Early Stopping iterative learning algorithms. Saving the computational resource (Frugal AI)

M2 – Intership proposal

Context Numerous approaches in statistical/machine learning require an extensive computational power: AI technologies such as deep neural networks (DNNs) heavily rely on GPU computations while more classical approaches are costly to run on tables with millions of rows. Frugal AI is a domain of AI encompassing all strategies aiming at reducing the computational costs of the learning strategies or designing new strategies able to perform optimally with a limited computational "budget".

All of this is precisely the context motivating the introduction of the co-called "early stopping rules". When applying an iterative learning algorithm such as gradient descent or any boosting algorithm, a key question (usually left unsolved to the end-user) is the choice of the number of steps to be performed. This number of steps reveals crucial in practice since too many steps lead to wasting space and time (computational resources) whereas too few steps would worsen the final statistical performance.

Designing an early stopping rule precisely consists in: (i) defining a data-driven number of iterations to be performed, (ii) proving the optimal performance of this data-driven number of iterations (early stopping rule) both theoretically and empirically on synthetic and real datasets. The question of designing new early stopping rule is related to the so-called *implicit regularization* phenomenon and can be also understood as an alternative model selection technique.

Up to now Early Stopping still remains a challenging question, although recent advances in the domain have been made [1,2,3,4]. Actually any valuable contribution to early stopping would impact a lot of famous learning algorithms such as LASSO, Projected and Proximal gradient descent, Quasi-Newton methods, boosting algorithms, PCA and SVD, density estimation with kernels, Stochastic Gradient Descent (SGD) and variants (that are ubiquitous when training DNNs), change-point detection, clustering procedures, ...

Expected achievements

1. Reading the state-of-the-art literature on early stopping and related topics such as optimization.
2. Implementing (Python) the procedure designed by [2] in the context of reproducing kernels and gradient descent.
3. Extending the procedure to more general gradient descent-like algorithms such as proximal gradient descent for instance, which is used in constrained optimization (projected gradient descent) or when optimizing non-smooth functions.
4. Studying how to adapt (empirically and theoretically) the strategy used with the previous algorithms to boosting algorithms that are performed from a set of weak learners.

Bibliography

1 Hucker, L., Reiß, M. (2024). Early stopping for conjugate gradients in statistical inverse problems. arXiv preprint arXiv:2406.15001.

2 Celisse, A., Wahl, M. (2021). Analyzing the discrepancy principle for kernelized spectral filter learning algorithms. Journal of Machine Learning Research, 22(76), 1-59.

3 Blanchard, G., Hoffmann, M., Reiß, M. (2018). Optimal adaptation for early stopping in statistical inverse problems. SIAM/ASA Journal on Uncertainty Quantification, 6(3), 1043-1075.

4 Raskutti, G., Wainwright, M. J., Yu, B. (2014). Early stopping and non-parametric regression: an optimal data-dependent stopping rule. The Journal of Machine Learning Research, 15(1), 335-366.

Advisor

- Alain Celisse, Professeur à l'université Paris 1 Panthéon-Sorbonne.
  `alain.celisse@univ-paris1.fr`

- Location: Laboratoire SAMM – Université Paris 1 Panthéon-Sorbonne
  Centre PMF, 90 Rue de Tolbiac Paris 13

- Duration: 4-6 month (M2 dissertation), followed by a potential PhD thesis (3 years)