Data-efficient causal effect estimation

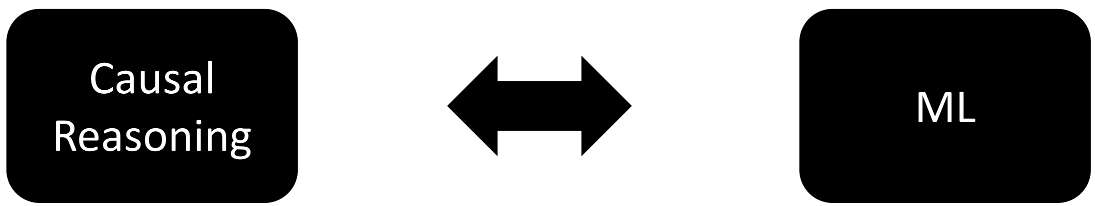
Adith Swaminathan
adswamin@microsoft.com

Joint work with Maggie Makar (MIT)
and Emre Kıcıman (MSR AI)

Brown TRIPODS
1.16.2019

Microsoft Research AI

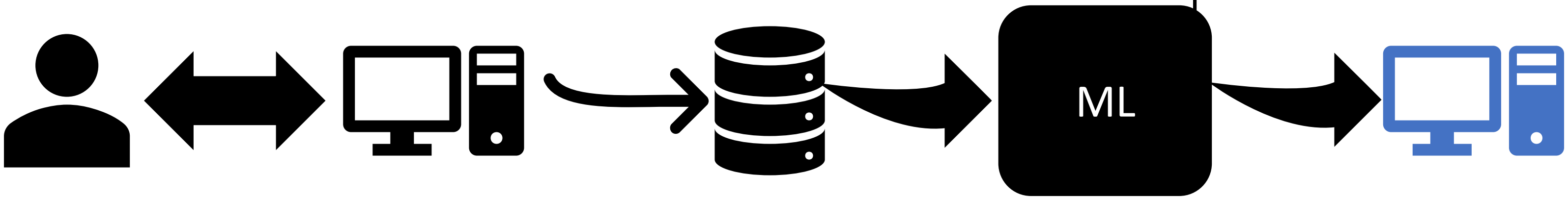
1. Improve ML applications using Causal Reasoning



2. Use ML tools to perform Causal Inference

1. Causal Reasoning -> ML

"Use logs collected from interactive systems to evaluate and train new interaction policies"

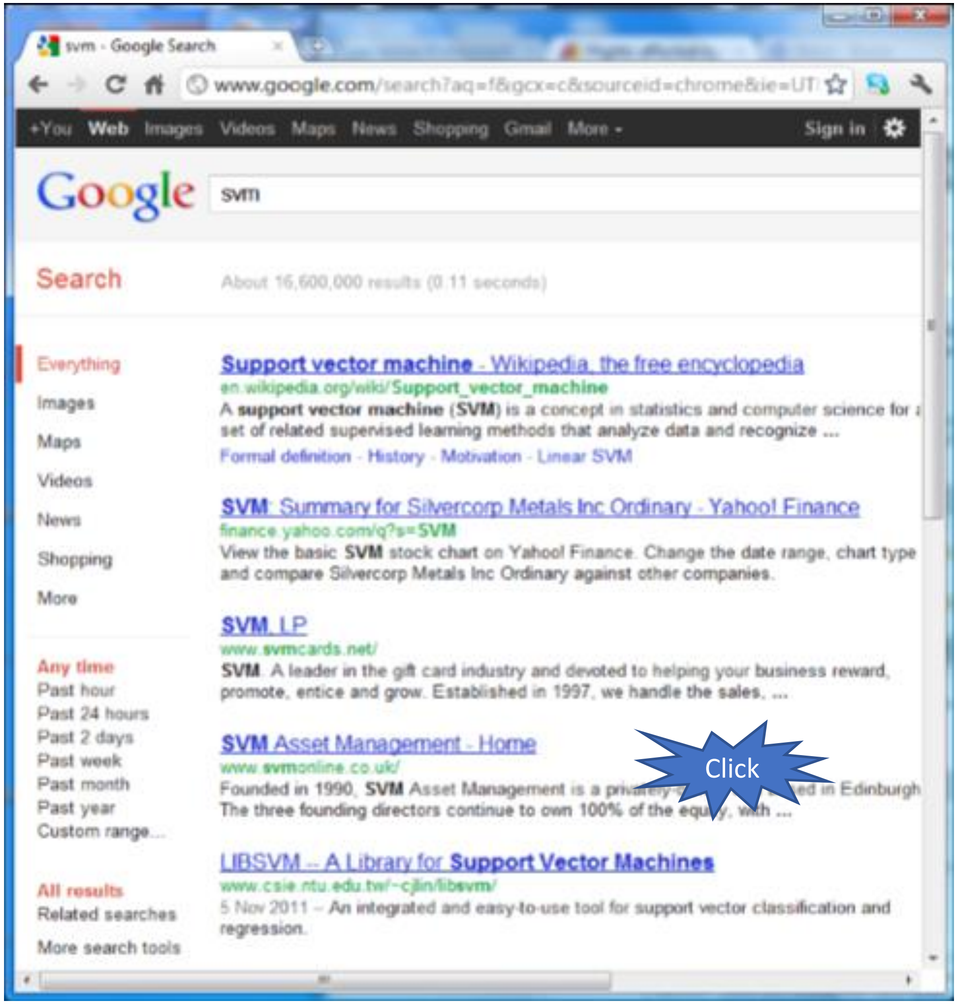


The data we collect from interactive systems…

Simple pragmatic fixes to address confounding!

