

Deep Learning - DS3040

Federated learning with heterogeneous data formats

SanthoshV(142302020)

SivaKumar(142302008)

Github link for resources(Research papers and codes for frameworks):

1. <https://github.com/monkl337/Aweome-Heathcare-Federated-Learning?tab=readme-ov-file>
2. <https://github.com/albarqouni/Federated-Learning-In-Healthcare?tab=readme-ov-file>
3. <https://github.com/adap/flower>(flower framework)
4. <https://github.com/FedML-AI/FedML>-(fedml framework)-<https://github.com/FedML-AI/FedML/tree/master/python/fedml>
5. <https://github.com/OpenMined/PySyft>(pysyft framework)

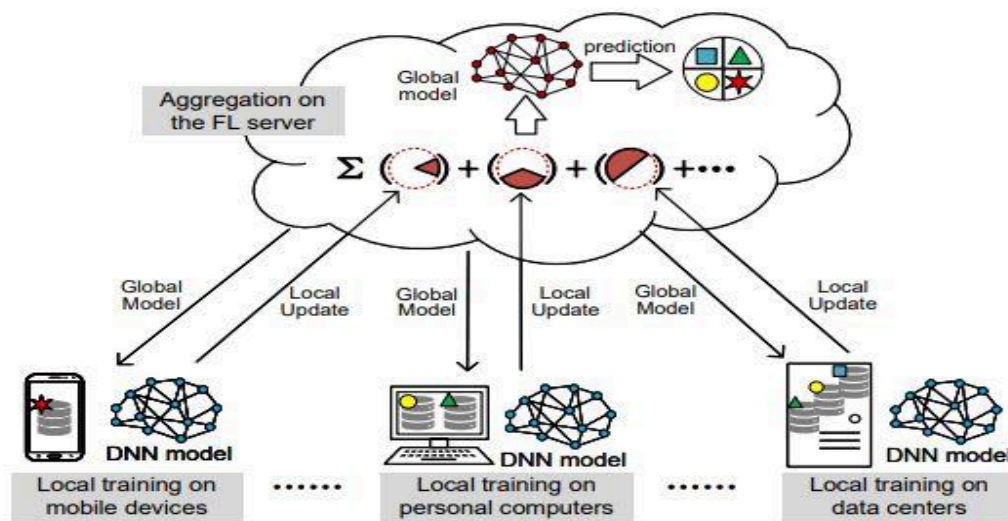
Federated learning?

Federated learning is a method to synthesize global models from local models trained on the edge. FL was first developed by Google in 2016 for their **Gboard application**, which incorporates the context of an Android user's typing history to suggest corrections and propose candidates for subsequent words. The solution that Google produced was a decentralized training approach where an iterative process would compute model training updates at the edge, aggregating these updates to produce the global update to be applied to the model. This core concept of aggregating model updates was key in allowing for a single, performant model to be produced from edge training.

The FL process

To better understand FL from an iterative process perspective, The steps for a round can be described as follows:

1. The aggregate global model parameters are sent to each user's device.
2. The received ML models located on the user devices are trained with local data.
3. After a certain amount of training, the local model parameters are sent to the central server.
4. The central server aggregates the local models by applying an aggregation function, producing a new aggregate global model.



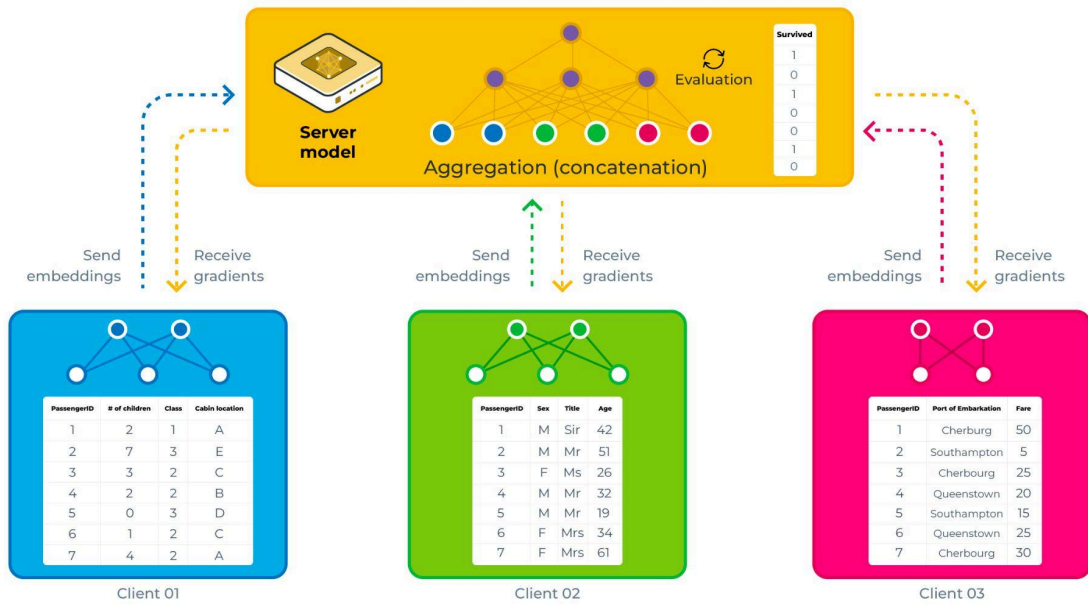
Horizontal Federated Learning

- Sample-based federated learning or homogenous federated learning
- Involves setting the data that has the same features but operates within a different sample space.

