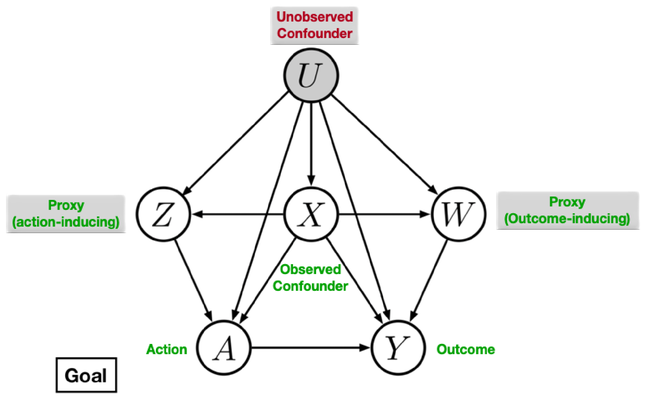
Proximal Causal Learning with Kernels: Two-Stage Estimation and Moment Restriction

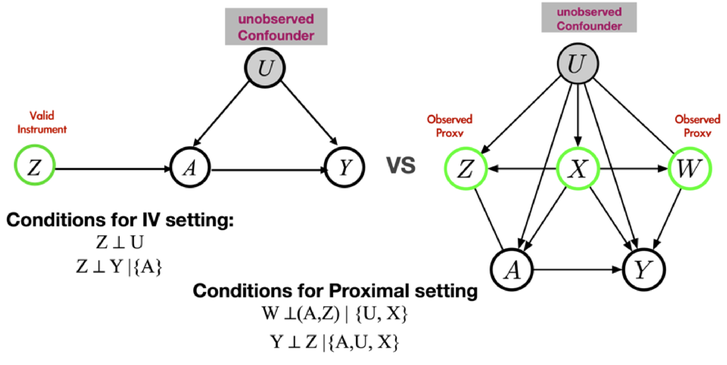
Afsaneh Mastouri* [1] Yuchen Zhu* [1] Limor Gultchin [2,4] Anna Korba [5] Ricardo Silva [1] Matt J. Kusner [1]
Arthur Gretton† [1] Krikamol Muandet† [3]

[1]University College London [2]University of Oxford [3]Max Planck Institute for Intelligent Systems [4]The Alan Turing Institute [5]ENSAE/CREST

Proximal Causal Learning



Proximal Setting vs. Instrumental Variable Setting



The Proximal Problem

$$\mathbb{E}[Y|A, X, Z] = \int_{\mathcal{W}} h(A, X, W)\, dF(W|A, X, Z)$$

Learn h, solution to above Fredholm integral equation. It follows:

$$\mathbb{E}[Y|do(A) = a] = \mathbb{E}_{X,W}\, h(a, X, W)$$

Contributions

Kernel Algorithms: KPV + PMMR

1 Proposed two kernel-based algorithms related by duality of the loss objective; derived their connection and demonstrated our methods can be applied to a more general class of inverse problems that involve a solution to a Fredholm integral equation.

2 Derive convergence guarantees for two proposed algorithms.

Algorithm 1: Two-stage Kernel Proxy Variable (KPV)

Stage 1 Learn expected value of feature of W as function of A, X, Z via Kernel Ridge Regression: $\mu_{W|A,X,Z} := \mathbb{E}[\phi(W) \mid A, X, Z]$

Stage 2 $\hat{\boldsymbol{h}} = \arg\min_h \mathbb{E}[(Y - \langle \boldsymbol{h}, \psi(A, X) \otimes \mu_{W|A,X,Z} \rangle_{\mathcal{H}_{AXW}})^2]$ via KRR

Flexibility h can be inferred from sub-samples A, X, W and A, X, Z

Algorithm 2: Proximal Maximum Moment Restriction (PMMR)

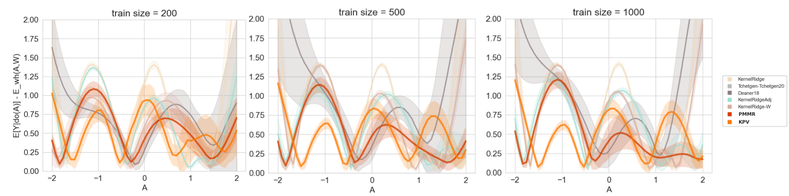
A single-step solution

Transfer minimising the discrepancy between the two sides of (1) into working with a weighted regression objective:

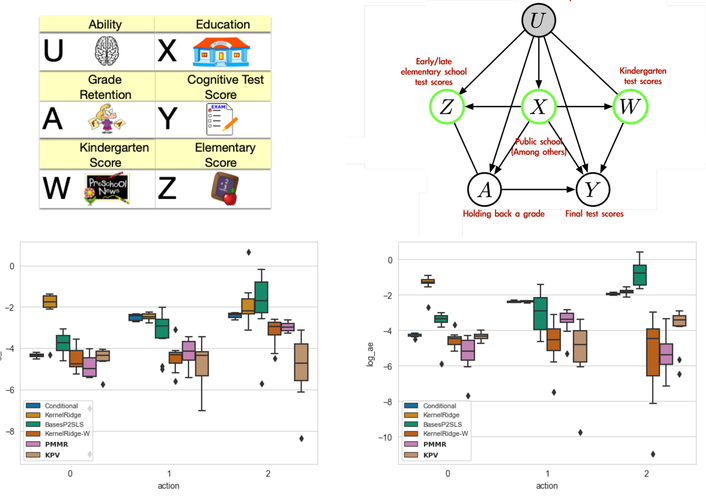
$$\min_h \mathbb{E}[(Y - h(A, W, X))(Y' - h(A', W', X'))k((A, Z, X), (A', Z', X'))]$$

Empirical Results

Example1: Synthetic Simulation



Example 2: Grade Retention on Cognitive Outcome



References

[1] Wang Miao, Zhi Geng, and Eric J Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987--993, 2018.

[2] Eric J Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao. An introduction to proximal causal learning. *arXiv preprint arXiv:2009.10982*, 2020.