JuliaCon 2024

Causal Machine Learning with CausalELM

DARREN COLBY

Outline

- Potential Outcomes Framework
- Motivating Examples
- General Workflow
- The ELM in CausalELM
- List of All Estimators
- Comparison with Other CausalML Packages
- Way Forward

Potential Outcomes Framework

- Define x_i as a vector of covariates for each unit i
- Define: $T_i \in \{0, 1\}$ as the treatment indicator
  - When unit i is treated, $T_i = 1$
  - When unit i is not treated, $T_i = 0$
- Define: $Y_i(T_i)$ as the outcome for unit i as a function of T_i
  - When unit i is treated, we observe $Y(1)$
  - When unit i is not treated, we observe $Y(0)$

Potential Outcomes Framework

- The impact of treatment for unit i is just: $\tau_i = Y_i(1) - Y_i(0)$
- Problem: we only observe $Y_i(1)$ or $Y_i(0)$ but not both potential outcomes
- Most models try to use assumptions, data structure, or research designs to overcome this problem

Motivating Example: Effect of 401(k) Plans on Net Worth

- What is the effect of US 401(k) plans on net worth?
  - $E[\tau_i] = E[Y_i(1) - Y_i(0)]$ = Average Treatment Effect (ATE)
- Data comes from a 1994 study by Poterba et. al

Motivating Example: Effect of 401(k) Plans on Net Worth

- x_i includes age, income in USD, family size, marital status, dual earner status, indicator for a defined benefits pension, IRA participation indicator, and home ownership indicator
- T_i denotes eligibility for a 401(k) plan
- Y_i represents net financial assets in USD

Motivating Example: Effect of 401(k) Plan on Net Worth

Could use linear regression but we don't know the functional form

Can also use double machine learning (Chernozhukov et. al, 2018)

- No need to specify functional form

Uses any ML models to get treatment and outcome residuals and combines them in a final linear model

Motivating Example: Effect of 401(k) Plan on Net Worth

- Instantiate a double machine learning model

```
dml_model = DoubleMachineLearning(covariates, treatment, outcome)
```

Motivating Example: Effect of 401(k) Plan on Net Worth

- Estimate the ATE

```
estimate_causal_effect!(dml_model)
```

```
8721.139649868901
```

Motivating Example: Effect of 401(k) Plan on Net Worth

- Get a summary from the model

```
1  summarize(dml_model)
```

```
Dict{Any, Any} with 11 entries:
  "Activation Function"    => swish
  "Quantity of Interest"   => "ATE"
  "Sample Size"            => 9915
  "Number of Machines"     => 50
  "Causal Effect"          => 8721.14
  "Number of Neurons"      => 24
  "Task"                   => "regression"
  "Time Series/Panel Data" => false
  "Standard Error"         => NaN
  "p-value"                => NaN
  "Number of Features"     => 6
```

Motivating Example: Effect of 401(k) Plan on Net Worth

- Validate model assumptions

```
1  validate(dml_model)
```

```
(Dict("0.1 Standard Deviations from Observed Outcomes" => 7903.4400
```

Motivating Example: Effect of Aid on Taliban Violence

- What is the effect of providing compensation to victims of violence in Afghan villages on future Taliban violence?
  - $E[Y_i(1) - Y_i(0)] \mid x_i, T_i]$ = Conditional Average Treatment Effect (CATE)
- Data comes from a 2019 study by Jason Lyall

Motivating Example: Effect of Aid on Taliban Violence

- Could subset by groups and use linear regression
  - Would get very tedious
  - Would not use all the data
- Can also use doubly robust estimation(Kennedy, 2022)
  - No need to specify functional form
- Uses any ML models predict treatments, conditional outcomes, and combines them into final ML model

Motivating Example: Effect of Aid on Taliban Violence

- Instantiate a doubly robust learner

```
dr_model = DoublyRobustLearner(covariates, treatment, outcome)
```

```
DoublyRobustLearner([0.13018867924528302 0.41723100075244546 … 1.0 2
```

Motivating Example: Effect of Aid on Taliban Violence

- Estimate the CATE

```
estimate_causal_effect!(dr_model)
```

```
1061-element Vector{Float64}:
 -1.1248208493165057
 -0.33658630048075666
  0.5502186364208397
  0.44192281327345584
 -0.525542635220144
 -0.012783507830087582
 -0.7809308817892282
 -1.3024783381733798
 -0.7028089822663802
 -0.1506935580311216
  ⋮
 -0.8777671389587851
 -0.9401054875993395
  1.0027280715847917
  0.6316318890842755
 -0.7387045473829954
 -0.6970800035383887
  1.1510660707945082
  0.2809723136292491
  0.1051366508528884
```

Motivating Example: Effect of Aid on Taliban Violence

- Get a summary from the model

```
1  summarize(dr_model)
```

```
Dict{Any, Any} with 11 entries:
  "Activation Function"    => fourier
  "Quantity of Interest"   => "CATE"
  "Sample Size"            => 1061
  "Number of Machines"     => 50
  "Causal Effect"          => [-1.12482, -0.336586, 0.550219, 0.441923, -0.5255…
  "Number of Neurons"      => 124
  "Task"                   => "regression"
  "Time Series/Panel Data" => false
  "Standard Error"         => NaN
  "p-value"                => NaN
  "Number of Features"     => 31
```

Motivating Example: Effect of Aid on Taliban Violence

- Validate model assumptions

```
validate(dr_model)
```

```
(Dict("0.1 Standard Deviations from Observed Outcomes" => 0.3994
```

General Workflow

Instantiate	Estimate	Summarize	Validate
Instantiate a model	Estimate the causal effect of interest	Get a summary of the model	Validate modeling assumptions

The ELM in CausalELM

- All CausalELM estimators need some kind of ML model
- Base model in CausalELM is extreme learning machine (Guang-Bin et. al, 2006)
  - $\mathbf{W}_{\mathbf{feats,\ neurons}} \sim U(-1, 1)$
  - $\mathbf{H} = g(\mathbf{X}\mathbf{W}^{T})$
  - $\boldsymbol{\beta} = \mathbf{H}^{+}\mathbf{Y}$
  - $f(\mathbf{X}_{Test}) = g(\mathbf{X}_{Test}\mathbf{W}^{T})\boldsymbol{\beta}$

The ELM in CausalELM

- Ensemble of ELMs
  - Uses bagging strategy
  - Provides regularization
  - Enables probabilistic outputs
- Activation functions
  - Provide regularization
  - Can greatly improve performance

List of All Estimators

Average Effect Estimators

- Interrupted Time Series
- G-computation
- Double Machine Learning

CATE Estimators

- S-learning
- T-learning
- X-learning
- R-learning
- Doubly Robust Estimation

Comparison with Other CausalML Packages

Simplicity	No required imports Only need to know four functions Randomization inference for all models Less flexibility
Interdisciplinary	G-computation and E-value from epidemiology Many estimators from econometrics literature Randomization inference from work on causal inference
Lightweight	Written in Julia standard library


POSSIBLY YOU!
INVESTORS?
INVESTORS? POSSIBLY YOU!
MakeAGif.com

Way Forward

Thank you!