AUTHOR: RHEEA-MARIA PRODAN
SUPERVISOR: RICKARD KARLSSON
RESPONSIBLE PROFESSOR: JESSE KRIJTHE

Analyzing the Impact of Depth and Leaf Size on CATE Estimation in Honest Causal Trees

A Study of Model Accuracy and Generalization Across Simulated and Real-World Data

TU Delft

1. Background

- standard ML: focus on predicting outcome from features
- causal ML: estimate treatment effects, how effects vary across individuals [1]

$$CATE = E[Y(1) - Y(0) \mid X = x]$$

- Y(1) - potential outcome if an individual receives treatment
- Y(0) - potential outcome if an individual does not receive treatment (control)
- X=x - individual's covariate profile

Honest Causal Trees

- adaptation of decision trees for causality [2]
- sample splitting:
  - half used for building the tree structure
  - half used for estimating treatment effects

2.Research Question

How do key hyperparameter choices, specifically maximum tree depth and minimum leaf size, affect the accuracy and tendency to overfit or underfit in CATE estimates produced by honest causal decision trees, across both simulated and real-world data settings?

3.Methodology

Evaluate the honest causal trees using synthetic data and the real-world data set IHDP

Simulated Data:

$Yi=\eta(Xi)+0.5(2Wi-1)\tau(Xi)+\epsilon i$

- i - individual
- Yi - observed outcome
- η(Xi) - mean outcome
- τ(Xi) - CATE
- Wi - binary treatment
- ϵi - noise term

Design ID	Description	K (Features)	Noise σ (Type)	$\eta(X_i)$ Formula	$\tau(X_i)$ Formula
		Athey and Imbens Designs			
1	Simple, Low Dim	2	0.1	$0.5 \cdot X_{i,0} + X_{i,1}$	$0.5 \cdot X_{i,0}$
2	Moderate Dim, Non-informative of Treatment Effect Covariates	10	0.1	$0.5 \sum_{k=0}^{1} X_{i,k} + \sum_{k=2}^{5} X_{i,k}$	$\sum_{k=0}^{1} \mathbf{1}\{X_{i,k} > 0\} \cdot X_{i,k}$
3	High Dim, Non-informative of Treatment Effect Covariates	20	0.1	$0.5 \sum_{k=0}^{3} X_{i,k} + \sum_{k=4}^{7} X_{i,k}$	$\sum_{k=0}^{3} \mathbf{1}\{X_{i,k} > 0\} \cdot X_{i,k}$
		Custom Designs			
4	Varying Dim, All Relevant Covariates	2, 5, 10, 15, 20, 25, 50	0.1	$0.5 \cdot X_{i,0} + X_{i,1}$	$0.5 \sum_{k=0}^{K-1} X_{i,k}$
5	Non-Linear CATE	2, 5	0.1	$0.5 \cdot X_{i,0} + X_{i,1}$	$\sin(X_{i,0}) + 2 \cdot \cos(X_{i,1})$
6	CATE with Interaction Terms	2, 5	0.1	$0.5 \cdot X_{i,0} + X_{i,1}$	$X_{i,0} \cdot X_{i,1}$
7	Varying Noise Levels	2	0.01, 0.1, 0.25, 0.5, 0.75, 1	$0.5 \cdot X_{i,0} + X_{i,1}$	$0.5 \cdot X_{i,0}$

Real-World Data:

- Infant Health and Development Program (IHDP) dataset - semi-synthetic version of a randomized control trial
- moderately-sized benchmark: 747 observations, 25 covariates

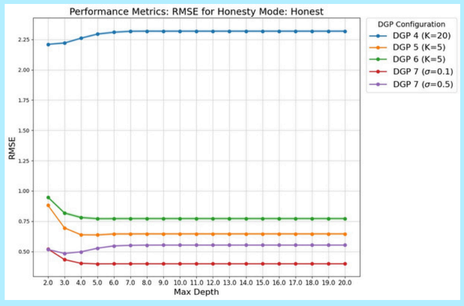
Experiment Setup:

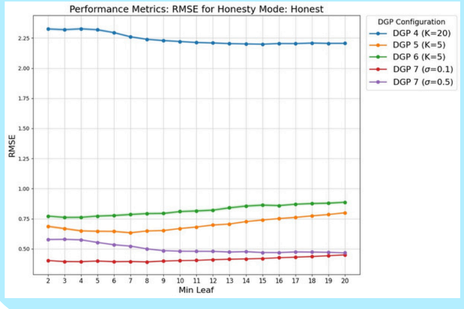
- Vary one hyperparameter at a time, with values between 2 and 20
- Use 3 models: Honest Causal Tree, Adaptive Causal Tree, and T-Learner
- Compute out-of-sample MSE, variance, and bias

Expected Results:

