Wasserstein Distributionally Robust Optimization for Machine Learning

Advisors: Franck Iutzeler (Univ. Grenoble Alpes) & Jérôme Malick (CNRS)

Practical informations:

- Host: the applied maths laboratory (LJK) of Univ. Grenoble Alpes, in a team of 10 permanents researchers and 13 Ph.D. students (including 4 former MVA students)
- Located in the campus of Grenoble, with mountains at your doorstep
- Starting date: March/April 2023 – potential Ph.D. position in October 2023
- Strong background in mathematics required; proficiency in Python or Julia appreciated
- Contact: franck.iutzeler@univ-grenoble-alpes.fr or jerome.malick@univ-grenoble-alpes.fr

Key-words: optimization, learning, optimal transport, variational analysis, stochastic algorithms

Context: Taking robust decisions is fundamental in many applications and is gaining importance with the rise of autonomous artificial intelligence systems (see eg. how to cheat neural networks and self-driving cars [1]). Indeed, real situations may differ from training data (due to attacks, lack of data, distributional shifts, or data biais). We would then seek for distributionally robust models that can perform well over all distributions that are close to the training data: denoting by $\ell(\theta, \xi)$ the error of a model parametrized by $\theta \in \mathbb{R}^d$ for a data point $\xi \in \Xi \subset \mathbb{R}^n$, we can learn a distributionally robust model by solving the min-max problem

$$\min_{\theta \in \mathbb{R}^d} \sup_{\mathsf{Q} \in \mathcal{U}} \mathbb{E}_{\xi \sim \mathsf{Q}}[\ell(\theta, \xi)] \tag{1}$$

that features a supremum over the measures in a neighborhood $\mathcal{U}$ of the empirical distribution stemming from the training data $(\xi_i)_{i=1}^m$. For its nice mathematical properties, Wasserstein distance is popular to define this neighborhood [2,3]: $\mathsf{Q} \in \mathcal{U}$ if and only if $W(\frac{1}{m}\sum_{i=1}^m \delta_{\xi_i}, \mathsf{Q}) \leq \rho$. Note that $\rho = 0$ gives the standard empirical risk minimization problem. However, the resulting Wasserstein distributionally robust optimization problem (1) is difficult to solve, in general, due to the constrained supremum over the infinite dimensional space of measures. Inspired by recent developments in optimal transport [4], we recently studied in [5] the entropic regularization of these problems, which naturally amounts to replacing the supremum in (1) by a log-sum-exp approximation.

Topic: In this context, we will consider the theoretical or practical aspects of Wasserstein distributionally robust optimization for machine learning and decision-making. A major question is the design, analysis and implementation of numerical optimization methods for solving entropic regularizations of (1). More theoretically, we will investigate generalization properties and behaviour of worst-case distributions. Finally, we are also interested in the application of these tools to improve fairness; see [6] in the context of federated learning.

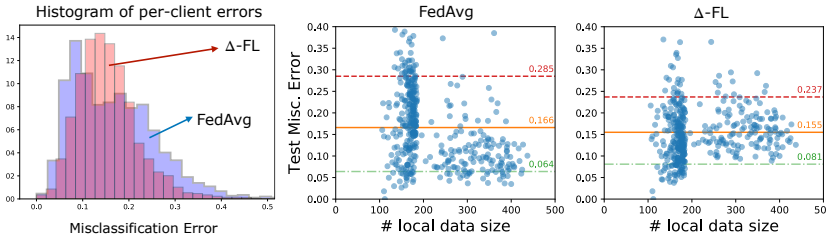

Figure 1: Illustration of distr. robust optimization from [6]: improved prediction of a robust model (Δ-FL) over the standard one (FedAvg) in a federated setting, on both non-conforming users and data-poor users.

References:

[1] https://adversarial-ml-tutorial.org/ and Kevin Eykholt et al. "Robust physical-world attacks on deep learning visual classification." CVPR 2018.

[2] Daniel Kuhn et al. "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations." Mathematical Programming, 2018

[3] Jose Blanchet, et al "Statistical analysis of Wasserstein distributionally robust estimators." Tutorials in Operations Research: Emerging Optimization Methods and Modeling Techniques. INFORMS, 2021

[4] Gabriel Peyré, and Marco Cuturi. "Computational optimal transport". FnT in Machine Learning, 2019

[5] Waïss Azizian, Franck Iutzeler, and Jérôme Malick. "Regularization for Wasserstein distributionally robust optimization." preprint arXiv:2205.08826, 2022.

[6] Krishna Pillutla, Yassine Laguel, Jérôme Malick, and Zaid Harchaoui. "Tackling Distribution Shifts in Federated Learning with Superquantile Aggregation." Spotlight at NeurIPS22 workshop on distrib. shifts, 2022.