Research Internship

Efficient optimization methods for outlier-robust machine learning

Supervisor: Tam Le — `tamle@lpsm.paris`

Location: Laboratoire de Probabilités, Statistique et Modélisation (LPSM), Université Paris Cité, Campus des Grands Moulins, 75013 Paris, France.

Keywords: median, machine learning, robustness, gradient descent, optimal transport.

Context. Machine-learning models are typically trained on large datasets using empirical risk minimization (ERM) and (stochastic) gradient methods [3]:

$$\min_{w \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} f(w, x_i), \tag{ERM}$$

where f measures prediction error and $x_1, \ldots, x_n$ are sample points. In practical situations, part of the dataset may be corrupted or of low quality. Such *outliers* can significantly degrade performance, motivating robust alternatives to standard ERM [5, 1]. This internship focuses on designing scalable stochastic algorithms for outlier-robust learning, using two notions of "median": *risk-level robustness* through the median-of-means principle (1), and *distributional robustness* through Wasserstein medians (2).

1. Median-of-means minimization. The first line of research thus investigates training procedures based on the *median-of-means* (MOM) estimator [10, 8]. The dataset is partitioned into K blocks of size $M = n/K$, and the objective becomes

$$\min_{w \in \mathbb{R}^p} \text{Median} \left\{ \frac{1}{M} \sum_{i \in I_k} f(w, x_i) : k = 1, \ldots, K \right\}.$$

This estimator is statistically robust and has proven effective in regression, ranking, and optimal transport [7, 6, 13]. However, its nondifferentiability and nonconvexity make standard gradient methods costly, as identifying the median block requires evaluating losses on the entire dataset. Toward efficient and scalable gradient-based methods, possible research directions during the internship include:

1. Developing stochastic gradient methods for MOM minimization using bias-reduction techniques to dynamically and efficiently identify median block [12, 9].

2. Designing fast optimization algorithms that exploit smooth structures hidden in the median landscape, extending recent advances in composite convex optimization [2].

2. Wasserstein median for outlier robustness. The second line of research aims to extend median-of-means approach to the space of probability measures via the *Wasserstein median* [4]:

$$\mathrm{Med}(\pi_1, \ldots, \pi_K) \in \mathrm{argmin}_\mu \sum \lambda_i W(\mu, \pi_i)$$

where $\pi_1, \ldots, \pi_K$ are probability measures and W is the Wasserstein distance [11]. This notion provides a robust distributional center and connects naturally with robust learning, clustering, and unsupervised tasks. During the internship, potential tasks to explore this notion in the context of outlier robustness may include:

1. Formulating gradient-based algorithms to compute Wasserstein medians of discrete measures.
2. Comparing the robustness and computational behavior of Wasserstein-medians with the median-of-means approach.

Prerequisites. Applicants should have a solid mathematical background, including one or more courses in statistics, machine learning, and optimization. Programming proficiency (preferably in Python and Pytorch) for prototyping algorithms is a strong plus. Prior experience with gradient methods and optimization is welcome but not mandatory.

Work Environment. The intern will receive a monthly stipend of approximately 600 euros (subject to standard adjustments). The research will be conducted at LPSM, Université Paris Cité, on the Campus des Grands Moulins. The intern will also participate in the activities of the Statistics, Data, Algorithm group at LPSM, which brings together researchers in statistics, optimization, and machine learning.

Application Procedure. Candidates are invited to send their application by email to Tam Le (`tamle@lpsm.paris`). Applications should include a curriculum vitae, recent transcripts of grades and a brief cover letter.

References

[1] C. C. Aggarwal, *Outlier Analysis*, Computer Science, Springer Cham, 2 ed., 2017.

[2] G. Bareilles, F. Iutzeler, and J. Malick, *Harnessing structure in composite nonsmooth minimization*, SIAM Journal on Optimization, 33 (2023), pp. 2222–2247.

[3] L. Bottou, F. E. Curtis, and J. Nocedal, *Optimization methods for large-scale machine learning*, SIAM Review, 60 (2018), p. 223–311.

[4] G. Carlier, E. Chenchene, and K. Eichinger, *Wasserstein medians: Robustness, pde characterization, and numerics*, SIAM Journal on Mathematical Analysis, 56 (2024), pp. 6483–6520.

[5] P. J. Huber, *Robust Statistics*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1248–1251.

[6] P. Laforgue, G. Staerman, and S. Clémençon, *Generalization bounds in the presence of outliers: a median-of-means study*, in Proceedings of the 38th International Conference on Machine Learning, M. Meila and T. Zhang, eds., vol. 139 of Proceedings of Machine Learning Research, PMLR, 18–24 Jul 2021, pp. 5937–5947.

[7] G. Lecué and M. Lerasle, *Robust machine learning by median-of-means: Theory and practice*, The Annals of Statistics, 48 (2020), pp. 906 – 931.

[8] G. Lugosi and S. Mendelson, *Regularization, sparse recovery, and median-of-means tournaments*, Bernoulli, 25 (2019), pp. 2075 – 2106.

[9] R. Mehta, V. Roulet, K. Pillutla, and Z. Harchaoui, *Distributionally robust optimization with bias & variance reduced gradients*, in The Twelfth International Conference on Learning Representations, 2024.

[10] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, Inc., New York, 1983. A Wiley-Interscience Publication. Translated from the Russian and with a preface by E. R. Dawson.

[11] G. Peyré and M. Cuturi, *Computational optimal transport: With applications to data science*, Found. Trends Mach. Learn., 11 (2019), p. 355–607.

[12] A. Ruszczyński, *A stochastic subgradient method for nonsmooth nonconvex multilevel composition optimization*, SIAM J. Control. Optim., 59 (2021), pp. 2301–2320.

[13] G. Staerman, P. Laforgue, P. Mozharovskyi, and F. d'Alché Buc, *When ot meets mom: Robust estimation of wasserstein distance*, in Proceedings of The 24th International Conference on Artificial Intelligence and Statistics, A. Banerjee and K. Fukumizu, eds., vol. 130 of Proceedings of Machine Learning Research, PMLR, 13–15 Apr 2021, pp. 136–144.