Intro to Federated Learning

Gianluca Mittone[1]

[1]University of Turin, Italy

Machine Learning (ML) is the branch of Artificial Intelligence focused on developing algorithms capable of adapting and improving their predictive or generative performance by feeding on data. Adapting or improving the system's behaviour based on the provided data is called learning since it is similar to the human learning process in many aspects. The same ML algorithm, usually referred to as a model, trained on different data will thus expose different capabilities and can, therefore, solve different tasks.

FL is a relatively recent distributed ML methodology [1] aiming to bridge the gap between the need to train ever bigger ML models on ever larger datasets and the individual and companies' will to protect and not share their private data. From another point of view, FL is also a way to distribute the training of an ML model even more than before. However, it should be considered that the learning performance of FL is usually lower than that of traditional centralised learning [2].

This course will start from Kairouz and McMahan's definition of FL [3]: "Federated learning is a machine learning setting where multiple entities (clients) collaborate in solving a machine learning problem, under the coordination of a central server or service provider. Each client's raw data is stored locally and not exchanged or transferred; instead, focused updates intended for immediate aggregation are used to achieve the learning objective." From this starting point, the most significant aspects of FL will be exposed and discussed.

This tutorial will particularly explore FL from both the learning [4] and computational [5] performance perspectives, investigating its pros and cons in a distributed ML setting. Since FL natively targets data privacy, some insights on how the FL process can be attacked and protected will also be discussed from a high-level perspective. Finally, a hands-on session will guide the participants in building a basic FL system, providing a better understanding of the major implementational difficulties of such a technique.

[1] Brendan McMahan et al. "Communication-Efficient Learning of Deep Networks from Decentralized Data." In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 2017.

[2] Gianluca Mittone et al. "A Federated Learning Benchmark for Drug-Target Interaction." In: Companion Proceedings of the ACM Web Conference, 2023.

[3] Peter Kairouz et al. "Advances and Open Problems in Federated Learning." In: Foundations and Trends in Machine Learning, 2021.

[4] Gianluca Mittone et al. "Model-Agnostic Federated Learning." In: Euro-Par 2023: Parallel Processing - 29th International Conference on Parallel and Distributed Computing, 2023.

[5] Gianluca Mittone et al. "Experimenting with Emerging RISC-V Systems for Decentralised Machine Learning." In: Proceedings of the 20th ACM International Conference on Computing Frontiers, 2023.