Example: Search *[https://arxiv.org/abs/1608.04468; WSDM'17]*

Model the propensity of clicks on documents to de-bias training set of learning-to-rank models



Pointers to recent results

"IPW fixes collaborative filtering for recommendations"

[Schnabel et al,ICML'16]

"Similar IPW-like ideas massively improve learning-to-rank for search"

[Joachims et al,WSDM'17 ***Best Paper****]*

"Important to reason about variance of IPW for counterfactual learning"

[Swaminathan&Joachims,ICML'15]

"We can do much better than IPW for structured treatments (slates)"

[Swaminathan et al,NIPS'17]

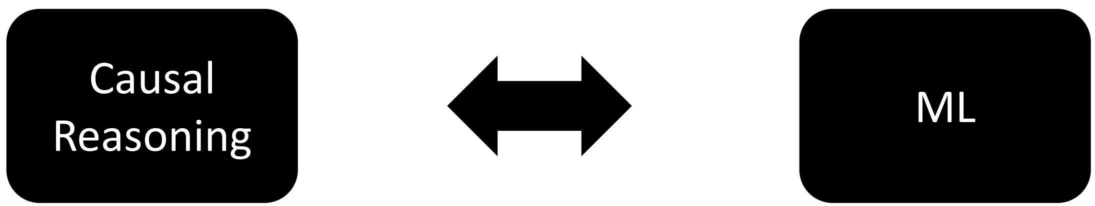
"Self-normalized estimators are better to use in these applications"

[Swaminathan&Joachims,NIPS'15]

"These techniques complement deep learning"

[Joachims et al,ICLR'18]

1. Improve ML applications using Causal Reasoning



2. Use ML tools to perform Causal Inference

2. ML -> Causal Reasoning

"Data efficient treatment effect estimation"

[AAAI'19]

Representation learning + Causal inference =
Bias-Variance Trade-off?

Problem Setting

Will my patient's blood pressure increase if I put her on medication A?

Challenges

- A question of *causal* nature
- *Limited data* at test time

Individual Treatment Effect (ITE)

- Estimate the causal effect of an intervention: if t changes, how does the outcome Y_t change?

- Target for estimation: $Y_1 - Y_0$

- Target is unobserved: the fundamental problem of causal inference

$$ITE: \tau(x) = \mathbb{E}_{Y_1 \sim \Pr(Y_1|x)}[Y_1] - \mathbb{E}_{Y_0 \sim \Pr(Y_0|x)}[Y_0]$$

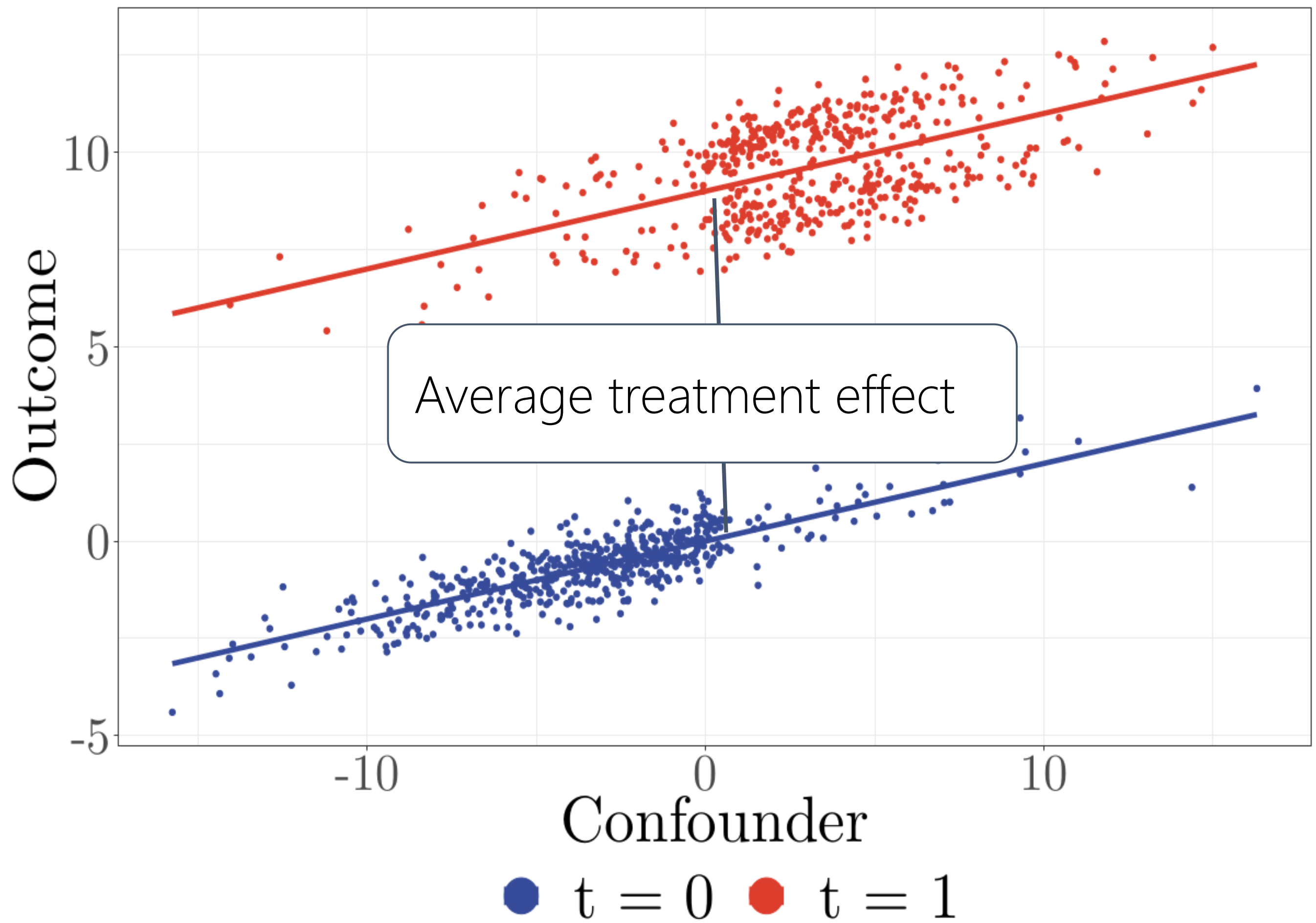
ITE estimation from obs. data

Two functions:

