NOW THE MATH OF

DECENTRALIZED LEARNING

- A bit of statistical learning
- the consensus problem
- Distributed Gradient
- Issues with networks
- Adversarials
- Federated and friends

$$\min_{x} f(x)$$

weights NN

error

$$(x, y)$$

features

Labels

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \ell\left(f_{\theta}(x_i), y_i\right)$$

NN

imag

loss function

weights

$$\min_{x} f(x)$$

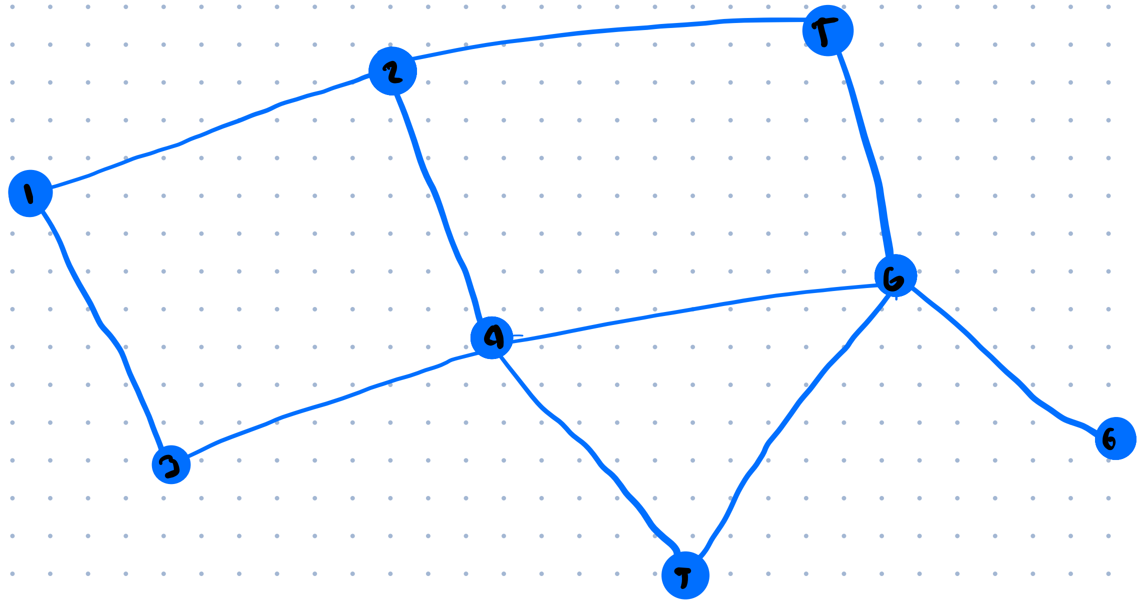
$$\min_{x} \frac{1}{n} \sum_{i=1}^{n} f_i(x)$$

$$\min_{x} \frac{1}{n} \sum_{i=1}^{n} f_i(x)$$

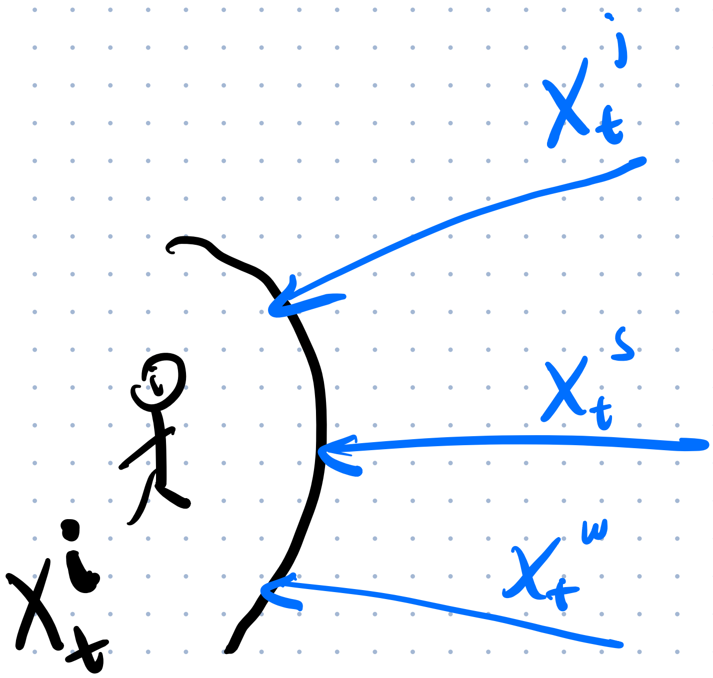
# of training data points.

THE CONSENSUS ALG.

