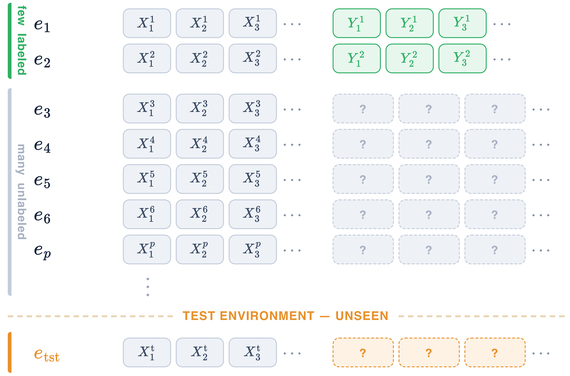
Anti-causal domain generalization: Leveraging unlabeled data

Sorawit (James) Saengkyongam†, Juan L. Gamella, Andrew C. Miller†, Jonas Peters‡, Nicolai Meinshausen‡ and Christina Heinze-Deml† †Apple, ‡ETH Zürich

Core Contributions

- **Semi-supervised domain generalization.**
  We formalize anti-causal domain generalization in a semi-supervised setting.
- **Robustness under anti-causal learning.**
  Under anti-causal learning, the environment shifts that affect the covariates do not propagate to the outcome. Moreover, the directions of these shift can be estimated from **unlabeled** data alone.
- **Data-driven robust regression.**
  We propose two regularization strategies MIR and VIR that exploit distributional variations in the unlabeled data to encourage robustness.
- **Worst-case optimality guarantees.**
  We show that the proposed estimators provably minimize worst-case risk over a class of environments whose size is controlled by the regularization strength.
- **Real-world evaluations.**
  We evaluate our methods on two naturally anti-causal datasets: a controlled physical system and a physiological signal prediction task.

Semi-supervised Domain Generalization

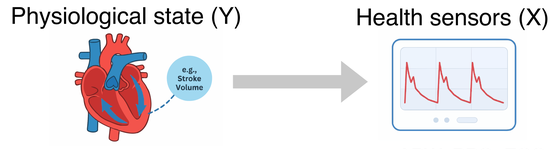


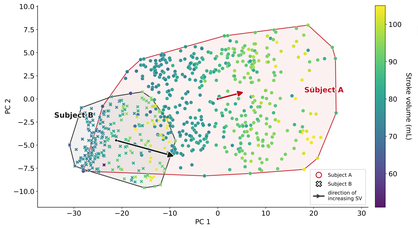
$$\mathcal{L} = \{e_1, \dots, e_k\} \subset \mathcal{E}^{tr} = \{e_1, \dots, e_p\}, \qquad e_{tst} \notin \mathcal{E}^{tr}$$

Labeled data are available from a few environments, and **unlabeled** data are available from many others

Goal: predict Y from X in an unseen environment

Motivating Example in ML for Health



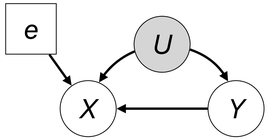


Each subject is one environment

Covariate shift
$\mathbb{P}_X^{s_A} \neq \mathbb{P}_X^{s_B}$

Concept shift
$\mathbb{E}^{s_A}[Y|X] \neq \mathbb{E}^{s_B}[Y|X]$

Anti-causal Learning



$$U := \varepsilon_U$$
$$S(e): \; Y := g_0(U, \varepsilon_Y)$$
$$X := f_0(Y, U, \varepsilon_X) + \varepsilon_e$$

Perturbations

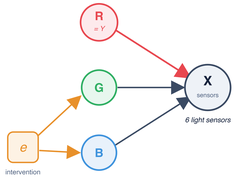
Covariate shift
$\mathbb{P}_X^{e_1} \neq \mathbb{P}_X^{e_2}$
$(X \perp\!\!\!\perp E)$

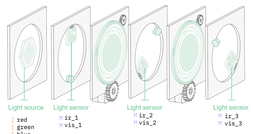
Concept shift
$\mathbb{E}^{e_1}[Y|X] \neq \mathbb{E}^{e_2}[Y|X]$
$(Y \perp\!\!\!\perp E | X)$