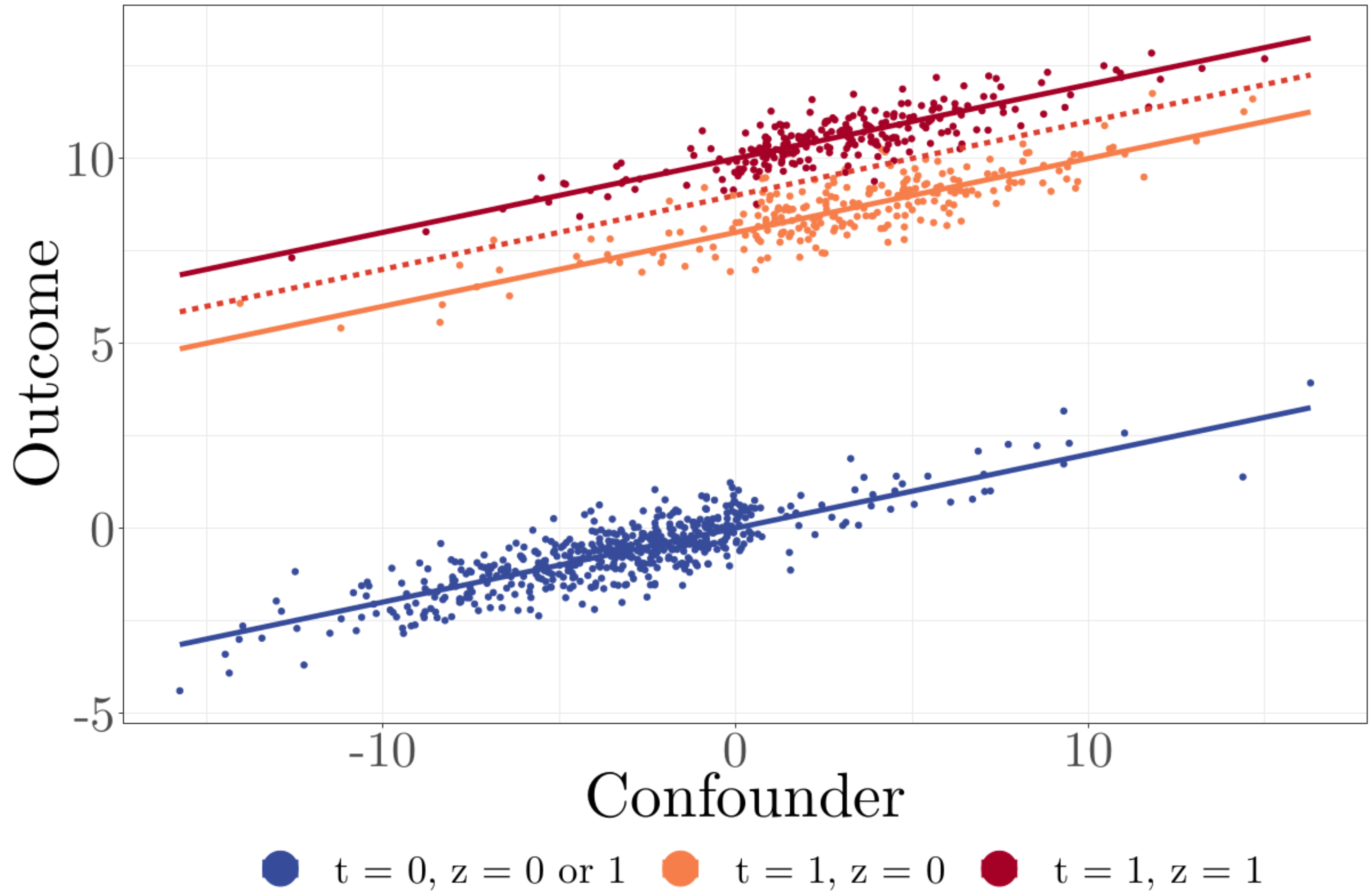
Adjustment for Confounding ← { Confounders }

