Bjarne Pfitzner, Prof. Dr. Bert Arnrich

Algorithms

- Federated learning parameters:
  - K: # of participating clients
  - C: Fraction of clients selected per global epoch
  - T: # of global epochs
- Federated Averaging (FedAvg) [1]

\[
\text{Algorithm 1 Federated Averaging (FedAvg) [1]}
\]

1. Server executes:
2. Initialise \( \theta^0 \)
3. \( m \leftarrow \max(C \times K, 1) \)
4. for \( t = 1 \) to \( T \) do
5. \( S^t \leftarrow (\text{random set of } m \text{ clients}) \)
6. for each client \( k \in S^t \) do
7. \( \theta_k^t \leftarrow \text{ClientUpdate}(\theta^t-1) \)
8. end for
9. \( \theta^t \leftarrow \frac{1}{m} \sum_k n_k \theta_k^t \)
10. end for
11. end for
12. ClientUpdate(\( \theta \)):
13. for each local iteration do
14. for each batch \( b_k \) in client’s split do
15. \( \theta_k \leftarrow \theta_k - \eta \nabla L(h_k; \theta_k) \)
16. end for
17. end for
18. return local model \( \theta_k \)

- LoAdaBoost [2]
  - Keep track of median client loss \( L_{\text{median}} \)
  - Train locally for \( E/2 \) epochs
  - While \( L_k > L_{\text{median}} \) do:
    - Repeat local training for \( (E/2 - \text{trial}) \) epochs
    - Break if total local epochs > \( 3E/2 \)
- Co-Learning [3]
  - Increasing local epochs (ILE), using:

\[
E^t = \begin{cases} 
E^0, & \text{if } t = 1, \\
2 \times E^{t-1}, & \text{if } t > 0 \text{ and } \frac{|\theta^t - \theta^{t-1}|}{|\theta^{t-1}|} \leq \epsilon, \\
E^{t-1}, & \text{if } t > 0 \text{ and } \frac{|\theta^t - \theta^{t-1}|}{|\theta^{t-1}|} > \epsilon
\end{cases}
\]

- Cyclical learning rate using: \( \eta^t = \eta^0 \times \lambda^{E/t} \)

Results

- Dataset: MNIST (handwritten digits)
- Model: CNN
- \( K=100, \)
- \( n_k=600, \)
- \( \text{LR}(\eta)=0.01, \)
- \( \text{LR decay}=0.001, \)
- Co-Learning parameter (\( \epsilon \))=0.01

- RoC: Rounds of communication to reach 97% accuracy

- “Infinite” batch size (full local dataset) performs poorly
- LoAdaBoost has overall slower convergence than the other two
- Generally higher \( E \) lead to faster convergence

Future Work

- Increase hyperparameter search space
- Include more complex algorithms & datasets
- No independent and identically distributed data