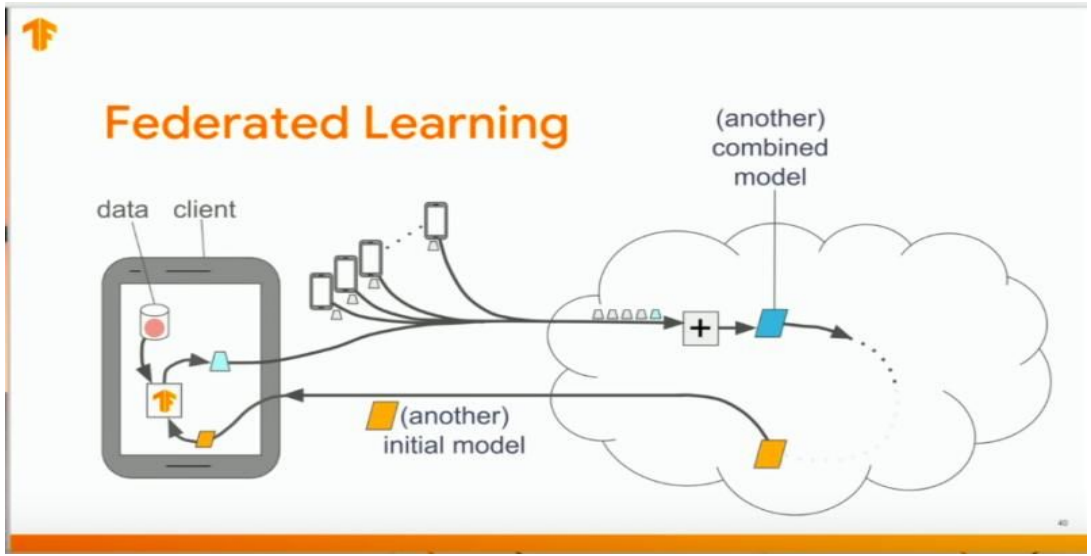
CS460 - Project Proposal

Fall Semester, 2021

Team:
Suraj Patel, SPS 18
Saswat Das, SMS 18

Topic: Exploring/Implementing Federated Learning and Possible Improvements to Existing Paradigms

- What is Federated Learning?



- Some (possible) applications of federated learning?
- What are some things to consider when discussing the efficiency and security of Federated Learning?

Relevant Papers:

1. **McMahan et al -** Communication-Efficient Learning of Deep Networks from Decentralized Data
2. **Konečný et al -** Practical Secure Aggregation for Privacy Preserving Machine Learning/Federated Learning: Strategies for Improving Communication Efficiency
3. **Rodriguez-Barroso et al -** Federated Learning and Differential Privacy: Software tools analysis, the Sherpa.ai FL framework and methodological guidelines for preserving data privacy

What we intend to do - a brief sketch

- Implement a Federated Learning algorithm (tentatively `FederatedAveraging`) after locally training models using algorithms like SGD (as suggested) on generated datasets, or even implementing some neural network(s) if possible using popular datasets (viz. FEMNIST, Shakespeare, Sentiment140, CIFAR 10) if we are feeling bold. We can use TensorFlow and SherpaAI for this if possible/needed (no promises).
- Using the (local) learning algorithm as a baseline, we can compare the results of our FL based trained-and-aggregated model with centralised learning with the same algorithm, and some other popular algorithms.
- Suggest improvements to the paradigm just implemented in terms of (1) the aggregating formula/algorithm; (2) minimising rounds of communication and computational cost per local device; and/or (3) enhancing privacy (viz. by adding differential privacy into the mix).
  Then we will see how said tweaks work in terms of these metrics and report any observed improvements/changes (fingers crossed).

Midway Target(s) and Work Division

- Successfully implement a FL model, and come up with some preliminary results.
- Suggest some tweaks and improvements to test out.
- If needed, present pertinent concepts from relevant literature.

Work Division

1. Coding/Privacy based considerations - Joint task, will divide up coding of different components fluidly as needed.
2. Data analysis and visualisation - Suraj
3. Mathematics, formulae, proofs - Saswat
4. Report writing and project website maintenance - Joint task

Expected Results

- Performance of FL based learning: almost as sound as if the learning was centralised, and maybe more nuanced.
- Either of these: Increased privacy, less communication rounds with the server, less/quicker local device computation, better aggregation.