Estimation of heterogeneity ← { Effect modifiers }

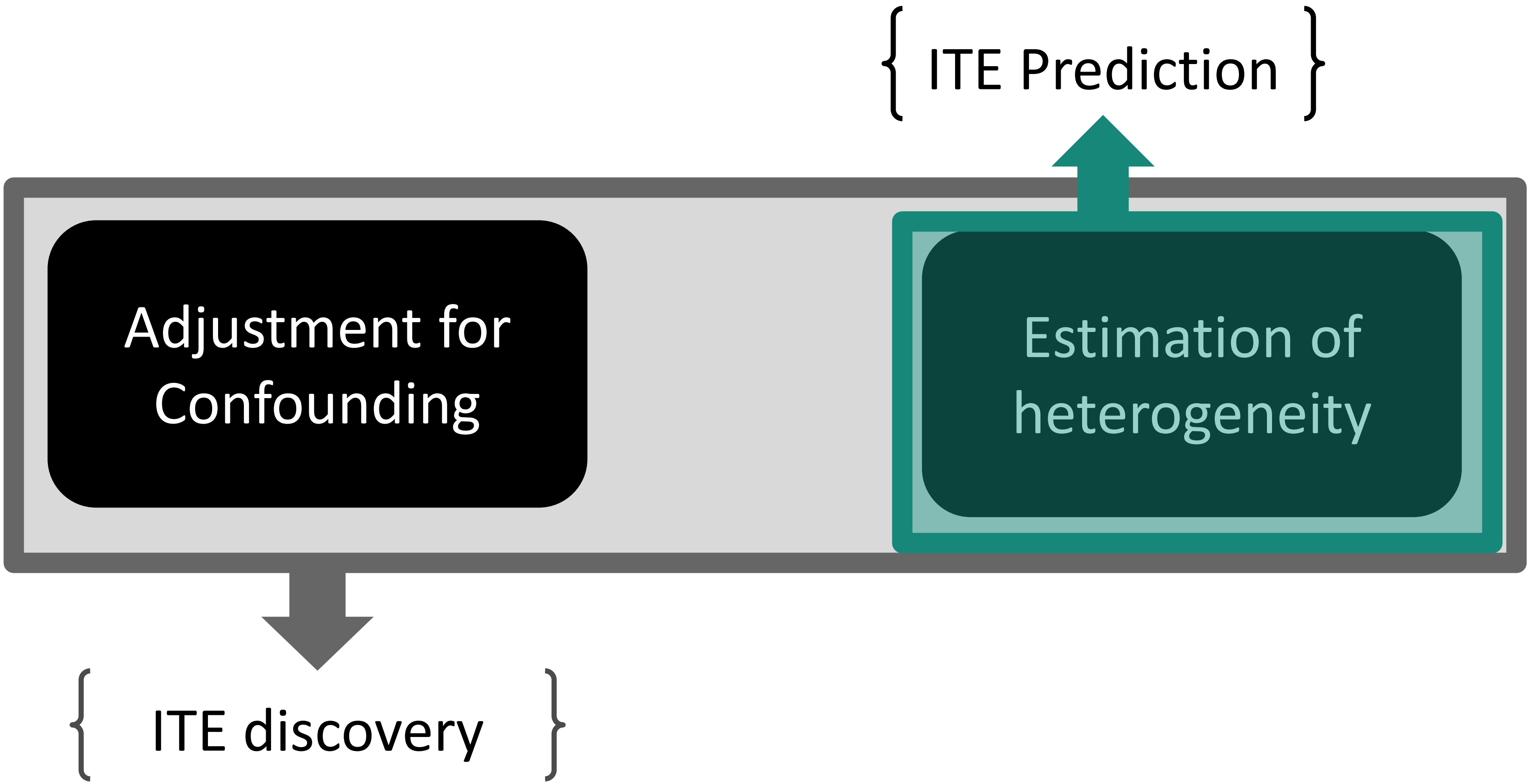
Confounders vs. Effect modifiers



Confounders vs. Effect modifiers

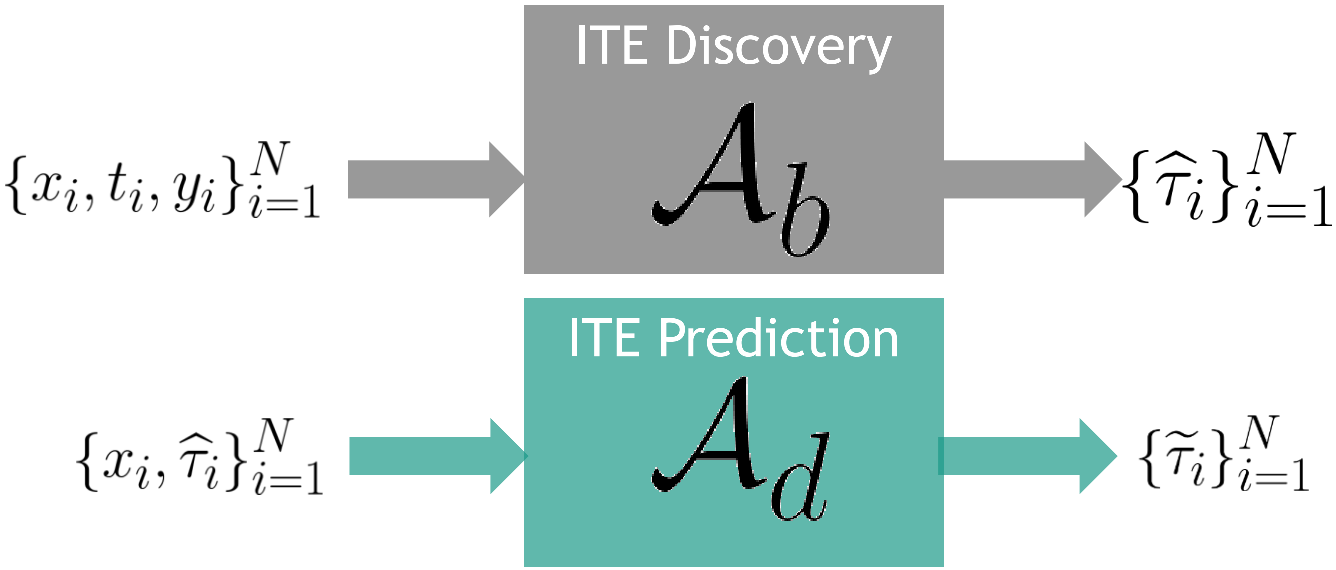


Data efficient ITE estimation

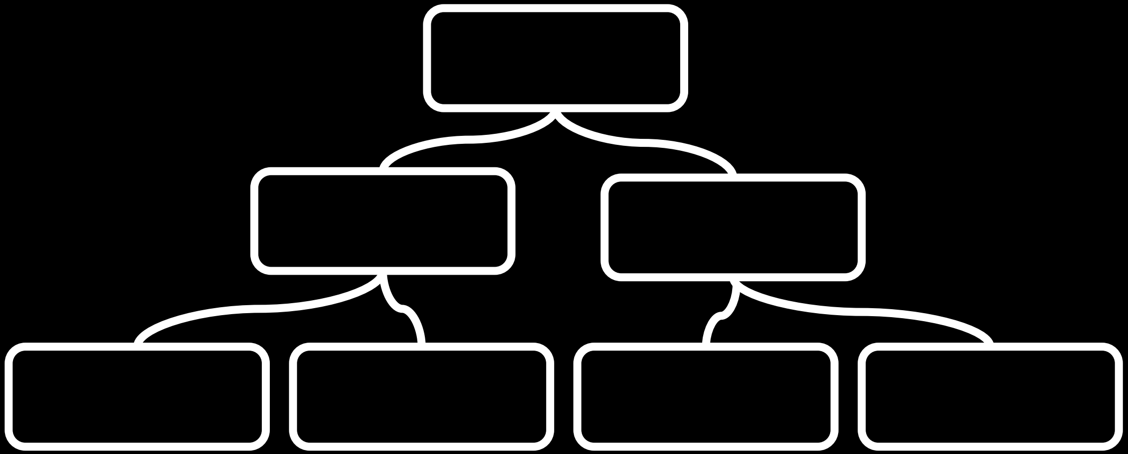


Insight

