Gradient-Masked Federated Optimization for OOD Generalization in FL

Irene Tenison, Sreya Francis, Irina Rish

Abstract

In real-world federated settings, the data will be distributed non-IID across clients. Which means, the data from any additional client clients will be out-of-distribution to the entire train data with which the server model was indirectly trained. Hence, OOD generalization is significant for the model performance on newer or non-participating clients.

Experiments

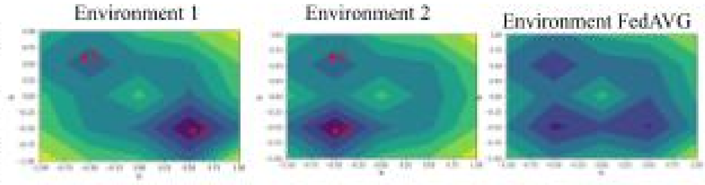
IID: Clients will have train samples from all digits

Non-IID: 80% of each client's train data will be of any two digits and rest 20% will be of the remaining digits

Test: Out-of-Distribution test dataset having inverted or modified spurious mechanism

Train OOD Test

Method



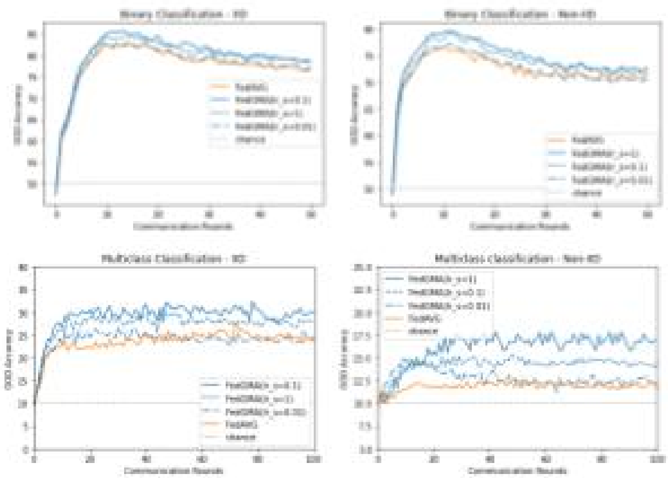
Point C is invariant across Env 1 and Env 2. For better OOD-Gen, C is to be pursued at the global model over the client global minima at A or B.

At Server after local updates:

mask = |avg(sign(client gradients))| ≥ agreement threshold
final gradients = mask*avg(|client gradients|)
Server params = avg(client params) + lr*final gradients

When more than agreement threshold number of clients have a consistent minima across clients, the global model takes an additional step in the direction of this consistent minima with a specified lr/step size.

Results



Conclusion and Future Work

- In this paper we propose FedGMA for OOD generalization and observe improved performance.
- For future work, we will provide theoretical explanations and experiment with more complex datasets (with focus on medical datasets)

Contact Info: Irene Tenison irene.tenison@mila.quebec

Distributed and Private Machine Learning, ICLR Workshop, May 7 2021