Optimizing the Performative Risk under Weak Convexity Assumptions

Yulai Zhao

NEURAL INFORMATION PROCESSING SYSTEMS

Abstract

- In performative prediction, a predictive model impacts the distribution that generates future data, a phenomenon that is being ignored in classical supervised learning.
- In this closed-loop setting, the natural measure of performance, performative risk (PR), captures the expected loss incurred by a predictive model *after* deployment.
- Prior work has identified general conditions on the loss and the mapping from model parameters to distributions that implies the convexity of the performative risk.
- In this paper, we relax these assumptions and focus on obtaining weaker notions of convexity, without sacrificing the amenability of the PR minimization problem for iterative optimization methods.

Problem

Generally, PR can not be optimized with no specific conditions. We are focusing on identifying these conditions and the extent to which the optimization is accessible. Specifically, we study how to optimize the performative risk through first-order information and analyze the suboptimality gap of performative risks.

How and Under what conditions could we optimize

the performative risks?

We answer this question affirmatively.

We aim to answer this question through two perspectives.

1. Validate several conditions in first-order optimization that could guarantee a linear convergence rate. Once recognizing such conditions, there are plenty implementations in literature. We pay special attention to an implication chain for smooth functions with Lipschitz-continuous gradients.

2. Connect the target performative optima points with stable points to take advantage of previous works.

Methodology

- There are various works in the optimization literature that aim to relax the strong-convexity on the objective function while maintaining favorable convergence properties. Examples include error bounds (EB), essential strong convexity (ESC), weak strong-convexity (WSC), restricted secant inequality (RSI), restricted strong-convexity (RSC), Polyak-Lojasiewicz (PL) inequality and quadratic growth (QG) condition.
- We build on this work to analyze performative prediction from an optimization perspective.

$$(SC) \to (ESC) \to (WSC) \to (RSI) \to (EB) \equiv (PL) \to (QG).$$

Setup

Performative risk is introduced when prediction causes a change in the distribution of the target variable, i.e.,

$$\mathbf{PR(\theta) = E_{z\sim D(\theta)} l(z;\theta)}$$

- Our ultimate goal is to find $\boldsymbol{\theta_{po} = \ argmin_{\theta}\ PR(\theta)}$, performative optima
- However, past work mainly focused on finding

$$\boldsymbol{\theta_{ps} = \mathrm{argmin}_{\theta}\, E_{z\sim D(\theta_{ps})} l(z;\theta)},$$ which are called performative stable points

Example: predicting credit default risk. A bank might estimate that a loan applicant has an elevated risk of default if he applied for a loan, and will act on it by assigning a high interest rate.

Results

1. Weak Strong Convexity for PR

Define DPR$(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$ = $\mathbf{E_{z\sim D(\theta_1)} l(z;\theta_2)}$ for decoupled performative risk.

We show: when DPR is WSC (weakly strong convex), PR is WC (weakly convex) to θ_{po}, namely

$$\mathbf{PR(\theta_{po}) \geq PR(\theta) + \langle \nabla PR(\theta), \theta_{po} - \theta \rangle}$$

Assumption 6 *For performative optimum θ_{PO} and its induced distribution $\mathcal{D} = \mathcal{D}(\theta_{\mathrm{PO}})$, suppose the optimal solution for minimizing $\mathrm{DPR}(\theta_{\mathrm{PO}}, \cdot)$ is θ^*. We say $\mathrm{DPR}(\theta_{\mathrm{PO}}, \cdot)$ satisfies μ−WSC, if for any $\theta \in \Theta$ it holds that*

$$\mathrm{DPR}(\theta_{\mathrm{PO}}, \theta^*) \geq \mathrm{DPR}(\theta_{\mathrm{PO}}, \theta) + \nabla_\theta \mathrm{DPR}(\theta_{\mathrm{PO}}, \theta)^\top (\theta^* - \theta) + \frac{\mu}{2}\|\theta^* - \theta\|^2. \quad (6)$$

2. Restricted Secant Inequality for PR

We show: when DPR is RSI (Restricted Secant Inequality), PR is RSI, namely

$$\boldsymbol{\langle \nabla PR(\theta), \theta - \theta_{po} \rangle \geq \mu' |\theta_{po} - \theta|^2}$$

