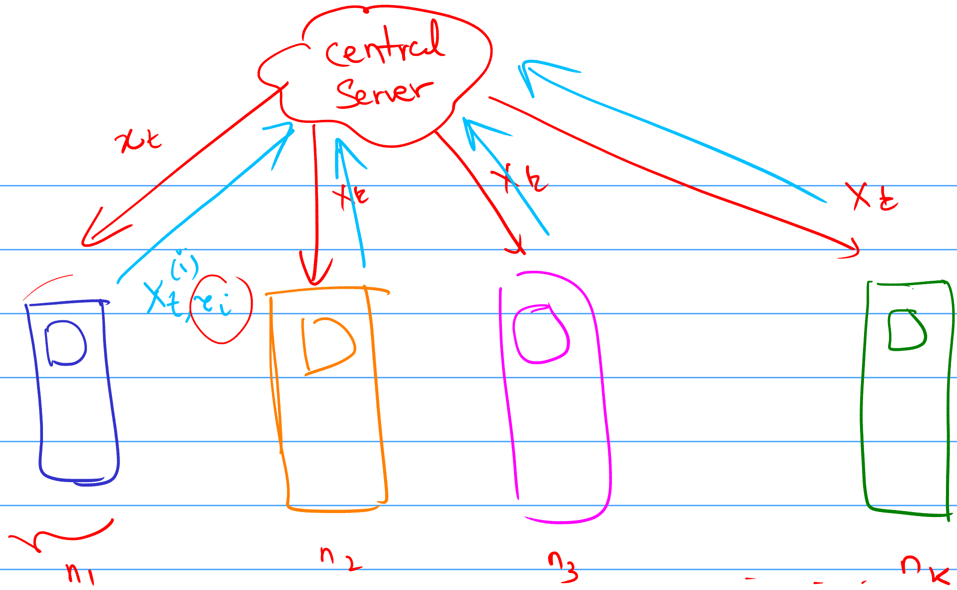


Every edge device makes few local updates $\rightarrow \tau_i$

$x^{(i)}_{t,j}$ — i^{th} edge device/client

$j \in \{0, \dots, \tau_i - 1\}$

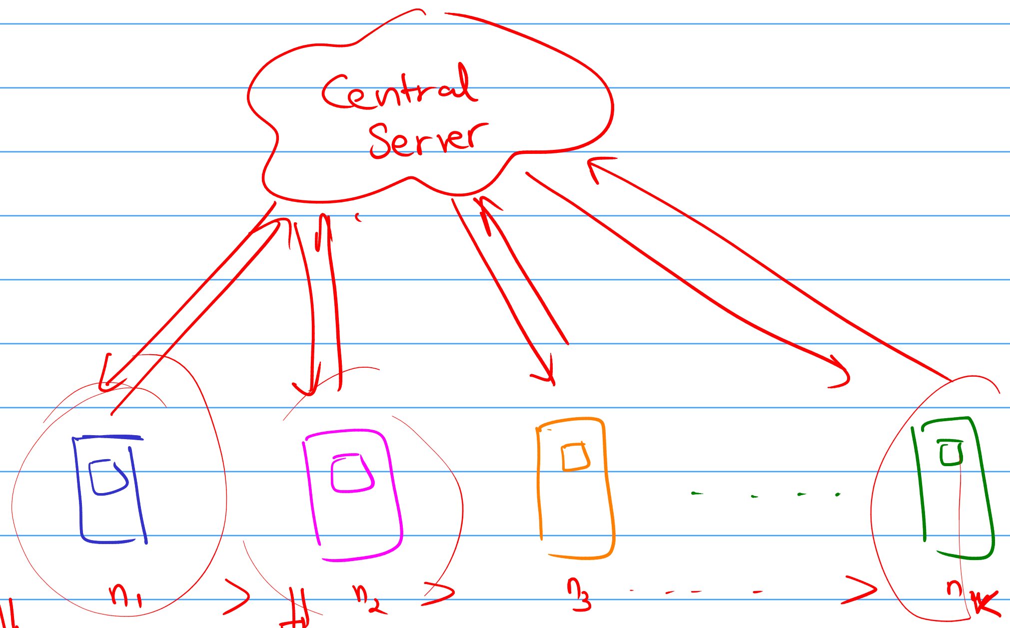
$$x^{(i)}_{t,0} = x_t$$

$$x^{(i)}_{t,j+1} = x^{(i)}_{t,j} - \eta\, g(x^{(i)}_{t,j} ; \xi_j)$$

$$x_{t+1} = \sum_{i=1}^{m} p_i \, x^{(i)}_{t,\tau_i}$$ } T communication rounds

$$p_i = \frac{n_i}{n}$$

Sources of Computational Heterogeneity in FL



If E and B are kept fixed across clients, then $\tau_1 > \tau_2$

τ_i: total number of local updates

$$\tau_i = \frac{E n_i}{B}$$

(Source of computational heterogeneity)

* If we fix τ across clients, then slow clients will take much longer to finish their updates bottlenecking each communication round.