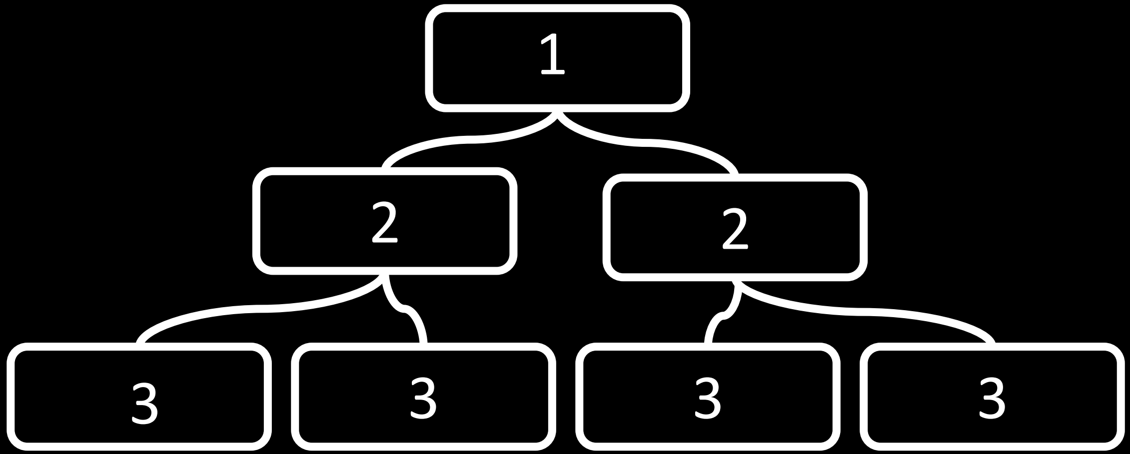
Leverage the *difference* between tasks at training and test time to reduce data collection burden at test time



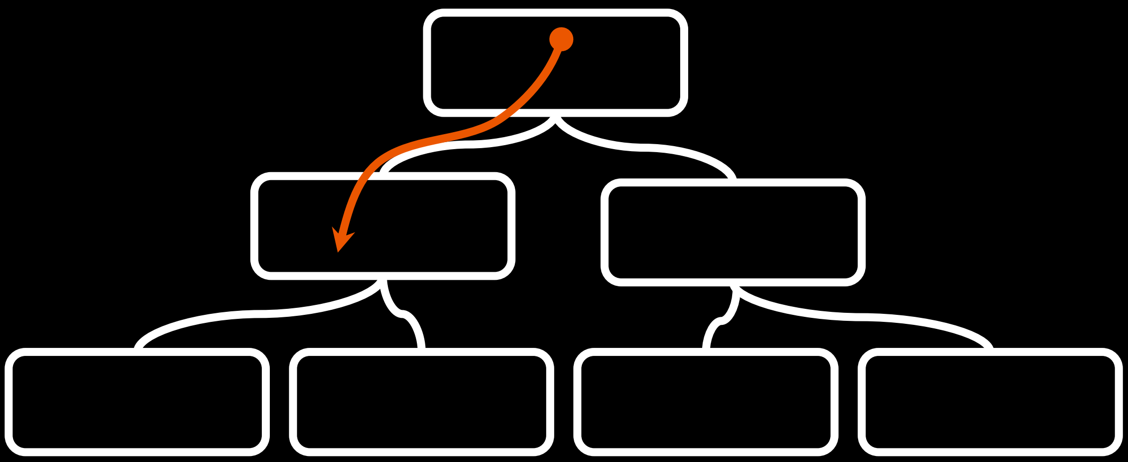
Why Trees?



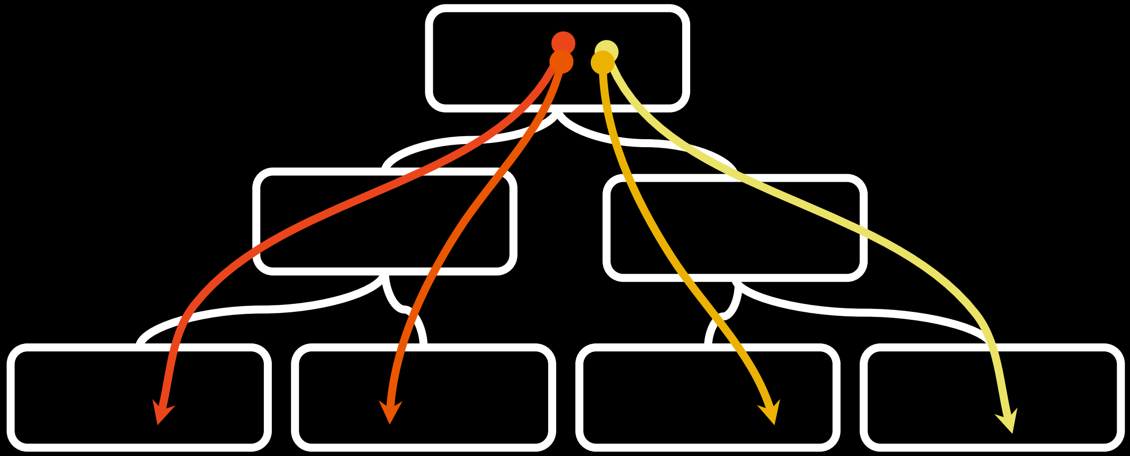
Trees identify the most important axes of heterogeneity



