Boosting for Predictive Sufficiency

Gowtham Reddy Abbavaram
CISPA Helmholtz Center for Information Security
Saarbrücken, Germany

Joint work with:

Rajeev Verma

Celia Rubio-Madrigal

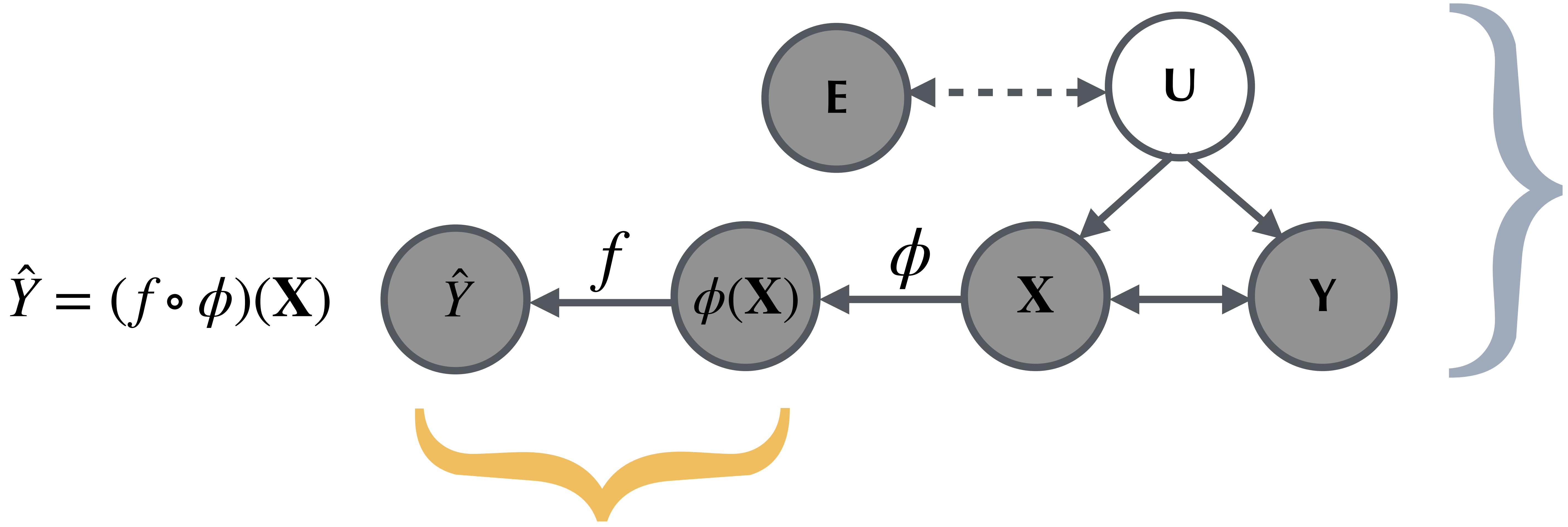
Krikamol Muandet

Rebekka Burkholz

ICLR

CISPA
HELMHOLTZ CENTER FOR
INFORMATION SECURITY

Causal Model



Data Generating Process

- ✳ $\mathbf{X}$ is the set of covariates
- ✳ Y is the outcome
- ✳ U is a continuous (hidden) confounder
- ✳ E is a discrete environment variable
- ✳ E encodes a shift in $\mathbb{P}(U)$ across envs.