X_t^i



$$\lim_{t \to \infty} X_t^i = \frac{1}{n} \sum_{j=1}^{n} X_0^j \quad \text{for all}$$



$$X_{t+1}^i = \text{AVE}\begin{pmatrix}\text{whoever} \\ i \\ \text{was} \end{pmatrix}$$
talking with

$$\frac{1}{2} a + \frac{1}{2} b$$

$$w_1 a + w_2 b$$

$$W_1 + W_2 = 1$$

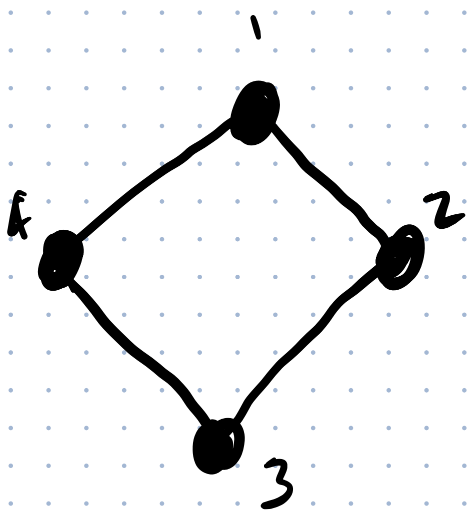
$$w_1, W_2 \geq 0$$

weighted average of $\{y_i\}$ $= \sum_{i=1}^{n} w_i y_i$

$$X^{i}_{t+1} = \sum_{j=1}^{n} \underline{\underline{w_{ij}}}\, X^{j}_{t}$$

w_{ij} = weight that agent i assigns to j

$$W = [w_{ij}] = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}$$



$$W = \begin{bmatrix} \boxed{1/3} & \boxed{1/3} & 0 & \boxed{1/3} \\ \boxed{1/3} & \boxed{1/3} & \boxed{1/3} & 0 \\ 0 & \boxed{1/3} & \boxed{1/3} & \boxed{1/3} \\ \boxed{1/3} & 0 & \boxed{1/3} & \boxed{1/3} \end{bmatrix}$$

$$X_{t+1}^{i} = \sum_{j=1}^{n} w_{ij} X_t^{j}$$

$$\begin{bmatrix} X_{t+1}^{1} \\ X_{t+1}^{2} \\ \vdots \\ X_{t+1}^{n} \end{bmatrix} = \begin{bmatrix} & & \\ & W & \\ & & \end{bmatrix} \begin{bmatrix} X_t^{1} \\ X_t^{2} \\ \vdots \\ X_t^{n} \end{bmatrix}$$

$$\begin{aligned} X_{t+1} &= W X_t \\ &= W(W X_{t-1}) \\ &= W \cdot W \cdots W X_0 \\ &= W^t X_0 \end{aligned}$$

$$X_{t+1} = W^t X_0$$

$$W^t = \frac{1}{n} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} X_0$$

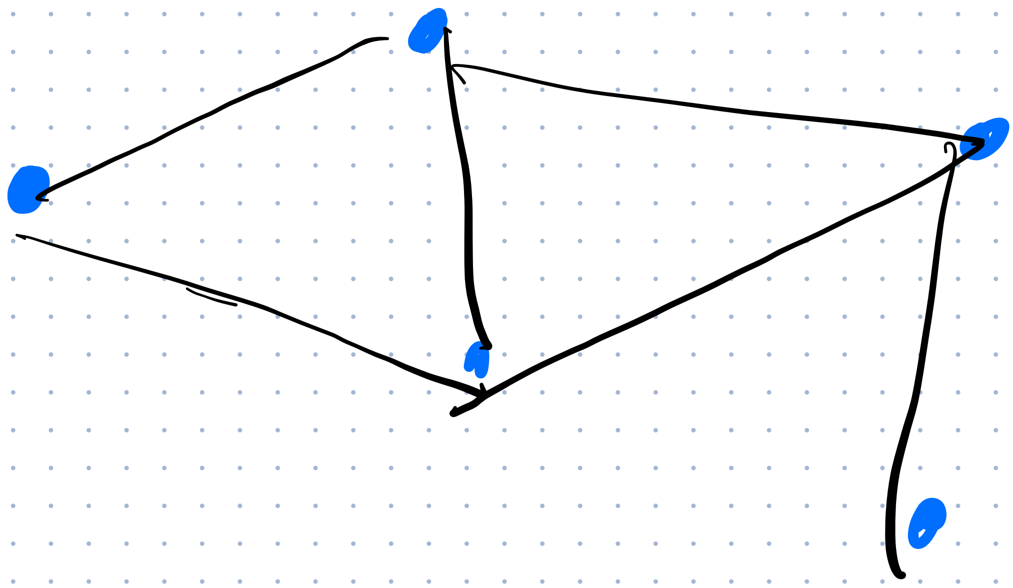
$$\lim_{t \to \infty} X_t^i = \frac{1}{n} \sum_{j=1}^{n} X_0^j$$

$$\begin{bmatrix} 1/n & \cdots & 1/n \\ \vdots & & \\ 1/n & \cdots & 1/n \end{bmatrix} \begin{bmatrix} \\ X_0 \\ \\ \end{bmatrix}$$