Trees can be traversed till querying ability is exhausted

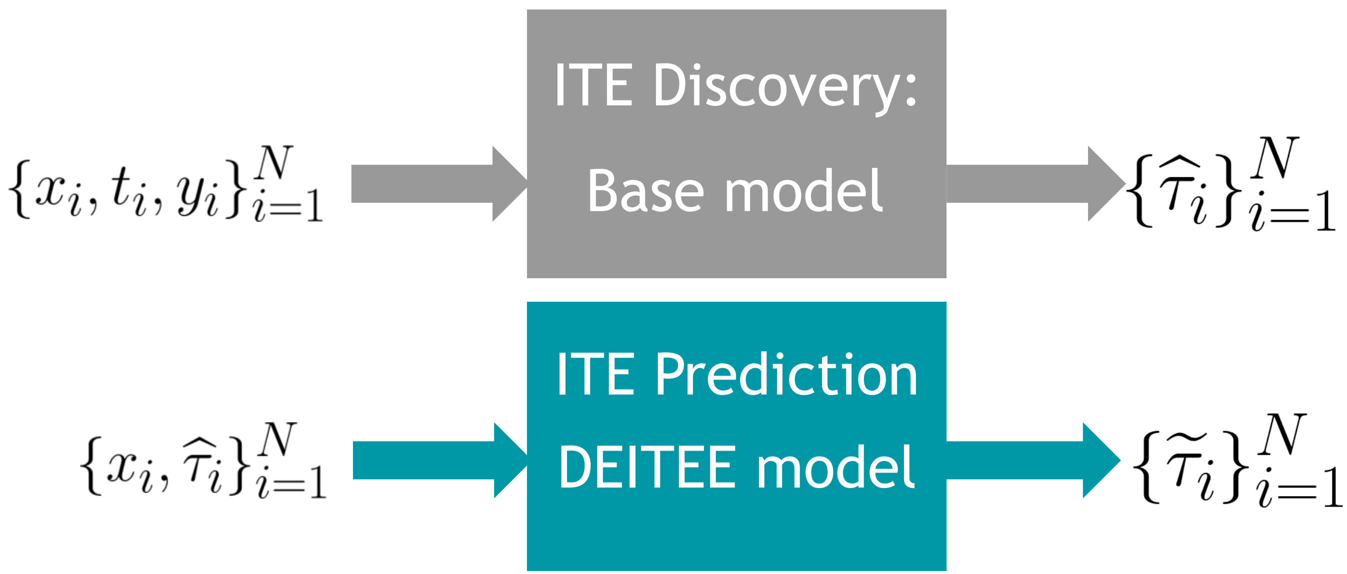


Different individuals → different queries



Algorithm: DEITEE

Data Efficient Individual Treatment Effect Estimator