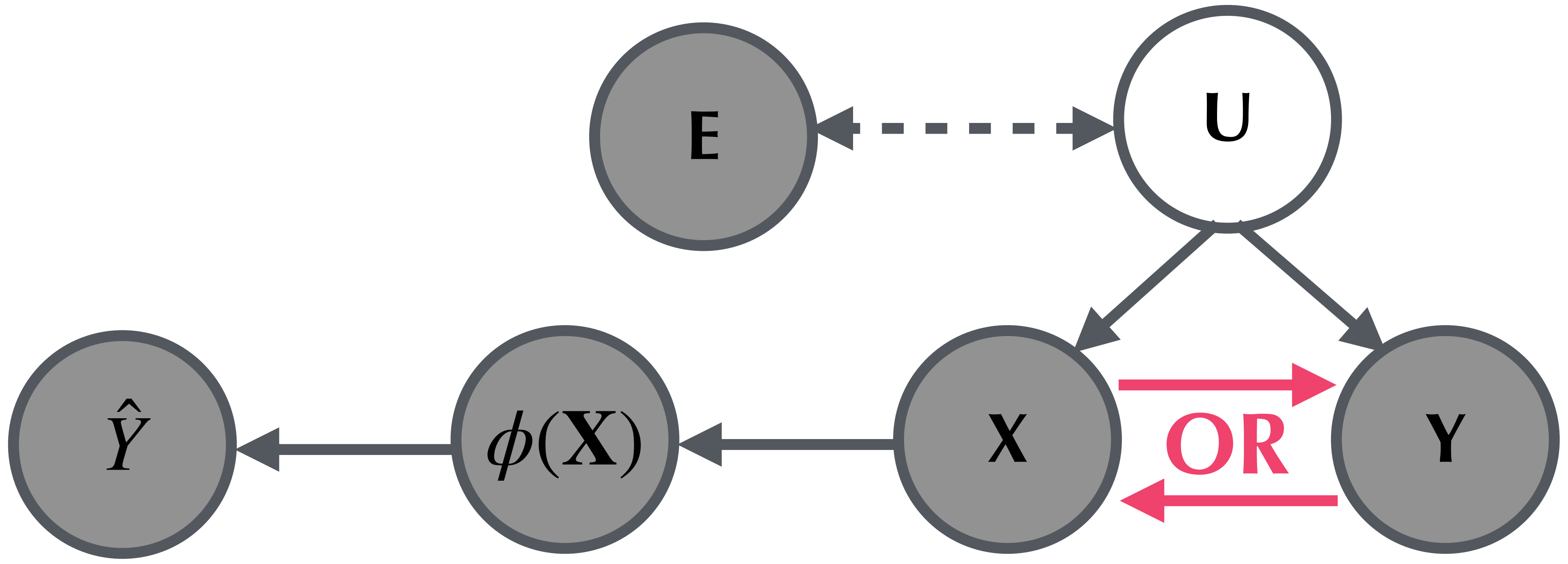
Computation Process

- ✳ $\phi(\mathbf{X})$ is a learned representation
- ✳ $\hat{Y}$ is the predicted Outcome

Goal

Achieve high predictive information $I(Y; \hat{Y})$

Predictive Information Under Hidden Confounding Shift[1]



$$I(Y;\hat{Y}) = \overbrace{I(\phi(\mathbf{X});Y\,|\,E)}^{\text{Cond. Inform.}} - \overbrace{I(\phi(\mathbf{X});Y\,|\,\hat{Y})}^{\text{Residual}}$$

Definition α - predictive sufficiency $\implies I(Y-\hat{Y};E \mid \hat{Y}) \leq \alpha$

Can Boosting achieve α-predictive sufficiency?

Proposition Predictive sufficiency and predictive information are related as:

$$I(Y-\hat{Y};E \mid \hat{Y}) = -I(Y;\phi(\mathbf{X}) \mid E,\hat{Y}) + I(Y;\phi(\mathbf{X}) \mid \hat{Y}) + I(Y;E \mid \phi(\mathbf{X}))$$

[1] Reddy, Abbavaram Gowtham, et al. "When shift happens-confounding is to blame." *International Conference on Learning Representations. 2026*

Standard Boosting Algorithm

Algorithm 1 Standard boosting algorithm

Require: Step size η, base predictor $\hat{Y}_0$, hypothesis class $\mathcal{H}$, reweighting rule to obtain $\mathcal{D}_t$

1: Initialize $\hat{Y}_0 \leftarrow h_0, t \leftarrow 0$ ▷ h_0 is often a constant predictor
2: **while** training error decreases **do**
3: Find weak learner $h_{t+1} \in \mathcal{H}$ that achieves $I(Y; h_{t+1}(\mathbf{X})) \geq \gamma$ under distribution $\mathcal{D}_t$
4: Update predictor: $\hat{Y}_{t+1} \leftarrow \hat{Y}_t + \eta h_{t+1}(\mathbf{X})$
5: Update distribution $\mathcal{D}_{t+1}$ using the reweighting rule and increment t by 1.
6: **end while**

Assumptions

∗ Assume h_t satisfies: $(i) \quad I(Y; h_t) \geq \gamma' \quad (ii) \quad I(Y; h_t \mid h_0 \ldots h_{t-1}) \geq \gamma$

