

# Machine Learning for Causal Inference

## Exercise based on Dorie et al. 2019

Ian Lundberg

# Automated versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition<sup>1</sup>

Vincent Dorie, Jennifer Hill, Uri Shalit, Marc Scott and Dan Cervone

**Abstract.** Statisticians have made great progress in creating methods that reduce our reliance on parametric assumptions. However, this explosion in research has resulted in a breadth of inferential strategies that both create opportunities for more reliable inference as well as complicate the choices that an applied researcher has to make and defend. Relatedly, researchers advocating for new methods typically compare their method to at best 2 or 3 other causal inference strategies and test using simulations that may or may not be designed to equally tease out flaws in all the competing methods. The causal inference data analysis challenge, “Is Your SATT Where It’s At?”, launched as part of the 2016 Atlantic Causal Inference Conference, sought to make progress with respect to both of these issues. The researchers creating the data testing grounds were distinct from the researchers submitting methods whose efficacy would be evaluated. Results from 30 competitors across the two versions of the competition (black-box algorithms and do-it-yourself analyses) are presented along with post-hoc analyses that reveal information about the characteristics of causal inference strategies and settings that affect performance. The most consistent conclusion was that methods that flexibly model the response surface perform better overall than methods that fail to do so. Finally new methods are proposed that combine features of several of the top-performing submitted methods.

**Key words and phrases:** Causal inference, competition, machine learning, automated algorithms, evaluation.

Vincent Dorie is Associate Research Scientist, Data Science Institute, Columbia University, 475 Riverside Drive, Room 320L, New York, New York 10115, USA (e-mail: [vdorie@gmail.com](mailto:vdorie@gmail.com)). Jennifer Hill is Professor of Applied Statistics and Data Science, Department of Applied Statistics, Social Science, and Humanities, New York University, 246 Greene Street, 3rd Floor, New York, New York 10003, USA (e-mail: [jennifer.hill@nyu.edu](mailto:jennifer.hill@nyu.edu)). Uri Shalit is Assistant Professor, Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, Technion City, Haifa 3200003, Israel (e-mail: [urishalit@technion.ac.il](mailto:urishalit@technion.ac.il)). Marc Scott is Professor of Applied Statistics, Department of Applied Statistics, Social Science, and Humanities, New York University, 246 Greene Street, 3rd Floor, New York, New York 10003, USA (e-mail: [marc.scott@nyu.edu](mailto:marc.scott@nyu.edu)). Dan Cervone is Director of

## 1. INTRODUCTION

In the absence of a controlled randomized or natural experiment,<sup>2</sup> inferring causal effects involves the difficult task of constructing fair comparisons between ob-

*Quantitative Research, Los Angeles Dodgers, Dodger Stadium, 1000 Vin Scully Ave., Los Angeles, California 90012, USA (e-mail: [dcervone@gmail.com](mailto:dcervone@gmail.com)).*

<sup>1</sup> Discussed in 10.1214/18-STS684; 10.1214/18-STS680; 10.1214/18-STS690; 10.1214/18-STS689; 10.1214/18-STS679; 10.1214/18-STS682; 10.1214/18-STS688

<sup>2</sup>We use natural experiment to include (1) studies where the causal variable is randomized not for the purposes of a study (for instance, a school lottery), (2) studies where a variable is randomized but the causal variable of interest is downstream of this (e.g., plays the role of an instrumental variable), and (3) regression discontinuity designs.

# Dorie et al. 2019<sup>1</sup>: Is Your SATT Where It's At?

---

<sup>1</sup>Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019). "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." *Statistical Science*, 34(1), 43-68. See also <https://jenniferhill7.wixsite.com/acic-2016/competition>

# Dorie et al. 2019<sup>1</sup>: Is Your SATT Where It's At?

- ▶ Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

---

<sup>1</sup>Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019).

“Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition.” *Statistical Science*, 34(1), 43-68. See also <https://jenniferhill17.wixsite.com/acic-2016/competition>

# Dorie et al. 2019<sup>1</sup>: Is Your SATT Where It's At?

- ▶ Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

- ▶ Simulated data. SATT was known to organizers

---

<sup>1</sup>Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019). "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." *Statistical Science*, 34(1), 43-68. See also <https://jenniferhill7.wixsite.com/acic-2016/competition>

# Dorie et al. 2019<sup>1</sup>: Is Your SATT Where It's At?

- ▶ Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

- ▶ Simulated data. SATT was known to organizers
- ▶ Confounders were defined by the organizers

---

<sup>1</sup>Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019).

“Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition.” *Statistical Science*, 34(1), 43-68. See also <https://jenniferhill7.wixsite.com/acic-2016/competition>

# Dorie et al. 2019<sup>1</sup>: Is Your SATT Where It's At?

- ▶ Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

- ▶ Simulated data. SATT was known to organizers
- ▶ Confounders were defined by the organizers
- ▶ Participants could use any algorithm to estimate SATT

---

<sup>1</sup>Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019).

“Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition.” *Statistical Science*, 34(1), 43-68. See also <https://jenniferhill7.wixsite.com/acic-2016/competition>

# Dorie et al. 2019<sup>1</sup>: Is Your SATT Where It's At?

- ▶ Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

- ▶ Simulated data. SATT was known to organizers
- ▶ Confounders were defined by the organizers
- ▶ Participants could use any algorithm to estimate SATT
- ▶ 30 teams attempted the task

---

<sup>1</sup>Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019).

“Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition.” *Statistical Science*, 34(1), 43-68. See also <https://jenniferhill7.wixsite.com/acic-2016/competition>

# Dorie et al. 2019<sup>1</sup>: Is Your SATT Where It's At?

- ▶ Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

- ▶ Simulated data. SATT was known to organizers
- ▶ Confounders were defined by the organizers
- ▶ Participants could use any algorithm to estimate SATT
- ▶ 30 teams attempted the task
- ▶ Today you will attempt it!

---

<sup>1</sup>Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019).

“Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition.” *Statistical Science*, 34(1), 43-68. See also <https://jenniferhill7.wixsite.com/acic-2016/competition>

# A few things to help you succeed

## 1. A few algorithms you might consider

- ▶ Ridge, LASSO, elastic net (glmnet)
- ▶ Random forest (ranger)
- ▶ Bayesian additive regression trees (BART)
- ▶ Super Learner (SuperLearner)

## 2. How do I choose a black-box algorithm?

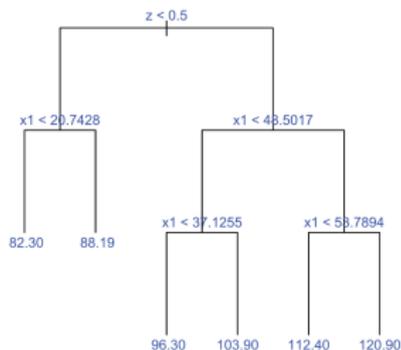
## 3. Overview of the data structure

## A few algorithms you might consider: glmnet

- ▶ Idea: With many coefficients, OLS can be high-variance.
- ▶ glmnet **penalizes** coefficients to reduce sample variance
  - ▶ Pushes coefficients toward 0
  - ▶ When we are uncertain about  $\hat{\beta}$ , better to keep  $\hat{\beta}$  small
- ▶ Three ways to penalize
  - ▶ Ridge penalty: Minimize  $\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 + \lambda \sum_j \beta_j^2$ 
    - ▶ Coefficients are pulled toward 0, but not exactly to 0
  - ▶ Lasso penalty: Minimize  $\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 + \lambda \sum_j |\beta_j|$ 
    - ▶ Some coefficients pushed exactly to 0 (dropped out entirely)
  - ▶ Elastic net: Penalize both  $\beta_j^2$  and  $|\beta_j|$

# A few algorithms you might consider: ranger<sup>2</sup>

- ▶ Random forest: A frequentist sum-of-trees model



(tree visualization from Hill 2011)

- ▶ Good at learning interactions among covariates
- ▶ Ranger is fast

---

<sup>2</sup>Wright, M. N., & Ziegler, A. (2017). [ranger: A fast implementation of random forests for high dimensional data in C++ and R](#). *Journal of Statistical Software*, 77(i01).

## A few algorithms you might consider: BART<sup>3</sup>

- ▶ Bayesian version of random forest
  - ▶ A prior regularizes estimates
  - ▶ Bonus: Free posterior variance estimates!
- ▶ Warning: A bit slower than ranger

---

<sup>3</sup>Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. “[BART: Bayesian additive regression trees](#).” The Annals of Applied Statistics 4.1 (2010): 266-298.

# A few algorithms you might consider: SuperLearner<sup>4</sup>

Why pick just one algorithm?

1. Fit many candidate learners  $f_1(), f_2(), \dots$
2. Predict out-of-sample (using cross-validation)
3. Learn a set of weights to take a weighted average

$$\hat{f}(\vec{\ell}, a) = \hat{\beta}_1 \underbrace{\hat{f}_1(\vec{\ell}, a)}_{\text{e.g. OLS}} + \beta_2 \underbrace{\hat{f}_2(\vec{\ell}, a)}_{\text{e.g. glmnet}} + \beta_3 \underbrace{\hat{f}_3(\vec{\ell}, a)}_{\text{e.g. ranger}}$$

---

<sup>4</sup>Original: Van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). [Super learner](#). *Statistical Applications in Genetics and Molecular Biology*, 6(1).  
Good intro paper: Naimi, A. I., & Balzer, L. B. (2018). [Stacked generalization: an introduction to super learning](#). *European Journal of Epidemiology*, 33(5), 459-464.

## A few algorithms you might consider

- ▶ You could use any of these
- ▶ You could use something else
- ▶ You could do the entire exercise with OLS
  - ▶ try various functional forms
- ▶ Choice is yours!

# Is your SATT where it's at? **Data structure**

- ▶ Simplified dataset on the course site: [dorie.csv](#)
- ▶ This is one simulation from the many in Dorie et al.
- ▶ Variables include
  - ▶ y: outcome (numeric)
  - ▶ z: treatment (binary)
  - ▶ x\_\*: confounders
  - ▶ set: I created this, coded train or test

Code in [dorie\\_example\\_code.R](#) can help you get started.

At the end of class, you will produce an  $\widehat{\text{SATT}}$ .  
I have the truth. We will see who is closest!