

# MVA Internship proposal on Federated Learning

Marie Garin  
Argyris Kalogeratos    Nicolas Vayatis  
Centre Borelli, ENS Paris-Saclay

December 20, 2021

## Abstract

This internship is articulated around distributed architecture and is especially focused on federated learning. The student will deal with distributed optimization, empirical risk minimization, and aggregation problems. Many directions of research are open depending on the interest of the candidate. The internship includes both theoretical and computational aspects.

**Context** Federated learning [1, 2] has stood out as a promising field of growing emphasis, advocating for an alternative to centralized learning. The latter widespread architecture is subject to many constraints, the crucial ones being privacy and latency. These shortcomings impelled machine learning community to consider more distributed approaches. Federated learning is one of them and relies on both local data storage and local model training on devices. Above and beyond the computational gain through the use of the processor of the devices, the benefit of keeping data locally is twofold. On the one hand, it preserves privacy, an apodictic value whose endangerment entails insidious societal effects such as self-censorship. On the other hand, it enables to collaboratively learn when releasing data is impractical - if not illegal.

## Mathematical elements

**Definition 1 (Federated setting)** *For a set of  $m$  nodes, the federated setting is defined as follows:*

- Each device  $i \in [m]$  owns a sample  $\{z_{ij} \in \mathcal{Z} : j \in [n_i]\}$  of size  $n_i$ , i.i.d. drawn from a distribution  $\mathcal{P}$ ;
- Given a loss function  $\ell : \Theta \times \mathcal{Z} \rightarrow \mathbb{R}_+$ , the risk is defined as  $R(\theta) = \mathbb{E}[\ell(\theta; Z)]$ , where  $\theta \in \Theta$  is a specific set of parameters for the learning model;
- Lastly, we set  $\theta^* \in \arg \min_{\theta \in \Theta} R(\theta)$ .

## Possible points to investigate

**A)** Breaking the aforementioned *i.i.d.* assumption. Here, each device owns a sample drawn from a distribution  $\mathcal{P}_i$  of its own. This new assumption impacts the definition of the risk, the optimal element, etc.

In one-shot federated learning, each device  $i$  estimates its local parameter  $\hat{\theta}_i \in \arg \min_{\theta \in \Theta} \hat{R}_i(\theta)$  defining the local empirical risk  $\hat{R}_i(\theta)$  as  $\frac{1}{n_i} \sum_j^{n_i} \ell(\theta; z_{ij})$ . The final aggregated parameter is usually the average, and since the sample size can vary a lot among the nodes, each local parameter is weighted *according to its sample size*, by a factor  $\frac{n_i}{\sum_k^m n_k} = \frac{n_i}{N}$ :

$$\hat{\theta}_S = \sum_{i=1}^m \frac{n_i}{N} \hat{\theta}_i. \quad (1)$$

The ongoing research focus on the optimization of the weights associated to a new aggregated parameter defined as follow:

$$\hat{\theta}_W = \sum_{i=1}^m w_i \hat{\theta}_i$$

**B)** Personalized models that return to each node a different  $\hat{\theta}^{(i)}$ .  
How to perform this personalization through the weights?

**C)** Another direction is to go beyond the one-shot approach and explore the multi-round optimization (*e.g.* stochastic gradient descent).

**Objectives:** The aim is to study one or more of the above problems, implement the designed methods, and set also realistic experiments to demonstrate comparative gains against existing approaches.

## References

- [1] J. Konečný, B. McMahan, and D. Ramage. Federated optimization: Distributed optimization beyond the datacenter. *CoRR*, abs/1511.03575, 2015.
- [2] J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik. Federated optimization: Distributed machine learning for on-device intelligence. *CoRR*, abs/1610.02527, 2016.