∗ Assume strong learner is a deterministic function of weak learners

∗ $I(U; \hat{Y}) \geq c \cdot I(Y; \hat{Y}); \quad c > 0$

Boosting for Predictive Sufficiency

Theorem

$\exists \; T < \infty$ such that $\hat{Y}_t$ learned by a boosting algorithm after $t \geq T$ rounds is α-predictive

$$T = \frac{H(Y) - H(Y \mid \mathbf{X}, E) - \alpha - I(Y; \hat{Y}_0)}{p \cdot \gamma}$$

If E is unknown:

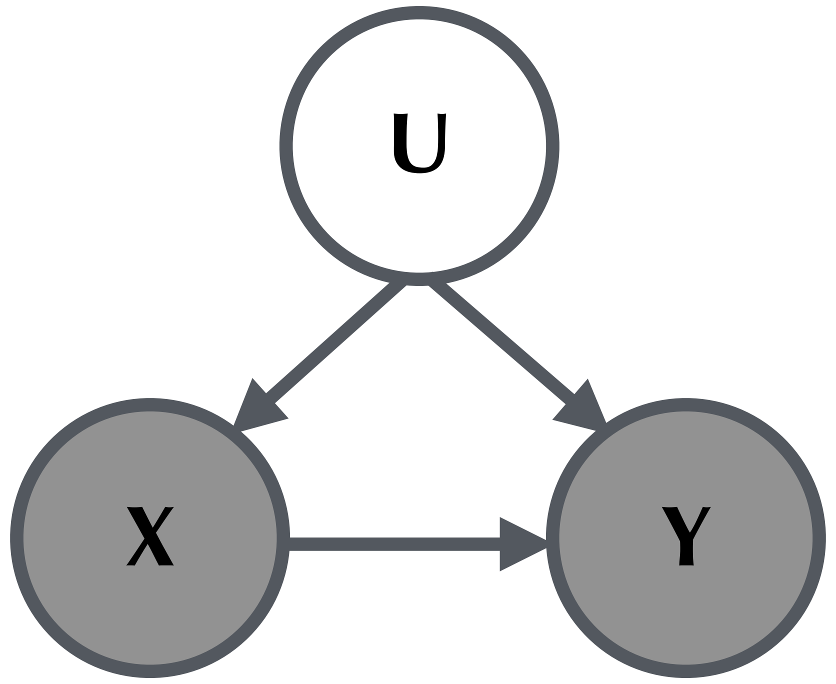
$$T = \frac{H(Y) - \alpha - I(Y; \hat{Y}_0)}{p \cdot \gamma}$$

