Federated Learning: Does everyone play fair?

1. Introduction

Data, data and more data. Undoubtedly, data plays a fundamental role in today's world. As a result, numerous applications based on Artificial Intelligence, or Machine Learning to be more specific, are emerging with the aim of using all the available information. Moreover, in all these applications there is a clear premise: the more (quality) data available, the better the performance of the implemented system. This fact invites us to think: is it possible for several entities to collaborate with their data to obtain a common result of higher quality? Obviously, the answer is yes, especially in problems common to all the participating entities, for example: medical diagnosis problems among several hospitals or fraud detection among different banking entities. However, current legislation and the existence of sensitive data that entities do not want to disclose make this idea difficult. To solve this problem, a new line of research has emerged in recent years: **Federated Learning**. This new concept of learning is used to train machine learning algorithms, for example deep neural networks, on multiple separate datasets contained in local nodes without sharing the data between them, thus keeping it private. In this project, we will focus on a deep learning image classification problem: What are the effects of training a model on just one sample of images from one node against multiple samples from a distributed grid of nodes? Is the data at each node kept private? **What would happen if a node has been attacked and has corrupt data?**

2. The problem

GMV is currently immersed in the development of Federated Learning-based solutions with the aim of implementing collaborative projects among different entities while preserving data privacy. However, as this is a relatively new area of research, the full effects that the nature of the problem may have on the system learning are not yet well-known. In this particular problem, GMV wants to know what would be the influence of a node with poisoned data on a Federated Learning model.

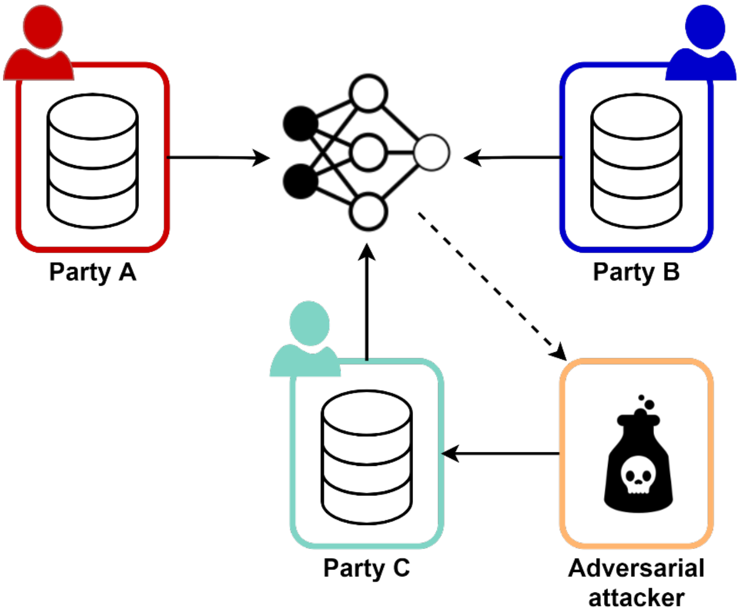

Figure 1. Problem description diagram.

As the main interest as of right now is research, we are going to provide **technical** documentation to get into the inner workings of the method and an open **dataset** to put the theory into practice. For this purpose, students will work with the well-known MNIST [1] dataset, which consists of handwritten digits from 0 to 9. The objective of the implemented models will be to classify each image with the corresponding digit. For the problem, GMV will provide two versions of this dataset:

- The original dataset distributed in three parties.
- A version of the same dataset in which one of the parties has suffered an attack and will therefore contribute poisoned data to the federated learning model.

Based on the above, the following objectives are proposed:

- **Understand** the idea and main concepts of **federated learning**.
- **Understand** the basic concepts of **adversarial attacks**.
- **Train a global model through Federated Learning** so that parties do not disclose their data.
- **Compare the results obtained with local models** (trained by each party with its own data) and study the effects: what are the advantages and disadvantages of using Federated Learning in the problem under study?
- **Train a federated model with poisoned data** and analyze the results:
    - Is it possible to identify the attacked party?
    - What are the effects of poisoned data from the attacked node on the global federated model?
    - Are there solutions to this problem?

3. Work plan and learning outcomes

At the beginning, the students will receive an introduction on the problem, methodology and the approach that is used for this matter. They will then be linked to some tutorials from GitHub [2] that will provide them a better understanding of the subject. The rest of the week will be devoted to analyze the datasets, train the algorithms in each of the cases explained above, study the effects of each scenario on the algorithms and extract conclusions from the results. We expect that at the end of the modelling week, the students will have a thorough understanding of the problem and the methods used to solve it, as well as means to continue digging deeper into the topic if they so wish.

4. Literature and references

[1] "MNIST dataset," [Online]. Available: http://yann.lecun.com/exdb/mnist/.

[2] "PySyft," [Online]. Available: https://github.com/OpenMined/PySyft.

[3] J. Konecny, H. B. McMahan, D. Ramage and P. Richtarik, "Federated Optimization: Distributed Machine Learning for On-Device Intelligence," 2016.

[4] T. Li, A. K. Sahu, A. Talwalkar and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," 2019.

[5] Q. Yang, Y. Liu, T. Chen and Y. Tong, "Federated Machine Learning: Concept and Applications".

[6] "Pytorch tutorial on adversarial example generation," [Online]. Available: https://pytorch.org/tutorials/beginner/fgsm_tutorial.html.

[7] Kurakin, A., Goodfellow, I., & Bengio, S. (2016). Adversarial machine learning at scale. *arXiv preprint arXiv:1611.01236*. Available: https://arxiv.org/pdf/1611.01236.pdf