* Key idea

$$W^t \longrightarrow \begin{bmatrix} \frac{1}{n} \end{bmatrix}$$

* Self weigths are positive.
* Weights add up to 1. and non negative.
* Network is connected.



$$X_{t+1}^{i} = \sum^{n} w_{ij} X_t^{j}$$

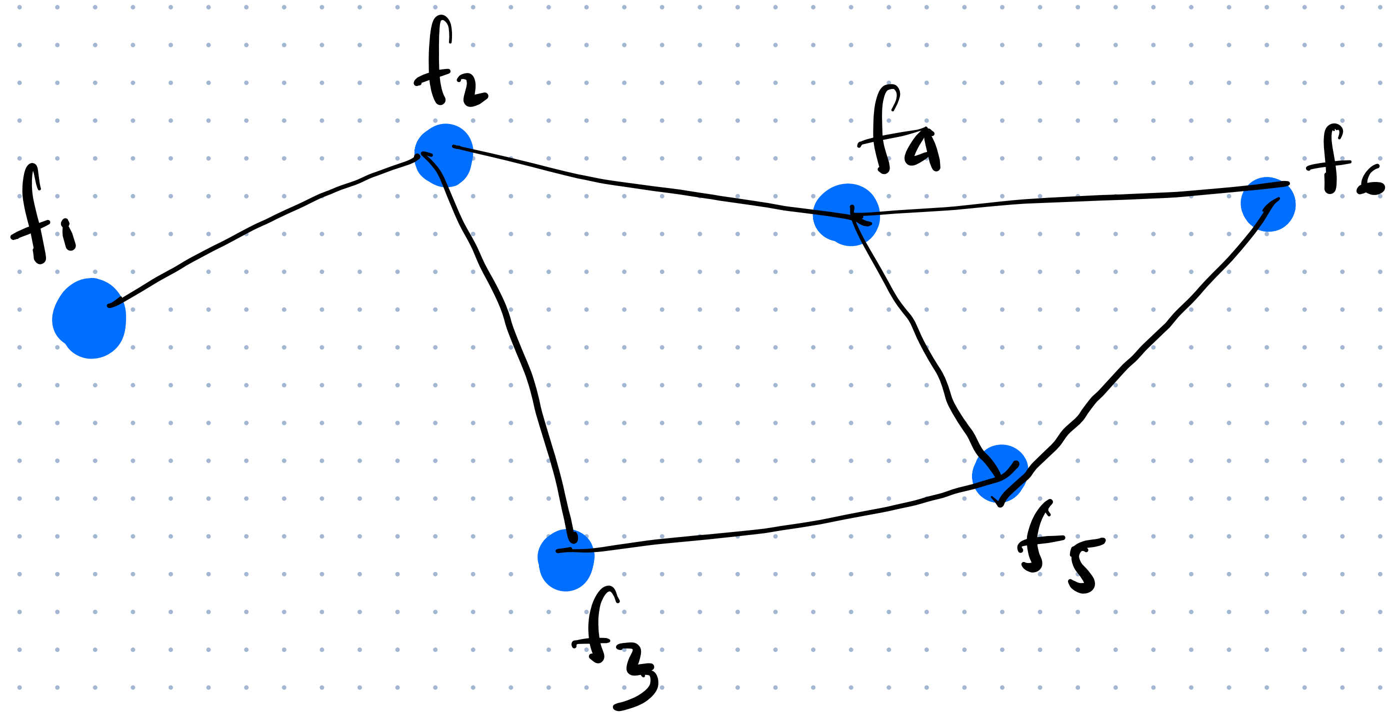
$$X_{t+1} = X_t - \alpha \nabla f(X_t)$$

Distributed Gradient Method

(Nedić, Ozdaglar, 2006)

$$X^i_{t+1} = \sum_{j=1}^{n} w_{ij} X^j_t - \alpha \nabla f_i(X^i_t)$$

$$\min \frac{1}{n} \sum_{i=1}^{n} f_i(x)$$



1. Consensus will happen. * under certain assumpt...)

$$\lim_{t \to \infty} X_t^i = c \quad \text{for all}$$

$$C = \arg\min \frac{1}{n} \sum_{j=1}^{n} f_j(x)$$

Time it takes to converge

$$O\left(\text{"time it takes centralizes"} \times \text{"Network topology"} \right)$$