HellingerIEP:
Making Machine Learning More Reliable When Data Changes

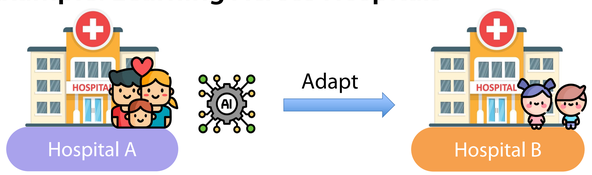
Tengyu Song · David Kepplinger · Anand N. Vidyashankar
Department of Statistics · George Mason University · tsong8@gmu.edu

TL;DR

A model trained on one population may perform poorly on another because the two populations differ. Contaminated or incorrect data can make this problem even worse. We develop a method that helps models adapt to new populations while reducing the influence of potentially misleading observations.

Introduction

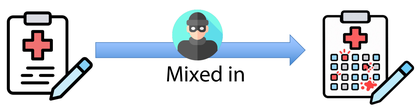
1 Example: Learning Across Hospitals



Past records,
Known outcomes
Mostly adults, some children

Patient details,
Outcomes unknown
Primarily children

2 New data can't be fully trusted



- Unusual cases
- Duplicated records
- Data errors

…

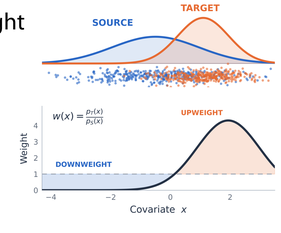
3 Why adaptation is needed

A model trained mostly on adults may struggle to predict outcomes for children because it cannot properly learn the distinguishing features. Domain adaptation improves predictions for the new population by strategically leveraging data from both.

Common strategy: Weighting Records

Weighting past samples

Patients at different hospitals may have different characteristics. We give more weight to past records that resemble new patients and less weight to those that do not.



Existing methods:
- KLIEP (minimizing KL divergence)
- uLSIF (minimizing least-squares/Pearson divergence)

Limitation: Not protected against contamination

Proposed Weighting: HellingerIEP

HellingerIEP: Our Approach

We estimate the density ratio by minimizing a variational Hellinger criterion:

$$J(w) = \mathbb{E}_{P_S}\left[\sqrt{w(X)}\right] + \mathbb{E}_{P_T}\left[w(X)^{-1/2}\right], \quad w : \mathcal{X} \to [0, \infty),$$

The weights are learned using an **efficient** optimization method. Unlike methods that trim unusual records, it does not require an extra setting for the proportion of records to remove.

Comparison with Other Density Ratio Estimation Methods

Method	Adaptivity	Robustness	Extra Robustness Tuning
Traditional Methods (KLIEP, uLSIF):	✓	✗	None
Existing robust Methods (Trimmed-KLIEP, etc.):	✓	✓	Required
HellingerIEP	✓	✓	None

Theoretical Results

1 Statistical Guarantees

For a fixed kernel dictionary and regularization level $\lambda > 0$, the estimator converges to the regularized population solution:

$$\hat{\beta} \xrightarrow{p} \beta^*_\lambda,$$
$$\sqrt{n}\left(\hat{\beta} - \beta^*_\lambda\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\text{asy}}\right).$$

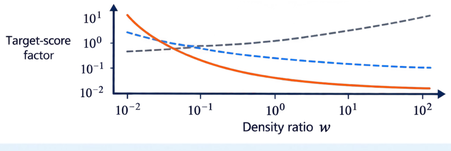
2 Why Our Approach Works: Less Sensitive to Unusual Records

Under a common square-root parameterization, the score contributions of large-ratio observations scale as shown below. Hellinger more strongly attenuates high-ratio regions than other methods.

HellingerIEP
Stronger damping

KLIEP
Less damping

uLSIF
May amplify

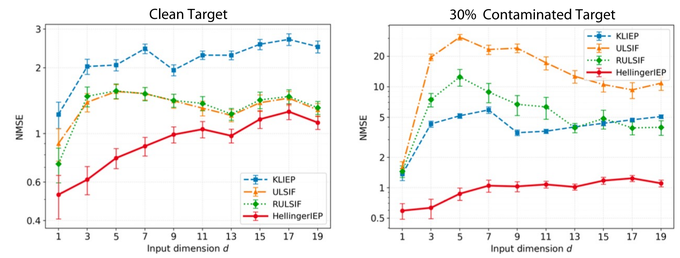


Takeaway
Robustness arises from divergence geometry: HellingerIEP provides stronger attenuation in high-ratio regions without requiring an extra robustness-specific tuning parameter.

Experimental Results

1 Can It Learn Weights Accurately from Messy Data?

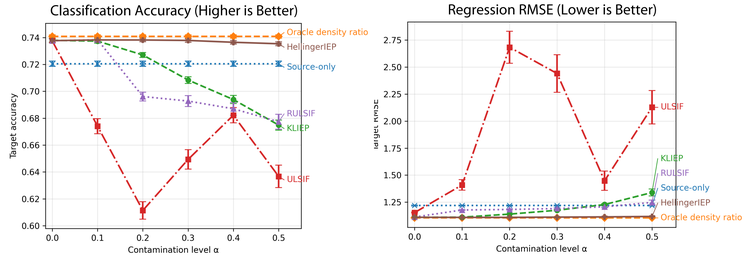
We measure density-ratio estimation error using normalized mean squared error (NMSE) when the target data are contaminated.



HellingerIEP is competitive in clean settings and also achieves the smallest NMSE under contaminated target covariates.

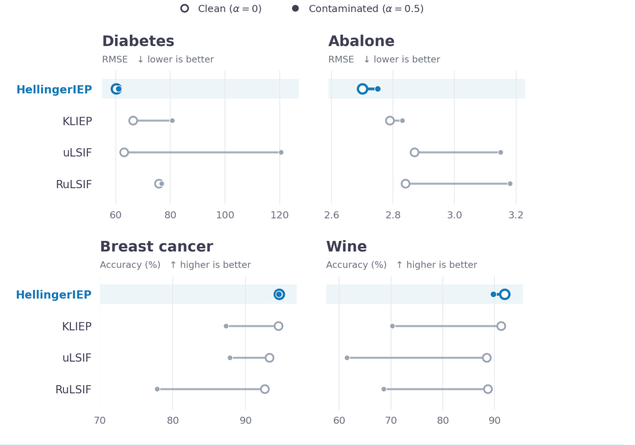
2 Do More Stable Weights Improve Predictions?

We train predictors using estimated importance weights and evaluate their predictive performance as target contamination increases.



Improved stability in weight estimation translates into more reliable downstream prediction.

3 Does It Work on Real-World Datasets?



Across benchmark datasets, HellingerIEP remains competitive in clean settings and degrades less under contamination.