for treatment and structural parameters

V Chernozhukov, D Chetverikov, M Demirer, E Dufo... - 2018 - academic.oup.com

We revisit the classic semi-parametric problem of inference on a low-dimensional parameter θ_0 in the presence of high-dimensional nuisance parameters η_0 . We depart from the ...

☆ Save Cite Cited by 3486 Related articles All 29 versions Web of Science: 1112

Applied DML: A Historical Perspective

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

Applied Causal Inference Powered by ML and AI: https://chapters.causalmachinelearning.org/CausalML_book_2022.pdf

3

3

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.causalmachinelearning.org/CausalML_book_2022.pdf

DML Package: <https://docs.doubleml.org/stable/index.html#>

Slides on DML: <https://bstewart.scholar.princeton.edu/sites/g/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**.

4

4

Agenda

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

5

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.causalm1-book.org/CausalML_book_2022.pdf
 DML Package: <https://docs.doubleml.org/stable/index.html#>
 Slides on DML: <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
- X : Measured confounders
- U and V are our error terms
- We assume zero conditional mean:

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

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

$$E[U \mid X, D] = 0 \quad E[V \mid X] = 0$$

6

6