Concentration of the geometry of empirical risks

Advisors: Franck Iutzeler (Prof. Univ. Toulouse III) & Paul Escande (CR CNRS)

Practical informations:

Host: Institut de Mathématiques de Toulouse (IMT), in the Statistics and Optimization team.

Location: Campus of the Université Paul Sabatier, Toulouse.

Duration: 4 to 6 months starting from March/April 2025 – potential Ph.D. position in October 2025.

Candidate profile: Strong background in mathematics required with an interest in machine learning and optimization.

Contact: franck.iutzeler@math.univ-toulouse.fr or paul.escande@math.univ-toulouse.fr

Keywords: Machine learning, statistical learning, empirical risk minimization, optimization

Context: A common strategy when training a machine learning model consists in minimizing the Empirical Risk which measures the error of the model on a finite training dataset. This approach is employed as a proxy for the minimization of the true risk (the average error over the distribution of the data) since the entire data distribution is unknown in most situations of interest. A fundamental question in statistical learning is to transfer guarantees on the empirical model to the one obtained via the minimization of the true risk. That is, comparing the theoretical guarantees on

$$\min_{\phi \in \mathcal{M}} \widehat{\mathcal{J}}(\phi) = \frac{1}{n} \sum_{i=1}^{n} l(\phi, X_i) \quad \text{to the ones on} \quad \min_{\phi \in \mathcal{M}} \mathcal{J}(\phi) = \mathbb{E}_{X \sim \mu}[l(\phi, X)],$$

where ϕ describes the parameters of the model (optimized over a set $\mathcal{M}$), l is the loss function and $(X_i)_{i=1}^n$ is the training data drawn from the distribution μ. These fundamental questions can be tracked back to the 1970's [5] and have led to a remarkably rich and mature theory. While most of the guarantees in the literature are derived on the excess risk or on the generalization error [6, 2], recent works tend to provide concentration guarantees on a distance between the sets of minimizers [4, 3]. These works suggest that understanding the concentration of empirical minimizers is intricate to the local geometry of the empirical risks around its minimizers.

Topic: In this context, we will investigate the properties of the local geometry of the empirical risks that are relevant to the concentration of the minimizers. The aim of this internship will also be to propose the assumptions under which the local geometry of the empirical risks around its minimizers mimics the one of the true risks, with respect to the properties identified above. Understanding this behavior is an open question in statistical learning, and may lead to significant advances in explaning the concentration occurring in complex machine learning models such as neural networks.

References:

[1] Waïss Azizian, Franck Iutzeler, and Jérôme Malick. “Exact generalization guarantees for (regularized) wasserstein distributionally robust models”. In: *Advances in Neural Information Processing Systems* (2023).

[2] Olivier Bousquet and André Elisseeff. “Stability and generalization”. In: *Journal of Machine Learning Research* (2002).

[3] Paul Escande. “On the Concentration of the Minimizers of Empirical Risks”. In: *Journal of Machine Learning Research* (2024).

[4] Christof Schötz. “Convergence rates for the generalized Fréchet mean via the quadruple inequality”. In: *Electronic Journal of Statistics* (2019).

[5] Vladimir Vapnik and Alexey Chervonenkis. “On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities”. In: *Theory of Probability and its Applications* (1971).

[6] Martin Wainwright. *High-Dimensional Dtatistics: A Non-Asymptotic Viewpoint*. Cambridge Univ. Press, 2019.