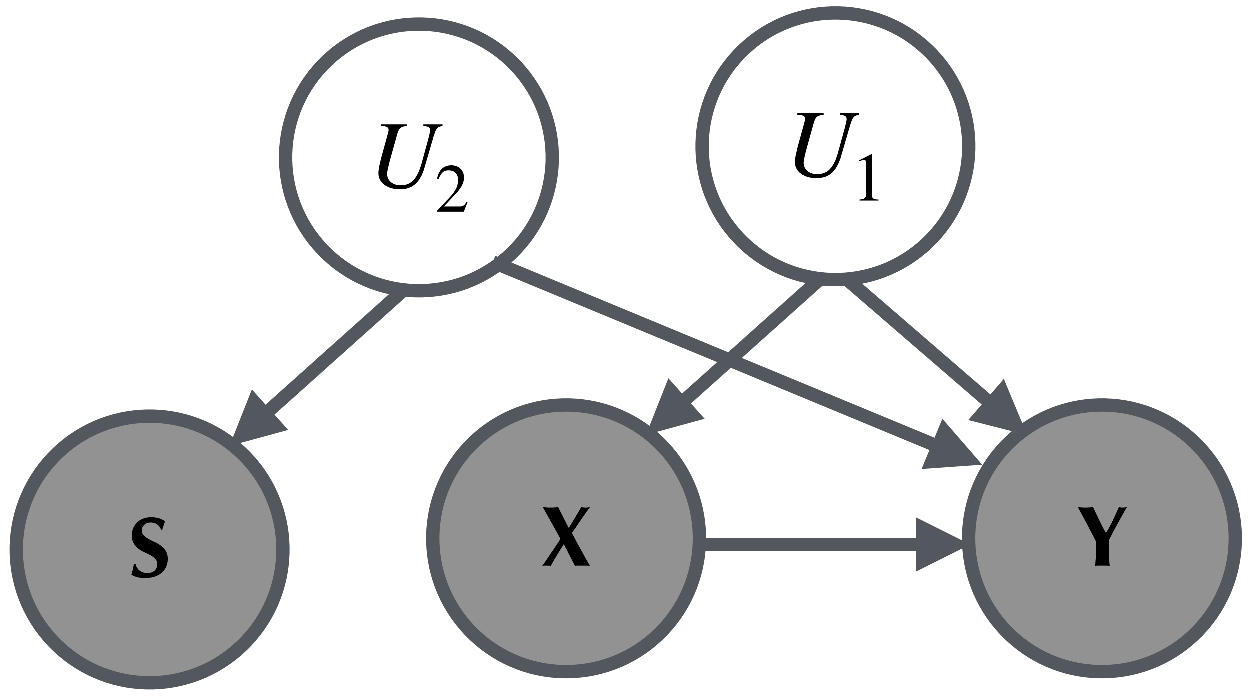
Corollary

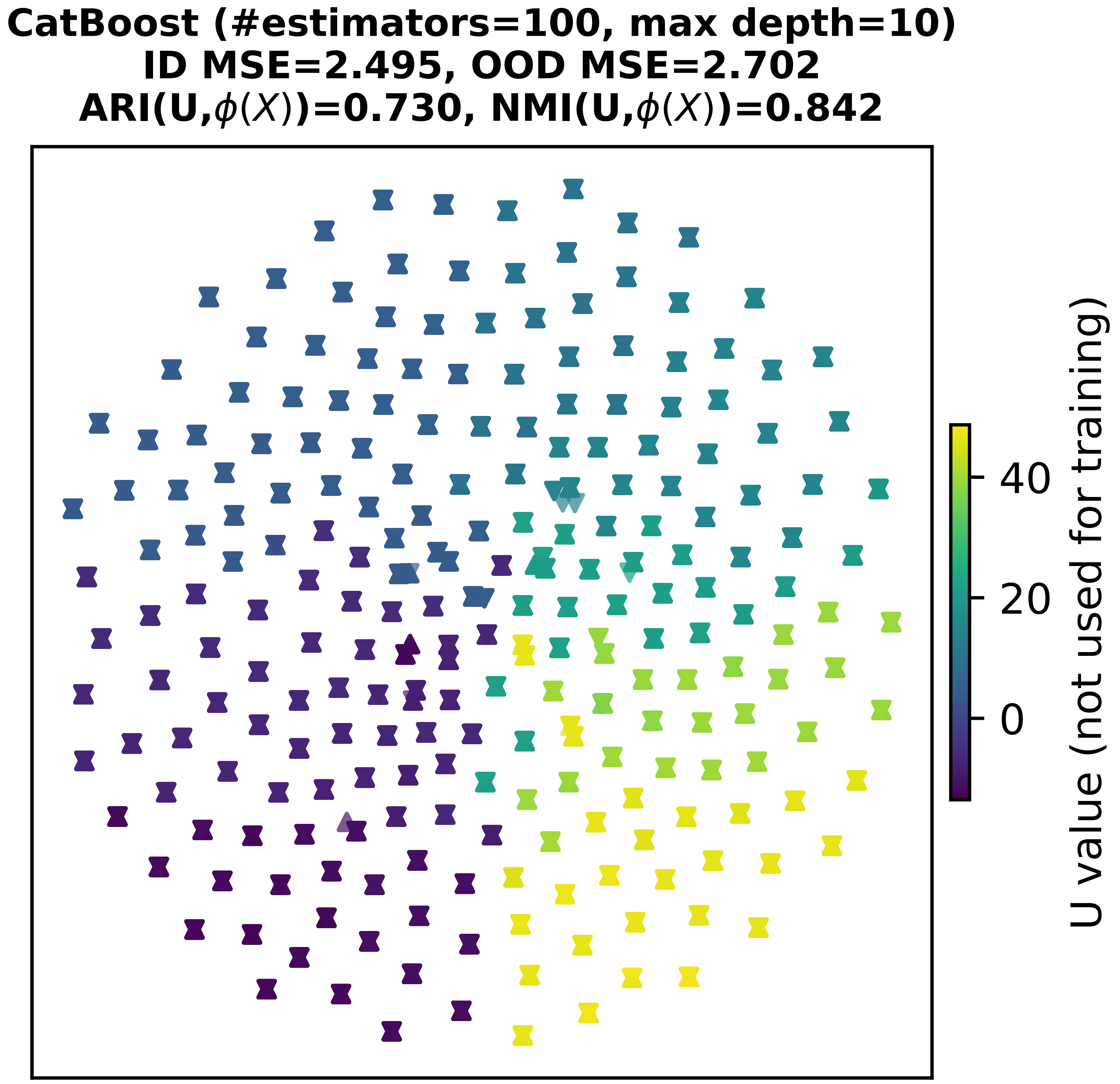
$\exists \; T < \infty$ such that $\hat{Y}_t$ learned by a boosting algorithm after $t \geq T$ rounds satisfies $H(U \mid \hat{Y}_t) \leq \delta$