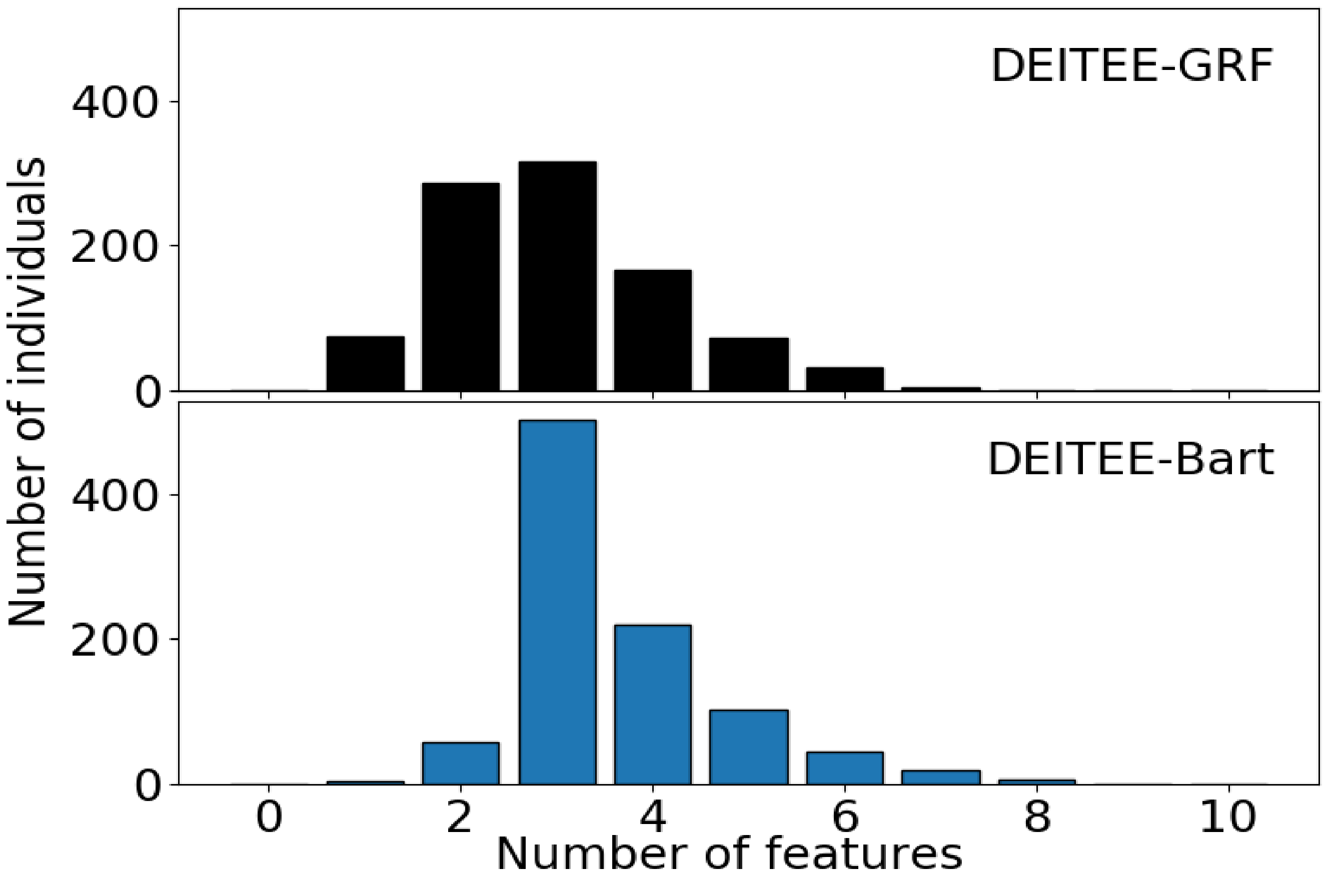


Experiments: Synthetic

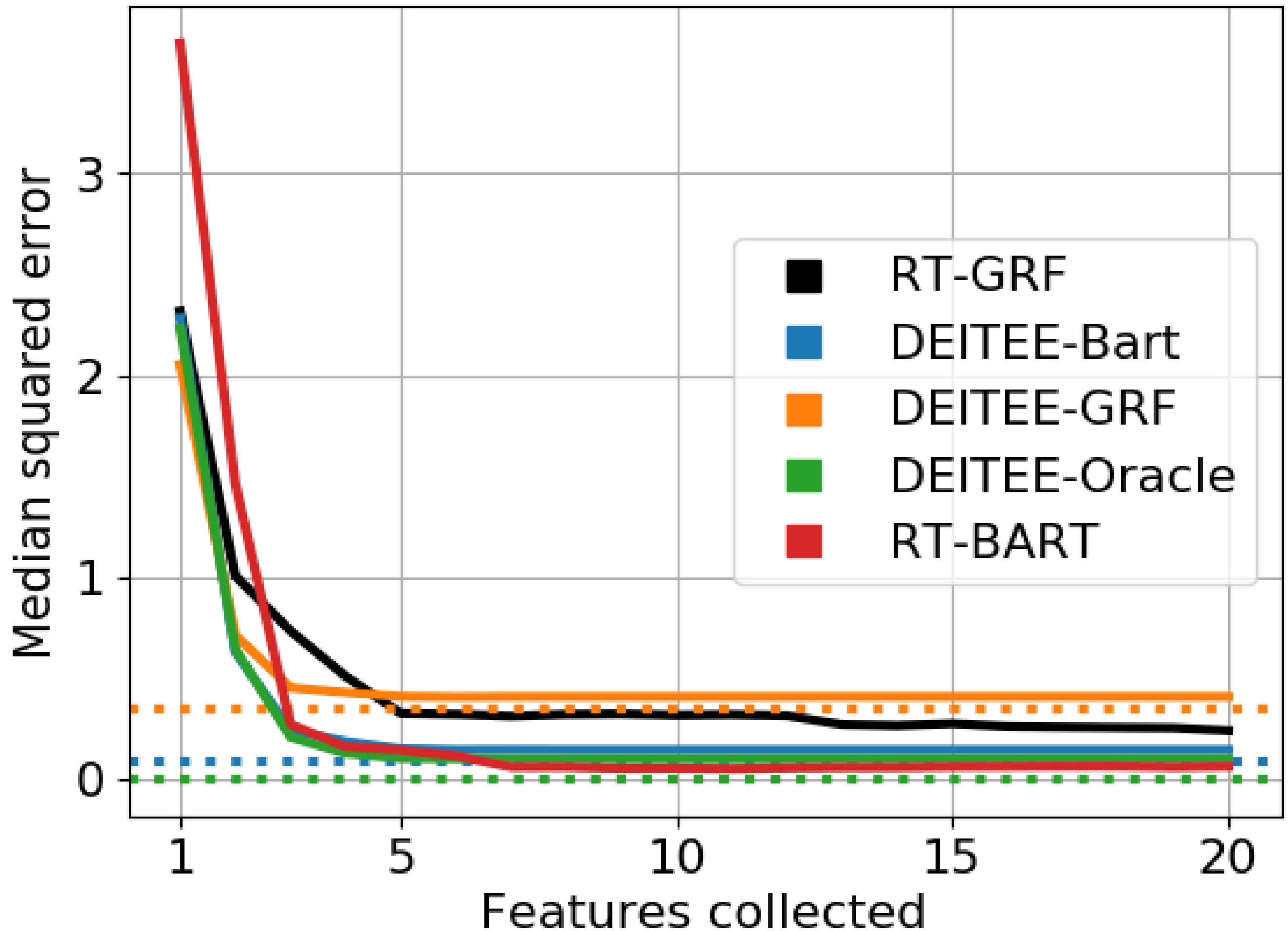
- Data: ACIC'17 simulated data ("semi-synthetic")
  N=5k; d=58
- Base models: BART and GRF
- Benchmarks: Train BART/GRF with feature regularization