* A quick way to eliminate the straggling effect is to fix total computational budget to T and allow clients to make as many updates as they can.

* However, faster clients will make more local updates.
[Source of computational heterogeneity]

* Variation in the hyperparameters, such as learning rates and momentum is another source of computational heterogeneity.

Objective Inconsistency Problem

Consider the setup,

We have two clients, $F_1(x) = (x-1)^2$, $F_2(x) = 2(x-5)^2$ } Local objective functions.

Global Objective function $\leftarrow$ $F(x) = 0.5F_1(x) + 0.5F_2(x)$

$$= \frac{1}{2}(x-1)^2 + (x-5)^2$$

$x^* = \frac{11}{3}$

1^{st} client.

$$x_{t,j+1}^{(1)} = x_{t,j}^{(1)} - 2\eta(x_{t,j}^{(1)} - 1)$$

$$(x_{t,j+1}^{(1)} - 1) = (1-2\eta)(x_{t,j}^{(1)} - 1)$$

$$(x_{t,j+1}^{(1)} - x_*^{(1)}) = (1-2\eta)(x_{t,j}^{(1)} - x_*^{(1)})$$

$$\Rightarrow (x_{t,\tau_1}^{(1)} - x_*^{(1)}) = (1-2\eta)^{\tau_1}(\underbrace{x_{t,0}^{(1)}}_{x_t} - x_*^{(1)})$$

$$\boxed{(x_{t,\tau_1}^{(1)} - x_*^{(1)}) = (1-2\eta)^{\tau_1}(x_t - x_*^{(1)})}$$ — 1st client

2^{nd} client:

$$x_{t,j+1}^{(2)} = x_{t,j}^{(2)} - 4\eta(x_{t,j}^{(2)} - 5)$$

$$(x_{t,j+1} - x_*^{(2)}) = (1-4\eta)(x_{t,j}^{(2)} - x_*^{(2)})$$

$$\boxed{(x_{t,\tau_2}^{(2)} - x_*^{(2)}) = (1-4\eta)^{\tau_2}(x_t - x_*^{(2)})}$$ — 2^{nd} client

At the central server (Server Update)

$$x_{t+1} = 0.5\, x^{(1)}_{t,\tau_1} + 0.5\, x^{(2)}_{t,\tau_2}$$

$$x_{t+1} = \frac{x_*^{(1)} + x_*^{(2)}}{2} + \frac{(1-2\eta)^{\tau_1}}{2}\left(x_t - x_*^{(1)}\right) + \frac{(1-4\eta)^{\tau_2}}{2}\left(x_t - x_*^{(2)}\right)$$

Let's analyze solutions to this equation

$$\tilde{x} = \frac{x_*^{(1)} + x_*^{(2)}}{2} + \frac{(1-2\eta)^{\tau_1}}{2}\tilde{x} - \frac{(1-2\eta)^{\tau_1}}{2}x_*^{(1)} + \frac{(1-4\eta)^{\tau_2}}{2}\tilde{x} - \frac{(1-4\eta)^{\tau_2}}{2}x_*^{(2)}$$

$$\tilde{x} = \frac{\left(1-(1-2\eta)^{\tau_1}\right)x_*^{(1)} + \left(1-(1-4\eta)^{\tau_2}\right)x_*^{(2)}}{\left(1-(1-2\eta)^{\tau_1}\right) + \left(1-(1-4\eta)^{\tau_2}\right)}$$

$$\lim_{\eta \to 0} \tilde{x} = \frac{\tau_1 x_*^{(1)} + 2\tau_2 x_*^{(2)}}{\tau_1 + 2\tau_2}$$

Depending on the number of local updates τ_1 and τ_2, this point can be arbitrarily different from the intended global minimum.

In FedAvg algorithm

$$F(x) = \sum_{i=1}^{m} \frac{n_i}{n} F_i(x)$$

Global objective function to be minimized

Mismatched objective function.

$$\tilde{F}(x) = \sum_{i=1}^{m} \frac{n_i \tau_i}{\sum_{i'=1}^{m} n_{i'}\tau_{i'}} F_i(x)$$

General Update Rule

NeurIPS '20

FedNova $\leftarrow$ $\boxed{x_{t+1,0} = x_{t,0} + \tau_{eff} \sum_{i=1}^{m} w_i \left(-\eta\, d_t^{(i)}\right)}$

Assume every agent performs just one local update.

$$x_{t+1,0} = x_{t,0} - \eta \sum_{i=1}^{m} w_i\, g(x_{t,0}; \xi_i)$$

Assume agent i performs two local updates.

$$x_{t,1}^{(i)} = x_{t,0}^{(i)} - \eta\, g(x_{t,0}^{(i)}; \xi_0)$$

$$x_{t,2}^{(i)} = x_{t,1}^{(i)} - \eta\, g(x_{t,1}^{(i)}; \xi_1)$$

$$= x_{t,0}^{(i)} - \eta g(x_{t,0}^{(i)}; \xi_0) - \eta g(x_{t,1}^{(i)}; \xi_1)$$

$$x_{t,2}^{(i)} = x_{t,0}^{(i)} - \eta \underbrace{\sum_{j=0}^{1} g(x_{t,j}^{(i)}; \xi_j)}_{d}$$