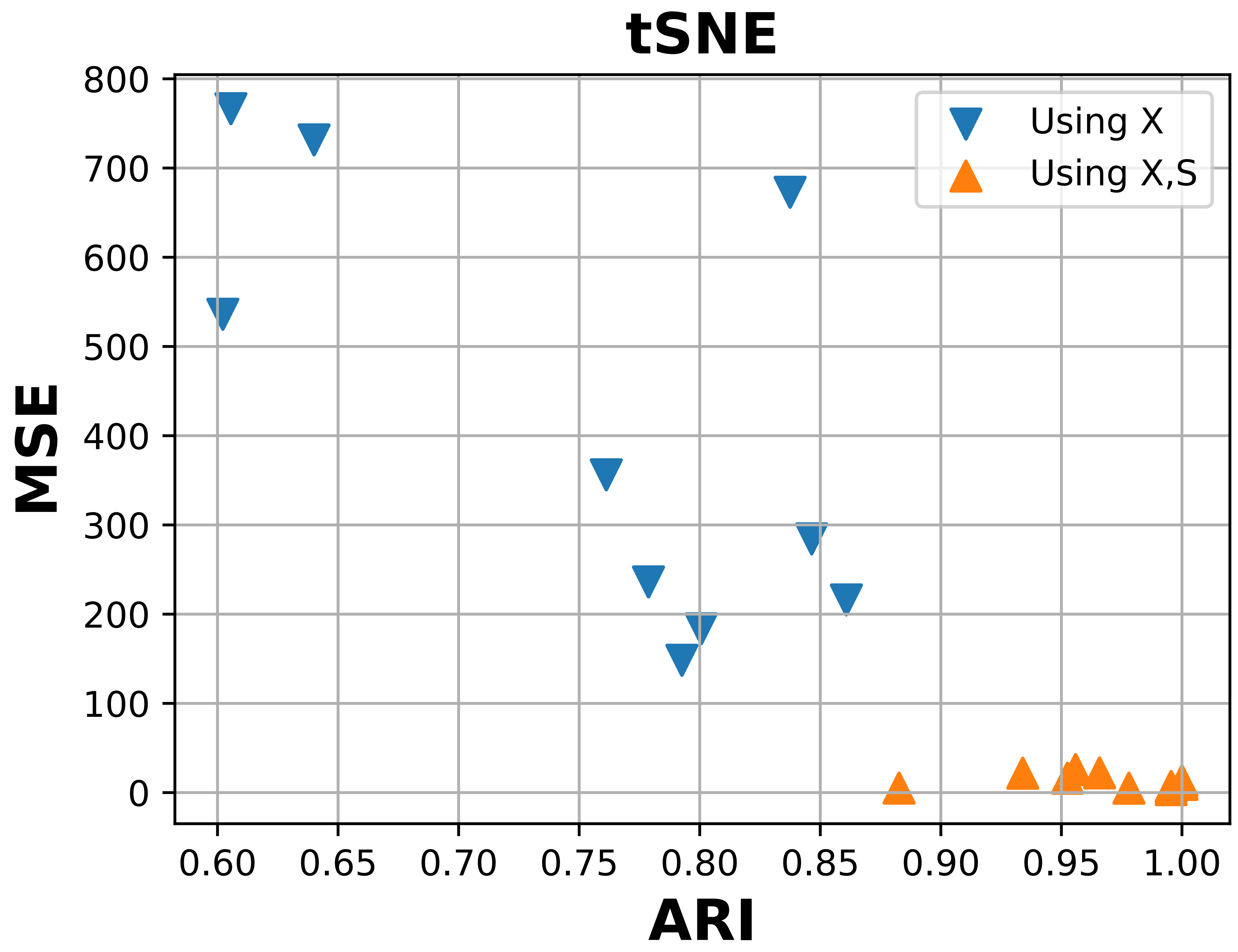
$$T = \frac{H(U) - \delta - c \cdot I(Y; \hat{Y}_0)}{c \cdot p \cdot \gamma}$$

