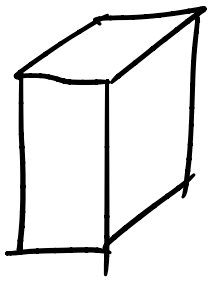
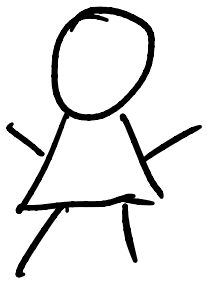
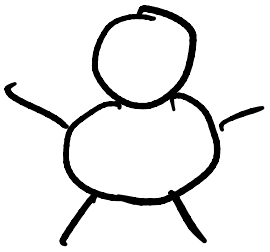
Centralized learning system.



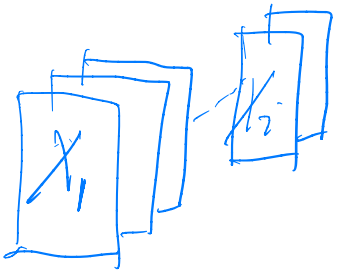
server



Alice



Bob



* The para. of the linear face recog. model is

$$w = [w; b] \in R^{MN+1}.$$

* The objective func. is

$$\mathcal{L}(w; X) = \sum_{i=1} \frac{1}{2}(y_i - w^T x_i)^2 + \eta \|w\|_2^2$$

$$x_i = [vec(X_i); 1] \in R^{MN+1}$$

* To find w that minimizes $\mathcal{L}(w; X)$.
the server conducts stochastic gradient descent (SGD)

$$w^t = w^{t-1} - \alpha \frac{\nabla \mathcal{L}(w; X)}{\nabla w}$$

$$\sum_i x_i (w^T x_i - y_i) + \lambda w$$

△ △

△: requires both Alice's & Bob's private data.

To address this issue, we observe that

$$\frac{\nabla \mathcal{L}(w; X)}{\nabla w} = \sum_i x_i (w^T x_i - y_i) + 2w$$

$$= \sum_i x_i^{Alice} \left(w^T x_i^{Alice} - y_i^{Alice} \right) + 2w$$

$$+ \sum_j x_j^{Bob} \left(w^T x_j^{Bob} - y_i^{Bob} \right) + 2w$$

$$= \text{Alice's local gradient} \quad (g_A)$$

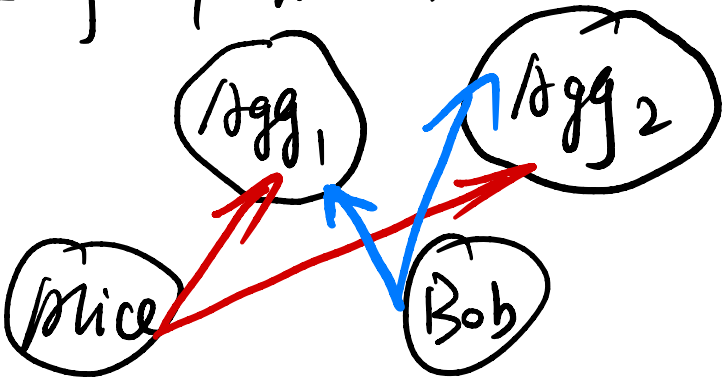
$$+ \text{Bob's local gradient} \quad (g_B)$$

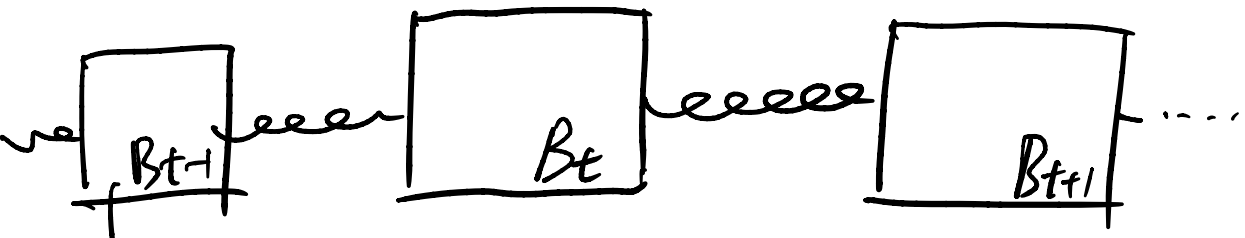
Thus, $\frac{\nabla \mathcal{L}(w; X)}{\nabla w} = g_A + g_B$

$$w^t = w^{t-1} - \alpha \left(g_A^{t-1} + g_B^{t-1} \right)$$

Put on Blockchain.

Toy example of 4 nodes.





① download w^{t-1} from B_{t-1} (Alice, Bob, Agg1, Agg2)

② Alice compute g_A^{t-1} locally

Bob ... g_B^{t-1} locally.

③ Alice and Bob broadcast g_A^{t-1}, g_B^{t-1} to the aggregation nodes in the network.

④ Agg1 and Agg2 competes with each other to solve a puzzle to win the right to do the aggregation and write the aggregated gradient, i.e.,

$$w^t = w^{t-1} - \alpha \left(g_A^{t-1} + g_B^{t-1} \right)$$

to Block B_t

⑤ repeat ① – ④, until converge.

(At time step $t=1$. w^0 can be randomized / initialized).