Distributionally Robust Losses for Latent Covariate Mixtures

John Duchi[a], **Tatsunori Hashimoto**[b], **Hongseok Namkoong**[c,*]

[a]Departments of EE & Statistics, Stanford University
[b]Department of Computer Science, Stanford University
[c]Decision, Risk, and Operations Division, Columbia Business School
*Corresponding author

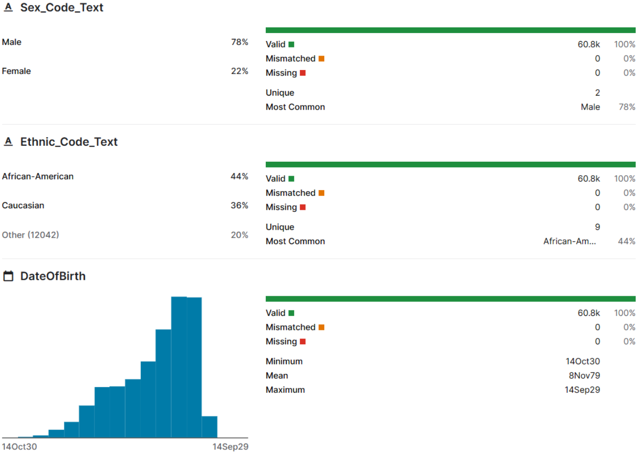
Reporter: Zheng Tan

September 9, 2023

Summary

FairML

COMPAS recidivism data set: to classify whether convicts commit another crime after release.



FairML

"Black defendants were often predicted to be at a higher risk of recidivism than they actually were."

- ERM: Splitting samples by demographic label
- Privacy concerns: preclude recording sensitive information (demographic).
- The joint estimation performs poorly in black subpopulation.

Table 1. Worst Case Error of Recidivism Prediction Models Across Demographic Subgroups

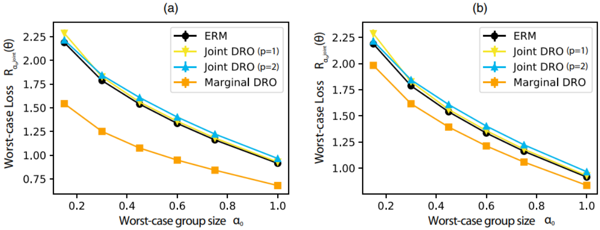
Method	Old	Young	Black	Hispanic	Other race	Female	Misdemeanor
ERM	37.7 ± 0.8	44.6 ± 1.0	37.7 ± 0.8	37.5 ± 0.9	37.9 ± 1.1	37.5 ± 0.9	37.6 ± 0.8
Joint ($p = 2$) − ERM	10.0 ± 2.0	4.0 ± 1.7	9.7 ± 1.7	9.6 ± 1.8	10.9 ± 2.3	9.2 ± 1.8	9.0 ± 1.6
Joint ($p = 1$) − ERM	5.8 ± 1.7	1.2 ± 1.7	5.2 ± 1.5	7.4 ± 1.9	6.6 ± 2.1	5.9 ± 1.7	5.1 ± 1.5
Marginal L = 0.01 − ERM	−0.7 ± 0.7	1.4 ± 1.1	−1.0 ± 0.7	−1.7 ± 0.9	−2.4 ± 1.1	−1.5 ± 0.8	−1.5 ± 0.7
Marginal L = 0.001 − ERM	−1.1 ± 0.7	0.4 ± 1.1	−1.4 ± 0.7	−2.1 ± 0.9	−2.6 ± 1.1	−1.9 ± 0.8	−1.8 ± 0.7

Note. The ERM row shows baseline worst case error; subsequent rows show error differences from baseline (negatives indicate lower error).

