Internship: Online graph inference for decentralized learning with heterogeneous data

In Machine Learning, decentralized and federated learning methods allow training from data stored locally by several agents (nodes) without exchanging raw data, in line with the increasing demand for more privacy-preserving algorithms (Kairouz et al., 2021). One of the key challenges in decentralized learning is to deal with data heterogeneity: as each agent collects its own data, local datasets typically exhibit different distributions. This internship will study this challenge in the context of fully decentralized learning algorithms, which provide a scalable and robust alternative to server-based approaches. Fully decentralized optimization algorithms, such as the celebrated Decentralized SGD (D-SGD) (Lian et al., 2017), operate on a graph representing the communication topology, i.e. which pairs of nodes exchange information with each other. The connectivity of the topology then rules a trade-off between the convergence rate and the per-iteration communication complexity of the algorithms. Choosing a good topology for fully decentralized machine learning is therefore an important question, and remains a largely open problem in the presence of data heterogeneity. In Le Bars et al. (2022), the authors show that an appropriate communication topology can reduce the detrimental effect of data-heterogeneity and they propose an algorithm to learn it. However, their theoretical results are only applied to the D-SGD algorithm and the communication topology in learned *offline* with a restriction to a specific classification framework. The goal of this internship is to extend their work to learn the topology in an *online* fashion, during the optimization process (e.g. using bi-level optimization (Dagréou et al., 2022)), while considering more general heterogeneous settings. An additional direction is to extend the theoretical results obtained for D-SGD to other decentralized algorithms (such as gradient push (Assran et al., 2019)).

Objectives: The main objectives of this 6 months internship are (i) to review some of the existing literature in the field, (ii) to design new (online) algorithms that learn the communication graph in decentralized learning and (iii) to derive theoretical guarantees for the proposed algorithms.

Requirements: Successful candidates should have a solid background in Machine Learning with an interest in Decentralized Optimization. Some knowledge in Statistical Learning Theory would be a plus but is not mandatory. A good understanding of the Python programming language is expected. Finally, good command of English is required as it will be one of the main working language.

Keywords: Machine Learning, Decentralized Learning, Learning Theory, Optimization

Contact: The recruited student will be based in the INRIA Research Center of Lille University and will be supervised by Aurélien Bellet and Batiste Le Bars.
The interested students should send an e-mail with a CV and a Motivation Letter to *aurelien.bellet@inria.fr* and *batiste.le-bars@inria.fr*.

References

Assran, M., Loizou, N., Ballas, N., and Rabbat, M. (2019). Stochastic gradient push for distributed deep learning. In *International Conference on Machine Learning*, pages 344–353. PMLR.

Dagréou, M., Ablin, P., Vaiter, S., and Moreau, T. (2022). A framework for bilevel optimization that enables stochastic and global variance reduction algorithms. *Neurips 2022*.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210.

Le Bars, B., Bellet, A., Tommasi, M., Lavoie, E., and Kermarrec, A. (2022). Refined convergence and topology learning for decentralized optimization with heterogeneous data.

Lian, X., Zhang, C., Zhang, H., Hsieh, C.-J., Zhang, W., and Liu, J. (2017). Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent. *Advances in Neural Information Processing Systems*, 30.