- U-shaped curve indicating bias-variance trade-off
- Causal Trees will exhibit a flatter curve
- Increasing dimensionality will increase the MSE
- Increasing the noise level will increase the MSE

4a. Results - Simulated Data

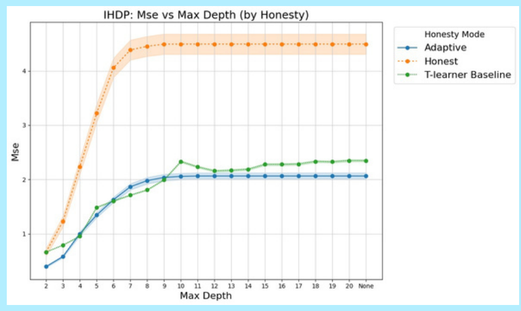


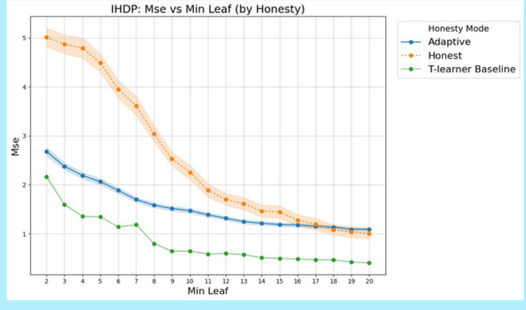


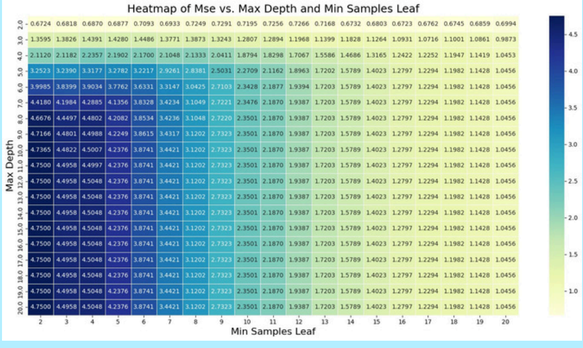
DGP Scenario	Metric	Honest Causal Tree	Adaptive Causal Tree	T-Learner Baseline
Design 4 ($K = 20$)	MSE	5.3085/5.0265	5.7711/5.4271	4.9431/4.5599
	$Bias^2$	0.0666/0.0627	0.0481/0.0460	0.0298/0.0244
	Var	1.4791/0.9999	2.9474/2.4512	2.9115/2.2071
Design 5 ($K = 5$)	MSE	0.4371/0.4939	0.2860/0.2772	0.1550/0.1810
	$Bias^2$	0.0114/0.0130	0.0047/0.0048	0.0008/0.0010
	Var	1.0694/0.9154	1.1196/1.0758	1.3863/1.3274
Design 6 ($K = 5$)	MSE	0.6152/0.6750	0.4079/0.4167	0.2745/0.2960
	$Bias^2$	0.0116/0.0144	0.0047/0.0052	0.0021/0.0023
	Var	0.6960/0.5284	0.8400/0.7502	0.8976/0.8066
Design 7 ($\sigma = 0.1$)	MSE	0.1662/0.1696	0.1092/0.0980	0.1237/0.1518
	$Bias^2$	0.0055/0.0072	0.0028/0.0024	0.0009/0.0012
	Var	0.2525/0.2094	0.2497/0.2305	0.3578/0.3884
Design 7 ($\sigma = 0.5$)	MSE	0.2956/0.2535	0.3230/0.2385	0.4927/0.2555
	$Bias^2$	0.0154/0.0167	0.0071/0.0062	0.0044/0.0029
	Var	0.3834/0.2928	0.4783/0.3932	0.7289/0.4825

- Honest Causal Trees: better for high-dimensionality and noisy environments
  - small maximum depth, large leaf size
- Adaptive Causal Trees: better for low noise levels and complex CATE structures
  - large maximum depth, small leaf size
- T-Learner: strong performance across most DGPs

4b. Results - Real-World Data







- Honest Causal Trees consistently yielded the highest MSE due to the sample splitting
- Deeper trees tend to overfit due to fewer observations in each leaf
- More aggregated leaves improved CATE estimation and generalization
- Grid search showed the lowest MSE(0.6790) is when both hyperparameters are set to 2

5. Discussion & Conclusion

- Optimal hyperparameter configurations are dependent on the data characteristics
- Accuracy of the CATE estimate can suffer due to honesty's sample-splitting

- Honest Causal Trees control variance well
- Adaptive Causal Trees and the T-Learner reduce bias more effectively

References
[1] Stefan Feuerriegel, David Frauen, Viktor Melnychuk, et al. Causal machine learning for predicting treatment outcomes. Nature Medicine, 30(5):958–968, 2024.
[2]Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences, 113(27):7353–7360, 2016.