Synthetic Data Results: Boosting for Predictive Sufficiency



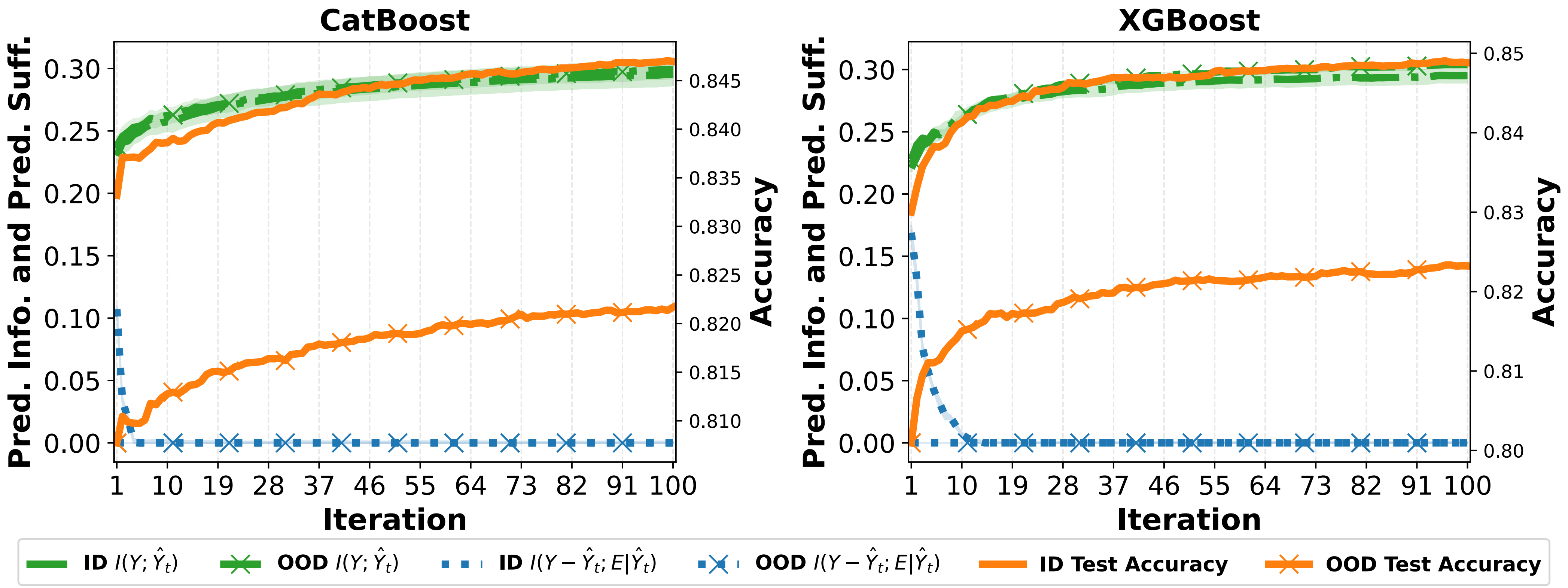