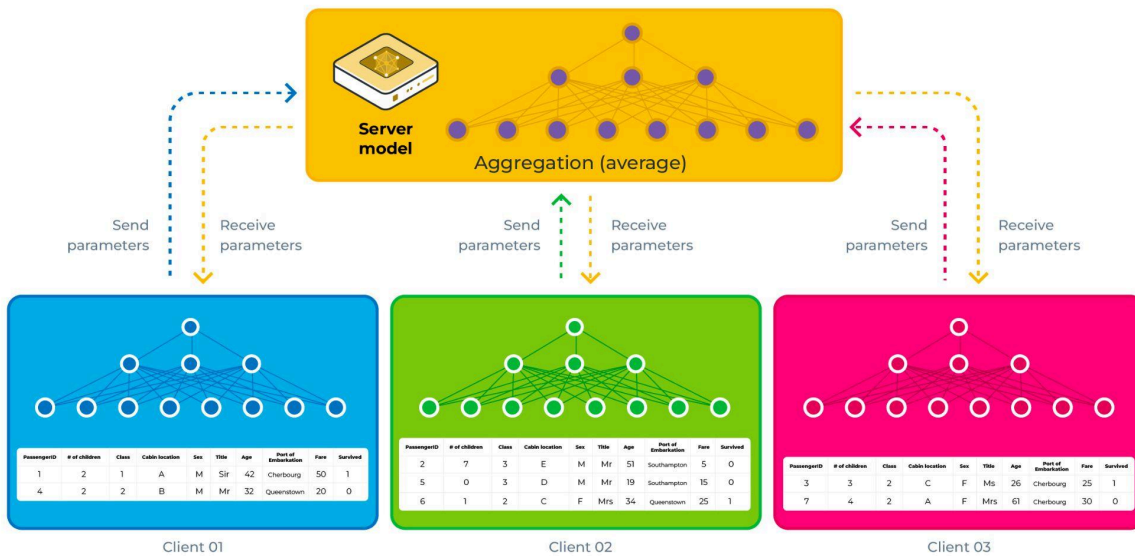
Vertical Federated Learning

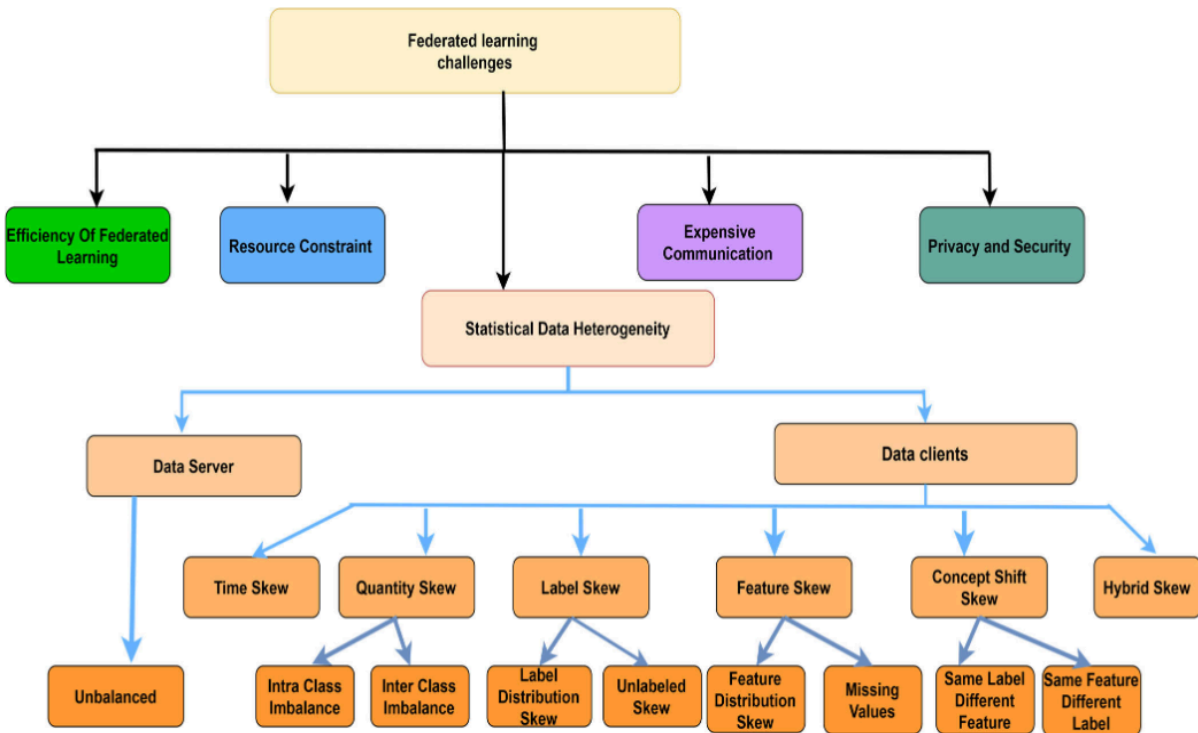
- Feature-based federated learning or heterogeneous federated learning.
- Data shares the same sample space, but different feature space

Vertical Federated Learning



Horizontal Federated Learning





The main challenges in Federated learning

Discussion was regarding the main idea about federated learning and how it works, what types of federated learning are there, and how to proceed further.

Key Points Discussed:

1. Objective of the project :

- Create a centralized and decentralized model for a dataset and compare the performance of these model
- Shuffle the features in the dataset with different local nodes and try to analyze the performance of this model using this modified dataset .

2. What dataset to use ?

As instructed by Mrinal Das sir, to start with a basic model for classification problems using logistic regression,FNN for both centralized and decentralized learning methods.

Things To be Addressed during next discussion:

- Deciding the dataset
- Check Existing works and code