

# Report

Project	<b>Convergence guarantees for gradient descent methods on optimization for non-convex function approximation</b>
Student	<b>Hoang Trung Hieu</b>
Supervisor	<b>Horváth Tomáš</b>
Program	MSc, Applied Mathematics
Semester	2020/2021, Autumn

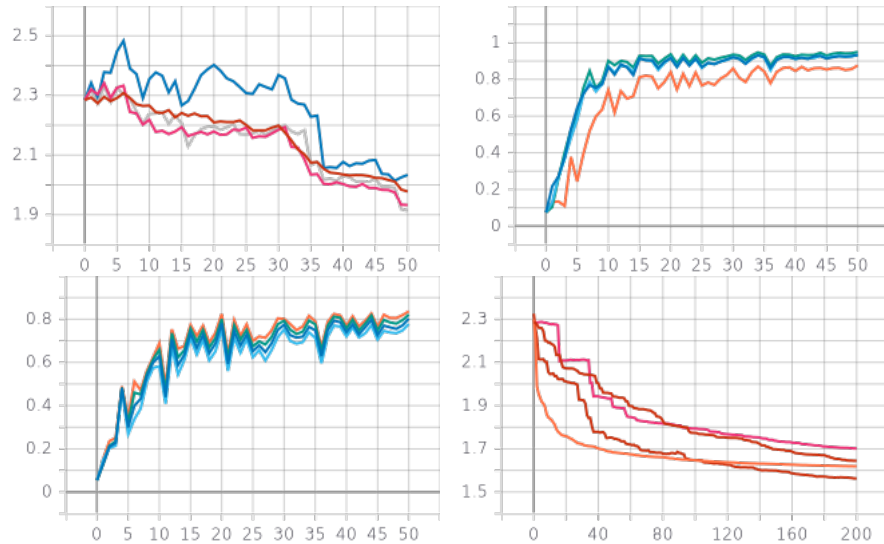
## Summary

Federated learning (FL) is a learning paradigm seeking to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself. But there are some limitations: Massively distributed, non-iid dataset, the data distribution's unbalancedness. To train a neural network in general we often use the gradient decent methods to optimize the loss function. In FL, Federated Averaging(FedAvg) is used most frequently. However, it is not perfect and it performs poor-quality in some cases. My project is to analyze whether it is possible to give convergence guarantees in case the data over each devices is unbiased and evaluating of divergence between local datasets.

In the first semester, my work is understanding the main concept of Federated Learning, FedAvg algorithm, and given convergence analyses when dataset is balanced or in convex problem. Besides, I also experiment on MNIST data set and Synthetic data and illustrate by meaningful plots.

## Process and Results

- Firstly, I took the Advanced Machine Learning Systems course [6], it gives a deeply analysis on the convergence of gradient decent methods. It also explains the meaning of adaptive learning rate and how it enhances the method.
  - These papers ([2],[3],[8]) give a general look of federated learning, the challenges and limitations .
  - In order to attempt to analyze the convergence of FedAvg, [7] is a paper analysing the convergence of convex approximation function on non-iid data and [1] present experiments on applying different popular optimization methods for training neural networks in a federated manner.
  - For experiment part, the code I use from [4] , [5], the dataset is MNIST and and the synthetic data is followed by [9]. The models are 2 hidden layers neural network and 2 convolutional layers neural network.
- In these figures, I plot the line graphs comparing the impact of different parameters (the number of device each update(top-left), the number of epoch (top-right)batch size (bottom-left)) and impact of unbalancedness (bottom-right).



## References

- [1] V. Felbab, P. Kiss, and T. Horváth, (2019), *Optimization in Federated Learning*, ITAT 2019
- [2] F. Sattler, S. Wiedemann, K. Müller, W. Samek, (2019), *Robust and communication-efficient federated learning from non-iid data*, IEEE transactions on neural networks and learning systems,
- [3] H. Brendan McMahan, Eider Moore, Daniel Ramage, (2017) , *Communication-Efficient Learning of Deep Networks from Decentralized Data*, <https://arxiv.org/abs/1602.05629>
- [4] *TensorFlow Federated*, <https://github.com/tensorflow/federated>
- [5] *On the Convergence of FedAvg on Non-IID Data repository*, <https://github.com/lx10077/fedavgpy>
- [6] *Advanced Machine Learning Systems Course* , (2017), <https://www.cs.cornell.edu/courses/cs6787/2017fa/>
- [7] X. Li, K. Huang, W. Yang, S. Wang, Z. Zhang, (2019), *On the Convergence of FedAvg on Non-IID Data*, <https://arxiv.org/abs/1907.02189>
- [8] J. Konečný, H. B. McMahan, D. Ramage, P. Richtárik, (2016), *Federated Optimization Distributed Machine Learning for On-Device Intelligence* <https://arxiv.org/abs/1610.02527>
- [9] A. K. Sahu, T. Li, M. Sanjabi, M. Zaheer, A. Talwalkar, and V. Smith, (2018), *Federated optimization for heterogeneous networks*, <https://arxiv.org/abs/1812.06127>