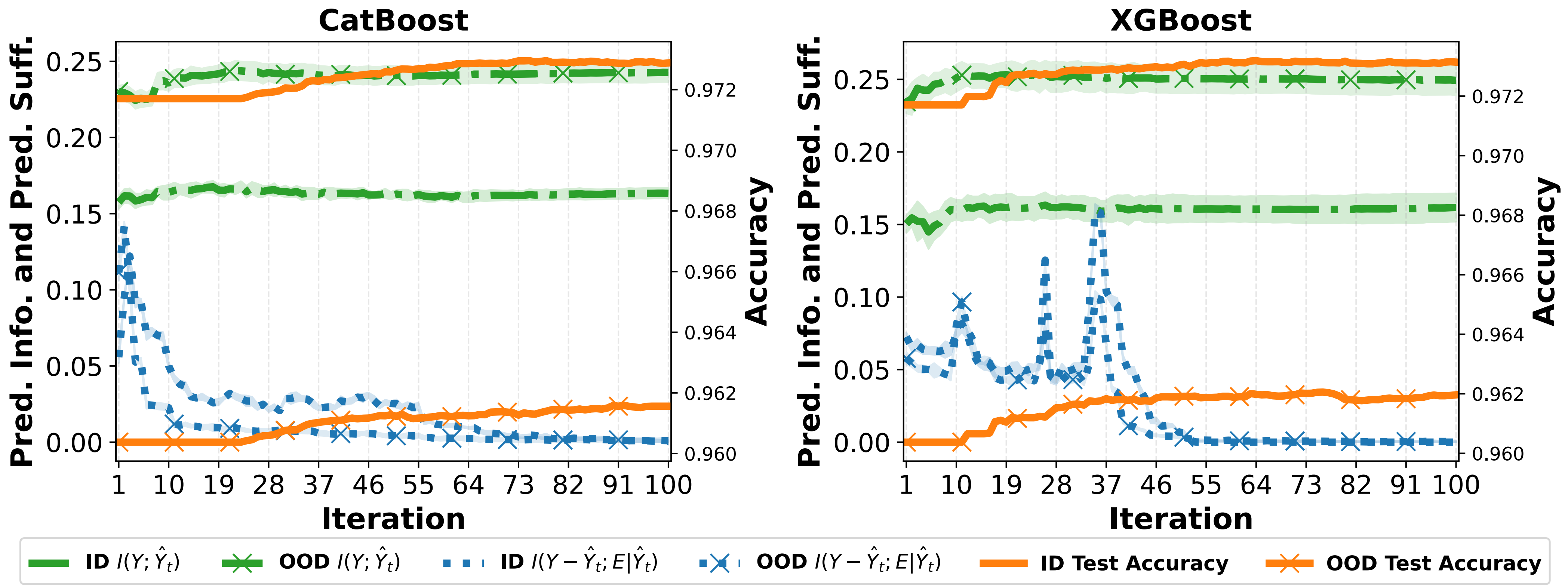


Real world Data Results: Boosting for Predictive Sufficiency

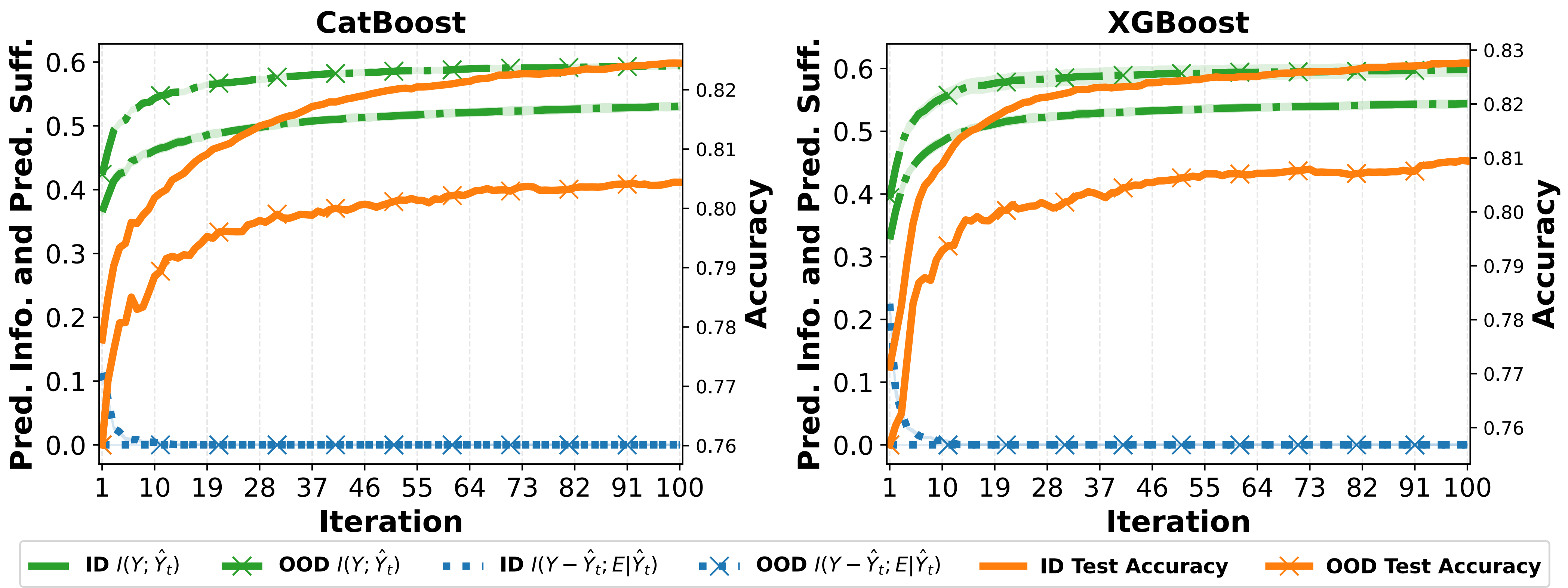