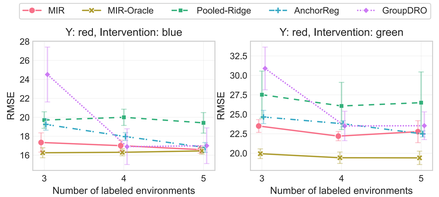
Invariance: $\forall e \in \mathcal{E} : \varepsilon_e \perp\!\!\!\perp Y$

Idea: learn the variations between $\varepsilon_{e_1}, \dots, \varepsilon_{e_p}$ and penalize in those directions of variations

Empirical Results: Causal Chamber (Light Tunnel)







- Pool-Ridge: Ridge regression on data pooled across training environments
- MIR-Oracle: MIR when tuning γ directly on the test environment
- AnchorReg (Rothenhäusler et al., 2021) • GroupDRO (Sagawa et al., 2020) • DRIG (Shen et al., 2026)

Data-driven Robust Regression

Capture variations between environment-perturbations

$$\beta_H^\gamma = \underset{\beta}{\operatorname{argmin}} \, \mathbb{E}^{tr}[(Y - \beta^\top X)^2] + \gamma \beta^\top H \beta$$

Mean-based Regularization (MIR)

$$H_{MIR} = \mathrm{Var}(K) = \frac{1}{p} K K^\top - \left(\frac{1}{p} K \mathbf{1}_p\right)\left(\frac{1}{p} K \mathbf{1}_p\right)^\top$$

where, $K := \left[\mathbb{E}^{e_1}[X] \quad \cdots \quad \mathbb{E}^{e_p}[X]\right]$

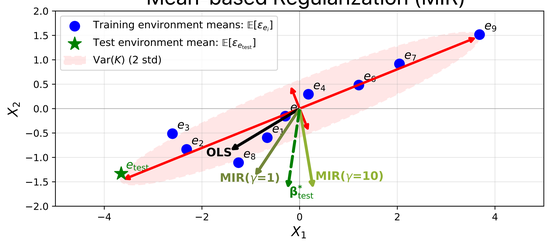
(Co)variance-based regularization (VIR)

$$H_{VIR} = \frac{1}{p} \sum_{i=1}^{p} (G_X^{e_i} - \bar{G}_X)^2$$

where, $G_X^e := \mathrm{Cov}^e(X)$ and $\bar{G}_X := \frac{1}{p} \sum_{j=1}^{p} G_X^{e_j}$

Distributional Robustness

Mean-based Regularization (MIR)

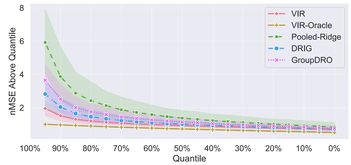


$$\beta_H^\gamma = \underset{\beta}{\operatorname{argmin}} \, \mathbb{E}^{tr}[(Y - \beta^\top X)^2] + \gamma \beta^\top H \beta$$
$$\in \underset{\beta}{\operatorname{argmin}} \, \sup_{e \in \mathcal{E}_{H,\gamma}} \mathbb{E}^e[(Y - \beta^\top X)^2]$$

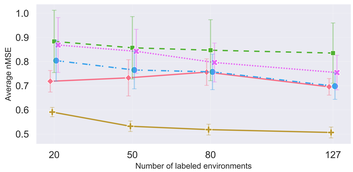
A class of perturbations defined by H and γ

Empirical Results: Stroke Volume Monitoring

Predicting **stroke volume** from arterial pressure waveform (**APW**) using data from the VitalDB dataset (Lee et al., 2022).



X : APW embeddings
Y : Stroke volume variation
e : Individual subjects (n=128)



Eval: Leave-one-subject-out

Metric: MSE normalized by subject's variance (nMSE)

- Lee, Hyung-Chul, et al. "VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients." Scientific Data 9.1 (2022): 279.
- Sagawa, Shiori, et al. "Distributionally Robust Neural Networks." In International Conference on Learning Representations, 2020.
- Rothenhäusler, Dominik, et al. "Anchor regression: Heterogeneous data meet causality." Journal of the Royal Statistical Society Series B: Statistical Methodology 83.2 (2021): 215-246.
- Shen, Xinwei, et al. "Causality-oriented robustness: Exploiting general noise interventions." Journal of the American Statistical Association (2026): 704-715.

TM and © 2026 Apple Inc. All rights reserved.