- Evaluation: (1) Accuracy relative to true ITE; (2) Number of features queried

DEITEE: Features queried



DEITEE doesn't sacrifice accuracy

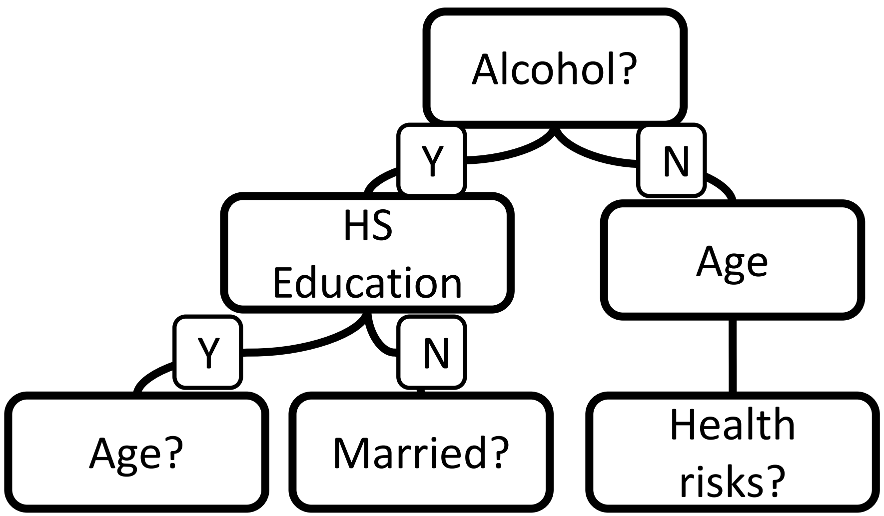


Experiment on real data

What is the effect of mother's habits on newborn's health?

1989 MA singleton births (CDC) N=90k; d=77

Mother's habit (treatment)	Mean Absolute Error relative to proxy ITE		Mean number of DEITEE features
	BART	DEITEE-BART	
Prenatal care	580.20	580.20	15.42
Smoking	587.62	587.62	16.2



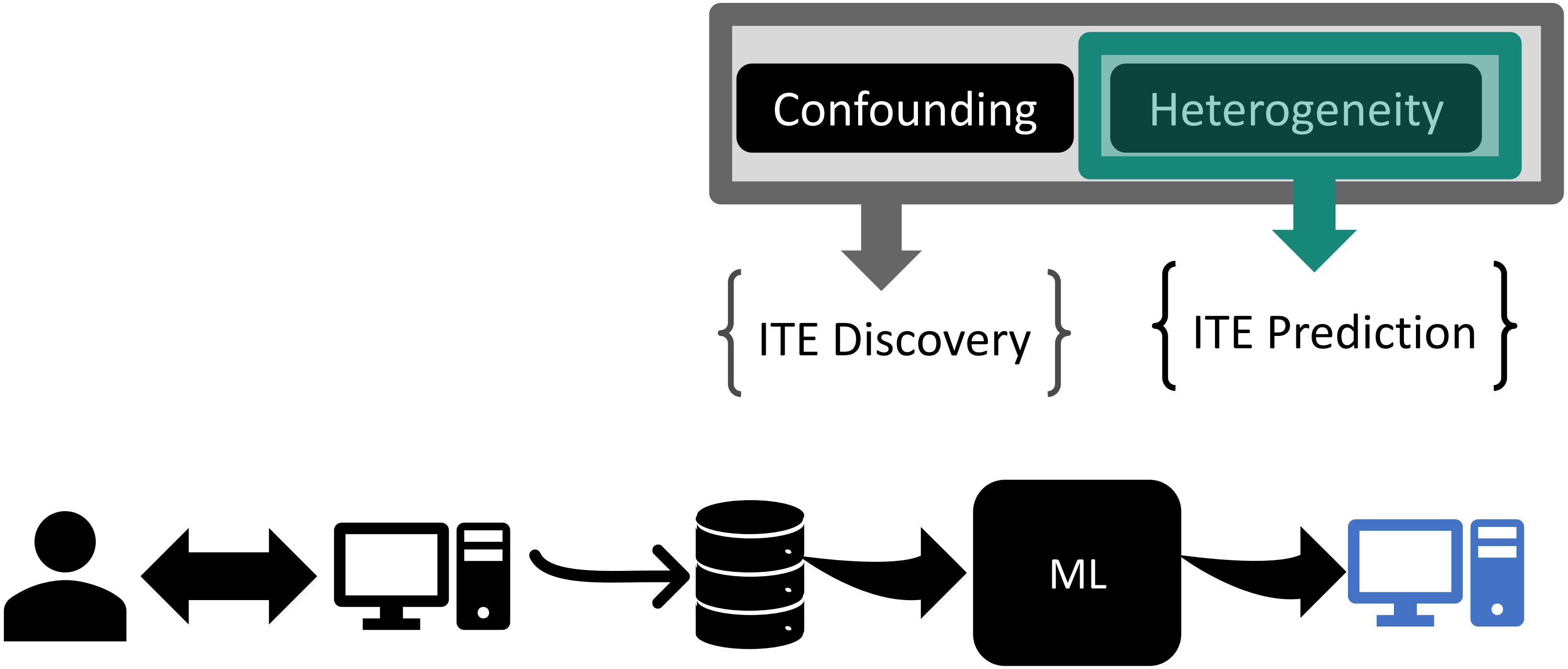
Conclusions

- DEITEE reduces the number of features required to estimate individual causal effects
    - ❖ Leverage difference between ITE discovery and ITE prediction

- Ongoing: Careful analysis of distillation error; guarantees on effect modifier discovery

- Need: Good robust method for model selection

Thanks!



adswamin@microsoft.com