Foodstamps



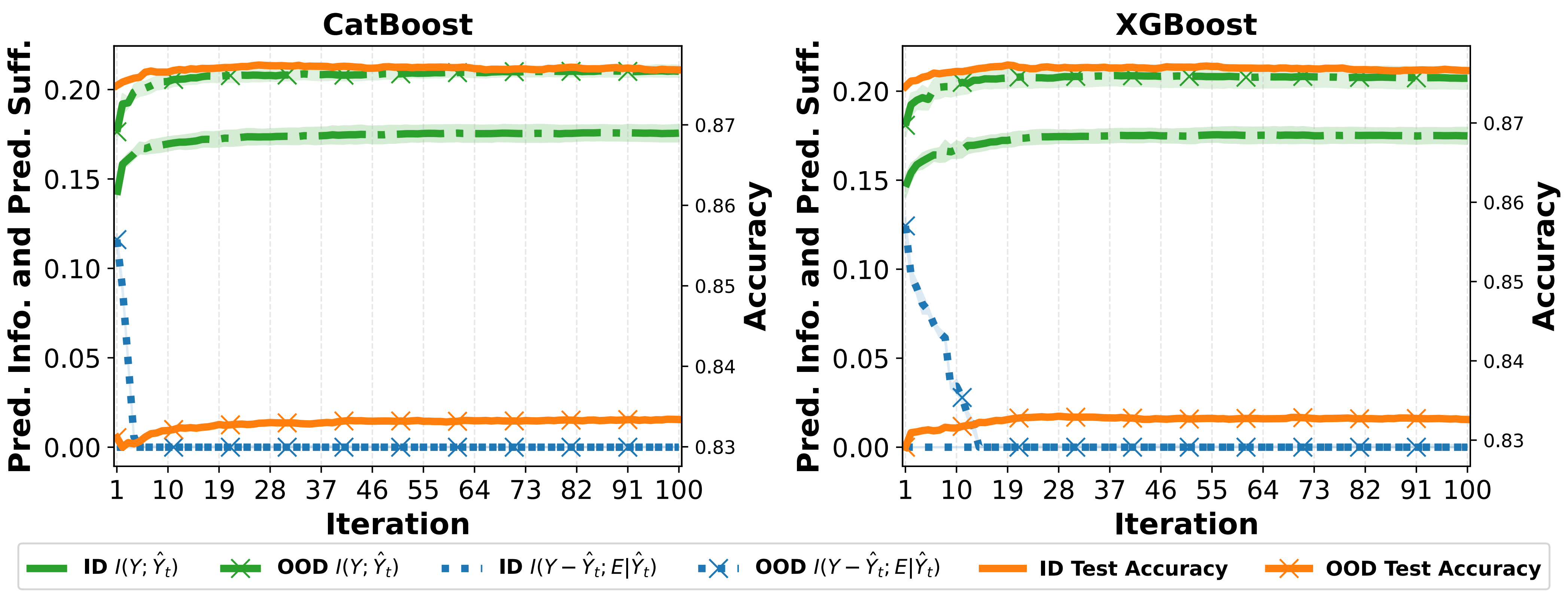
Unemployment



Income



Diabetes



Readmission