Covariate Shift

Wine dataset: predict quality assessment;
Samples: union of subgroups of white and red wines;

* Training samples: red, with real distribution Q_1
* Testing samples: white, with real distribution Q_0
* A certain sample X generates from a mixture distribution $P_X = \alpha Q_0 + (1-\alpha)Q_1$



Ambiguity Set

An assumption: $X \sim P_X = \alpha Q_0 + (1-\alpha) Q_1$, with the latent subpopulation loss:

$$\min_{\theta \in \Theta} \mathbb{E}_{X \sim Q_0}[\mathbb{E}[l(\theta; (X, Y)|X]]$$

With unknown Q_0, Q_1, unknown α in reality!!

Postulate a lower bound $\alpha_0 \in (0, 1/2)$, construct ambiguity set:

$$\mathcal{P}_{\alpha_0, X} = \{Q_0 : P_X = \alpha Q_0 + (1-\alpha) Q_1, \\ \text{for some } \alpha \geq \alpha_0 \text{ and distribution } Q_1 \text{ on } X\}$$

Objective

A joint form(Duchi et al. 2021)

$$\min_{\theta\in\Theta} \sup_{Q_0\in\mathcal{P}_{\alpha_0,(X,Y)}} \mathbb{E}\big[l(\theta);(X,Y)\big]$$

A marginal version, assume that $P_{Y|X}$ is known and does not change across groups:

$$\min_{\theta\in\Theta}\left\{\mathcal{R}(\theta) = \sup_{Q_0\in\mathcal{P}_{\alpha_0,X}} \mathbb{E}_{X\sim Q_0}\big[\mathbb{E}\big[l(\theta);(X,Y)|X\big]\big]\right\}$$

Duality Form & Conditional Risks

A CVaR duality form:

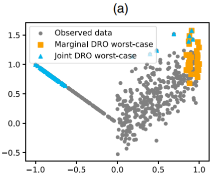
$$\mathcal{R}(\theta) = \sup_{Q_0 \in \mathcal{P}_{\alpha_0, X}} \mathbb{E}_{X \sim Q_0}\left[\mathbb{E}\left[l(\theta); (X, Y)|X\right]\right]\Big\}$$
$$= \inf_{\eta}\{\frac{1}{\alpha_0}\mathbb{E}_{X \sim P_X}[\mathbb{E}([l(\theta; (X, Y)|X] - \eta)_+] + \eta\}$$

Conditional risks:

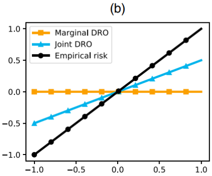
$$\mathbb{E}[\mathbb{E}([l(\theta; (X, Y)|X] - \eta)_+] = \sup_h \mathbb{E}_P[h(X)(l(\theta; (X, Y)) - \eta)]$$

A Numerical Example: Why Marginal Loss Better

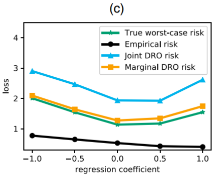
- $Z \sim \text{Bernoulli}(0.15), X = (1 - 2Z) \cdot \text{Uniform}([0, 1])$
- $Y = |X| + \mathbf{1}\{X \geq 0\} \cdot \epsilon$
- $l(\theta, (x, y)) = |\theta x - y|$



Data for a 1-dimensional regression problem (circle) and the worst case distribution Q_0 for joint and marginal DRO (triangle/square).



Best fit lines according to each loss. Only marginal DRO selects a line which fits both the $X > 0$ and $X < 0$ groups.



Loss under different regression coefficients. Unlike Marginal DRO, ERM dramatically underestimates and joint DRO overestimates the worst case loss (2).

New Product Procurement

Ban et al. 2019:

* New, short-life-cycle product under demand uncertainty;
* Unknown demand of new products, but has data on similar products sold in the past;
* A scenario tree method.

Pre-disaster Resource Allocation

Li et al. 2022:

* A new measure called "Shortage Severity Measure" to evaluate the severity of the shortage;
* Equity modeling: horizontal vs vertical.

(a) Horizontal equity (b) Vertical equity

Thank You!