After τ_i local updates $\leftarrow$ $\boxed{x_{t,\tau_i}^{(i)} = x_{t,0}^{(i)} - \eta\left(\sum_{j=0}^{\tau_i-1} g(x_{t,j}^{(i)}; \xi_j)\right)}$

$$x_{t+1,0} = \sum_{i=1}^{m} p_i\, x_{t,\tau}^{(i)} \qquad \text{with } \sum p_i = 1$$

$$= \sum_{i=1}^{m} p_i \left(x_{t,0} - \eta\left(\sum_{j=0}^{\tau_i-1} g(x_{t,j}^{(i)}; \xi_j)\right)\right)$$

$$\Rightarrow \boxed{x_{t+1,0} = x_{t,0} + \sum_{i=1}^{m} p_i \left(-\eta\left(\sum_{j=0}^{\tau_i-1} g(x_{t,j}^{(i)}; \xi_j)\right)\right)}$$

* In vanilla FedAvg :

$$x_{t,\tau_i}^{(i)} = x_{t,0} - \eta \underbrace{\left(\sum_{j=0}^{\tau_i - 1} g(x_{t,j}^{(i)}) \right)}_{\substack{\text{Accumulated gradient} \\ \Delta_t^{(i)}}}$$

* We define a normalized gradient :

$$d_t^{(i)} = \frac{\sum_{j=0}^{\tau_i-1} a_j^{(i)} g(x_{t,j}^{(i)})}{\sum_{j=0}^{\tau_i-1} a_j^{(i)}} = \frac{G_t^{(i)} a^{(i)}}{\|a^{(i)}\|_1}$$

here $a^{(i)}$ is a non-negative vector

* If the agents share normalized gradient,

$$x_{t+1,0} = x_{t,0} + \sum_{i=1}^{m} w_i \left(-\eta \frac{G_t^{(i)} a^{(i)}}{\|a^{(i)}\|_1} \right)$$

* In case of vanilla FedAvg:

$$a^{(i)} = \underbrace{[1 \; \ldots \; 1]}_{\tau_i}$$

$$\boxed{\|a^{(i)}\|_1 = \tau_i}$$

$$x_{t+1,0} = x_{t,0} + \sum_{i=1}^{m} \frac{w_i}{\tau_i} \left(-\eta \sum_{j=0}^{\tau_i-1} g(x_{t,j}^{(i)}) \right)$$

If $\boxed{w_i = p_i \tau_i}$ $\times$ $\boxed{w_i = \frac{p_i \tau_i}{\sum_{i'=1}^{m} p_{i'} \tau_{i'}}}$

$$\boxed{\tau_{eff} = \sum_{i=1}^{m} p_i \tau_i}$$

FedAvg

$$\boxed{\|a^{(i)}\|_1 = \tau_i}$$

In FedAvg:

$$a^{(i)} = [\underbrace{1, \dots\dots 1}_{\tau_i}]$$

$$w_i = \frac{p_i \tau_i}{\sum_{i'=1}^{m} p_{i'} \tau_{i'}}$$

$$\tau_{eff} = \sum_{i=1}^{m} p_i \tau_i$$

$$\longrightarrow d_i^{(t)} = \frac{G_t^{(i)} a^{(i)}}{\|a^{(i)}\|_1}$$

FedNova:

* Client i normalizes the total accumulated update $\Delta_t^{(i)}$ by the number of updates τ_i

$$-\eta d_t^{(i)} = \frac{\Delta_t^{(i)}}{\tau_i}$$

$$\sum_{i=1}^{m} p_i \frac{\Delta_t^{(i)}}{\tau_i}$$ and multiplies by $\tau_{eff} = \sum_{i=1}^{m} p_i \tau_i$

$$X_{t+1,0} = X_{t,0} + \tau_{eff} \sum_{i=1}^{m} p_i \frac{\Delta_t^{(i)}}{\tau_i}$$

Fairness

Measures of fairness in FL

1. Variance: $\mathrm{Var}(F_1(x), \dots, F_K(x)) = \frac{1}{K} \sum_{i=1}^{K} \left(F_i(x) - \frac{\sum_{i'=1}^{K} F_{i'}(x)}{K}\right)^2$

2. Entropy: $-\sum_{i=1}^{K} \frac{F_i(x)}{\sum_{i'=1}^{K} F_{i'}(x)} \log\left(\frac{F_i(x)}{\sum_{i'=1}^{K} F_{i'}(x)}\right)$

3. Jain's Fairness Index: $\frac{\left(\sum_{i=1}^{K} F_i(x)\right)^2}{K \sum_{i=1}^{K} (F_i(x))^2}$