Assumption 7 *For performative optimum θ_{PO} and its induced distribution $\mathcal{D}$, suppose the optimal solution for minimizing $\mathrm{DPR}(\theta_{\mathrm{PO}}, \cdot)$ is unique and is denoted as θ^*. We say $\mathrm{DPR}(\theta_{\mathrm{PO}}, \cdot)$ satisfies μ−RSI, if for any θ it holds*

$$\langle \nabla_\theta \mathrm{DPR}(\theta_{\mathrm{PO}}, \theta), \theta - \theta^* \rangle \geq \mu \|\theta^* - \theta\|^2 \quad (8)$$

Open Problems

1. Showing PL inequality for PR.

2. Understanding when and how (e.g., some structural properties of loss function or a natural set of distributions), it holds that

$$W\big(D(\theta), D(\theta')\big) \leq C \,||\, \nabla_{\theta'} DPR(\theta, \theta') ||^2$$

Such a condition characterizes local properties of DPR near performative stable points, it could be more common.

3. What is the impact of data pre-processing steps on the implications of performative shifts?

Acknowledgements

YLZ gratefully acknowledges Aurelien Lucchi and Celestine Mendler Dunner for numerous helpful discussions and advice during the internship at ETH Zurich.

References

J´er ˆome Bolte, Trong Phong Nguyen, Juan Peypouquet, and Bruce W Suter. From error bounds to the complexity of first-order descent methods for convex functions. Mathematical Programming, 165(2):471–507, 2017.

Chris Clarke. Financial engineering, not economic photography: Popular discourses of finance and the layered performances of the sub-prime crisis. Journal of Cultural Economy, 5(3): 261–278, 2012.

Franck Cochoy, Martin Giraudeau, and Liz McFall. Performativity, economics and politics: An overview. Journal of Cultural Economy, 3(2):139–146, 2010.

Dmitriy Drusvyatskiy and Adrian S Lewis. Error bounds, quadratic growth, and linear convergence of proximal methods. Mathematics of Operations Research, 43(3):919–948, 2018.

Dmitriy Drusvyatskiy and Lin Xiao. Stochastic optimization with decision-dependent distributions. Mathematics of Operations Research, 2022.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In Proceedings of the 2016 ACM conference on innovations in theoretical computer science, pages 111–122, 2016.

Keegan Harris, Hoda Heidari, and Steven Z Wu. Stateful strategic regression. Advances in Neural Information Processing Systems, 34:28728–28741, 2021.

Nil Kamal Hazra, Mithu Rani Kuiti, Maxim Finkelstein, and Asok K Nanda. On stochastic comparisons of maximum order statistics from the location-scale family of distributions. Journal of Multivariate Analysis, 160:31–41, 2017.

Zachary Izzo, Lexing Ying, and James Zou. How to learn when data reacts to your model: performative gradient descent. In International Conference on Machine Learning, pages 4641–4650. PMLR, 20

Meena Jagadeesan, Celestine Mendler-D ¨unner, and Moritz Hardt. Alternative microfoundations for strategic classification. In International Conference on Machine Learning, pages 4687–4697. PMLR, 2021.

Meena Jagadeesan, Tijana Zrnic, and Celestine Mendler-D ¨unner. Regret Minimization with Performative Feedback. In International Conference on Machine Learning, pages 9760–9785. PMLR, 2022.

Hamed Karimi, Julie Nutini, and Mark Schmidt. Linear convergence of gradient and proximalgradient methods under the polyak-łojasiewicz condition. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 795–811. Springer, 2016.

Qiang Li and Hoi-To Wai. State dependent performative prediction with stochastic approximation. In International Conference on Artificial Intelligence and Statistics, pages 3164–3186. PMLR, 2022.

Ji Liu and Stephen J Wright. Asynchronous stochastic coordinate descent: Parallelism and convergence properties. SIAM Journal on Optimization, 25(1):351–376, 2015.

Ji Liu, Steve Wright, Christopher R´e, Victor Bittorf, and Srikrishna Sridhar. An asynchronous parallel stochastic coordinate descent algorithm. In International Conference on Machine Learning, pages 469–477. PMLR, 2014.