we have K clients, random fraction C is selected

→ global model sent → perform computation on edge

↓

send update

P_k = set of indexes of data points on client k

$n_k = |P_k|$

Let $f(w)$ be our objective function that we minimize

$$f(w) = \sum_{i=1}^{N} f_i(w) \quad , \text{here } f_i(w) = \ell(x_i, y_i; w_i)$$

In our case

$$f(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w) \quad , \quad F_k(w) = \frac{1}{n_k} \sum_{i \in P_k} f_i(w)$$

Let $C = 1$,

→ each client compute $g_k = \nabla F_k(w_t)$

→ Then after then for the central server Update

$$\Rightarrow w_{t+1} \rightarrow w_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} g_k$$

Algorithm 1 FederatedAveraging. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:

initialize w_0

for each round $t = 1, 2, \dots$ **do**

$m \leftarrow \max(C \cdot K, 1)$ select m out of K

$S_t \leftarrow$ (random set of m clients)

for each client $k \in S_t$ **in parallel do**

$w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$ update each client w_t^k

$m_t \leftarrow \sum_{k \in S_t} n_k$ Total No of data points for that epoch

global w $w_{t+1} \leftarrow \sum_{k \in S_t} \frac{n_k}{m_t} w_{t+1}^k$ *// Erratum*[4]

ClientUpdate(k, w): *// Run on client k*

$\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)

for each local epoch i from 1 to E **do**

for batch $b \in \mathcal{B}$ **do**

$w \leftarrow w - \eta \nabla \ell(w; b)$

return w to server

Now, there exist another algorithm that can replace SGD here and that is

Just have f as the objective function to optimize.

Nesterov's Accelerated GD

$f : \mathbb{R}^d \to \mathbb{R}$

$$f = \frac{1}{2} x^T A x - b^T x, \quad A := \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}, \quad b := \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Here f is $\overset{m}{\text{strongly convex}}$ and $\overset{M}{\text{smooth}}$

M and m

☆ Here, Define: f is m-strongly convex if, there exists $m > 0$ such that $\forall\ x, y \in \mathbb{R}^d$ and EVERY $\lambda \in [0,1]$

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{m}{2} \lambda (1-\lambda) \|x-y\|^2$$

We will use its equivalent characterizations to prove SHOW

$$\|\nabla f(x) - \nabla f(y)\| \geq m \|x - y\|$$

Now for $\nabla f(x)$, $\Rightarrow \nabla f(x) = \frac{1}{2}(A + A^T)x - b^T$

As A is symmetric, $\Rightarrow \nabla f(x) = Ax - b^T$

$\therefore \Rightarrow \nabla f(x) - \nabla f(y) = Ax - b^T - (Ay - b^T)$
$= A(x-y)$

$\Rightarrow \|\nabla f(x) - \nabla f(y)\| = \|A(x-y)\|$

Regarding A, $\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}$, lets consider for p and $q \neq 0$ simultaneously

$\Rightarrow [p \ q] \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix}$

$\Rightarrow [4p+q \quad p+3q] \begin{bmatrix} p \\ q \end{bmatrix}$

$\Rightarrow 4p^2 + pq + pq + 3q^2$

$\Rightarrow 4p^2 + 2pq + 3q^2$

$\Rightarrow \underbrace{(p+q)^2 + 3p^2 + 2q^2}_{>0 \ +ve}$

Hence A is $+ve$ definite

$\therefore \quad A \geq \lambda_{min}(A) I \rightarrow (1)$ where $\lambda_{min}(A)$ is the smallest eigenvalue

For eigenvalues, $(A - \lambda I)x = 0$

$|A - \lambda I| = 0$

$\Rightarrow \begin{vmatrix} 4-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = 0$

$\Rightarrow (4-\lambda)(3-\lambda) - 1 = 0$

$\Rightarrow 12 - 7\lambda + \lambda^2 - 1 = 0$

$\Rightarrow \lambda^2 - 7\lambda + 11 = 0$

$\lambda = \dfrac{7 \pm \sqrt{49-44}}{2} = \dfrac{7 \pm \sqrt{5}}{2}$

$\lambda_{min} = \dfrac{7-\sqrt{5}}{2}$

Using (1) $\Rightarrow \|\nabla f(x) - \nabla f(y)\| \geq \underbrace{\lambda_{min}}_{\downarrow m} \|x - y\|$

$\boxed{m = \dfrac{7-\sqrt{5}}{2}}$

Hence f is m-strongly convex

★ Define: $f : \mathbb{R}^d \to \mathbb{R}$ satisfies

$$\| \nabla f(x) - \nabla f(y) \|_2 \leq M \| x - y \|_2$$

$M > 0 \Rightarrow$ M-smooth.

From before

$$\Rightarrow \| A(x-y) \|$$

Since $A \leq \lambda_{max}(A) I$

$\therefore \Rightarrow \| A(x-y \| \leq \| \lambda_{max}(A) I \| \; \| x-y \|$

$$\leq \lambda_{max}(A) \; \| x-y \|$$

Hence proved.

$M = \lambda_{max}(A)$

$$\boxed{M = \frac{7+\sqrt{5}}{2}}$$