June 12, 2019

ICML 2019

Classifying Treatment Responders Under Causal Effect Monotonicity

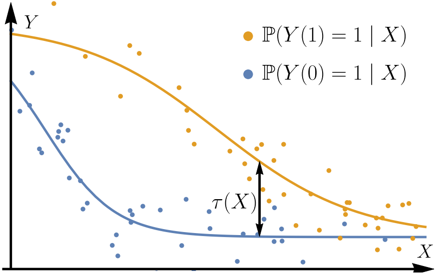
Nathan Kallus

ORIE and Cornell Tech, Cornell University

Heterogeneous Treatment Effect Estimation

X_{Age}	X_{Weight}	X_{BMI}	X_{SysBP}	T (Anticoagulant)	Y (Hemorrhage)
49	106	31		Warfarin	1
54	89	26		None	0
43	130	38		None	1
⋮	⋮	⋮	⋮	⋮	⋮

Fit CATE $\tau(X) = \mathbb{E}[Y(1) - Y(0) \mid X]$ to data on X, T, Y



E.g.:
Causal Forest (Wager & Athey '17),
TARNet (Shalit et al. '17),
...

Often Outcome is **Binary**

Treatment (T)	Outcome Observed (Y)
Give anticoagulant	Hemorrhage?
Personalized discount	Buy?
Target job training	Employed in 6 months?
Homelessness prevention program	Re-enter?
Recidivism prevention program	Recidivate?
Support for minority CS students	Drop out?

Often We Want to Predict **Response**

Treatment (T)	Individual Label of Interest ($Y(1) - Y(0)$)
Give anticoagulant	Hemorrhage iff medicated
Personalized discount	Would buy iff discounted
Target job training	Would get job iff trained
Homelessness prevention program	Re-enter iff not targeted
Recidivism prevention program	Recidivate iff not targeted
Support for minority CS students	Drop out iff not targeted

Classifying Responders: The Problem

- Each unit consists of
  - Features X
  - Potential outcomes $Y(1), Y(0) \in \{0, 1\}$
- "Non-responder" has $Y(0) = Y(1)$
  - Would've bought (or, not bought) *regardless* of discount
  - Would've hemorrhaged (or, not) *regardless* of anticoagulant
- "Responder" has $Y(1) = 1 > 0 = Y(0)$
  - Would've bought *if and only if* offered discount
  - $R = \mathbb{I}\left[Y(1) > Y(0)\right]$
  - Ground truth *NOT* observed in X, T, Y data
- Want classifier $f : \mathcal{X} \to \{0, 1\}$ with small loss

$$
\begin{aligned}
L_\theta(f) &= \theta \mathbb{P}\left(\text{false positive}\right) + (1-\theta)\mathbb{P}\left(\text{false negative}\right) \\
&= \theta \mathbb{P}\left(f(X) = 1, R = 0\right) \\
&\quad + (1-\theta)\mathbb{P}\left(f(X) = 0, R = 1\right).
\end{aligned}
$$

Monotonicity

- **Monotone treatment response assumption**:

$$Y(1) \geq Y(0)$$

  - Discount never causes a would-be buyer to *not* buy
  - Job training never causes someone to *not* get employed?

Monotonicity

- **Monotone treatment response assumption**:

$$Y(1) \geq Y(0)$$

  - Discount never causes a would-be buyer to *not* buy
  - Job training never causes someone to *not* get employed?
- Under monotonicity, $R = Y(1) - Y(0) \in \{0, 1\}$
  - So,

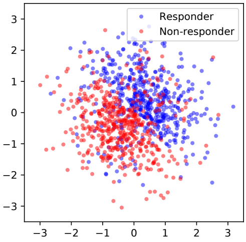
$$\mathbb{P}(R = 1 \mid X) = \tau(X) = \mathbb{E}[Y(1) - Y(0) \mid X]$$

  - $f(X) = \mathbb{I}[\tau(X) \geq \theta]$ minimizes $L_\theta(f)$
  - Can take plug-in approach using any CATE estimator $\hat{\tau}$
  - Question: any value to a direct classification approach?

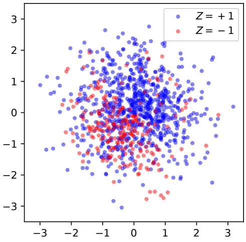
Classifying Responders

- For simplicity, consider completely randomized data with $\mathbb{P}(T = 1) = 0.5$
- Let $Z = \mathbb{I}[Y = T]$ (observable!)
  - $R = 1 \implies Z = 1$
  - $R = 0 \implies Z \sim \text{Bernoulli}(0.5)$
- Z is like a corrupted observation of R
  - Seeing $Z = 0$ is **more informative** about R
- Using Z as a surrogate label for R leads to new direct approaches to the classification problem
  - Two instantiations of this are RespSVM, RespNet

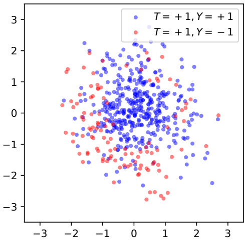
Empirical Results: Synthetic



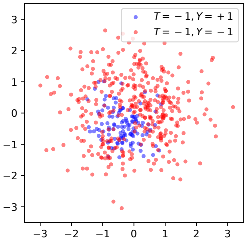
The true label R



The observable label Z

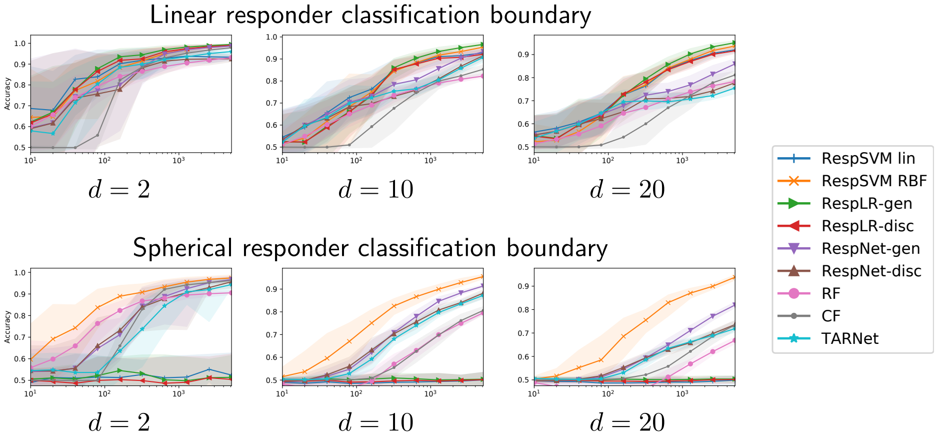


$T = +1$



$T = 0$

Empirical Results: Synthetic



Empirical Results: Census Data

- Predict whether the sex-at-birth of mother's first two kids being the same influences her decision to have a third
  - Follows data construction by Angirst & Evans '96
  - Covariates: ethnicity of mother and father; their ages at marriage, at census, at 1st kid, and at 2nd kid, year of marriage, and education level

Method	L_θ (in 0.01)	% 1st	% 2nd	% 3rd
RespSVM lin	49 ± 2.7	100%		
RespLR-gen	57 ± 2.4		100%	
RespLR-disc	58 ± 2.3			2%
LR	58 ± 2.3			92%
RF	58 ± 2.3			6%

Thank you!

Poster: Today 6:30pm @ Pacific Ballroom #74