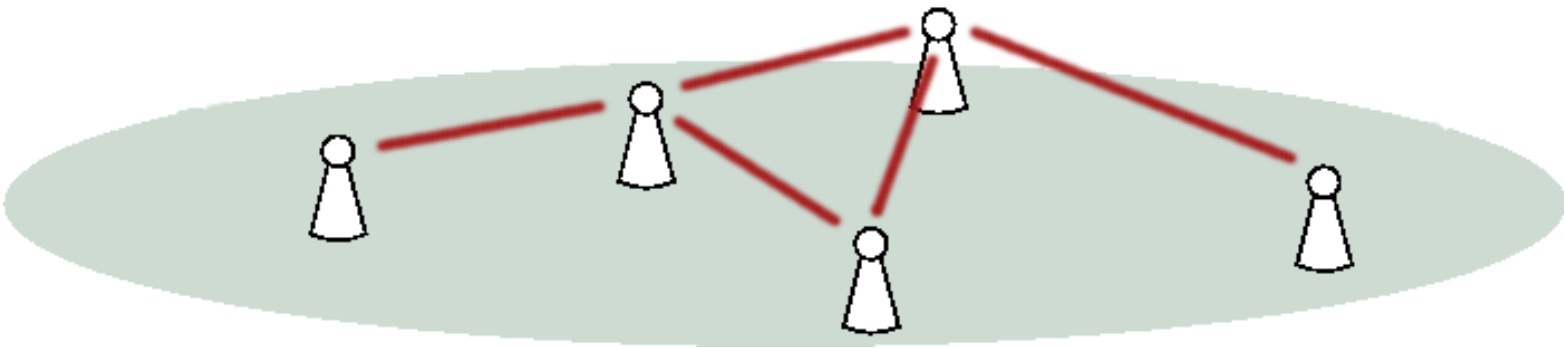
Context

Consider a network composed of N agents



- Agents process *local* data
- Agents cooperate to estimate some *global* parameter

A regression example

Data set formed by n samples (X_i, Y_i) $(i = 1 \dots n)$

- Y_i = variable to be explained
- X_i = explanatory features

Looking for a model.

Linear regression example:

$$\min_x \sum_{i=1}^{n} \| Y_i - x^T X_i \|^2$$

A regression example

Data set formed by n samples (X_i, Y_i) $(i = 1 \dots n)$

- Y_i = variable to be explained
- X_i = explanatory features

Looking for a model.

$$\min_x \sum_{i=1}^{n} \ell(x^T X_i, Y_i) + r(x)$$

A regression example

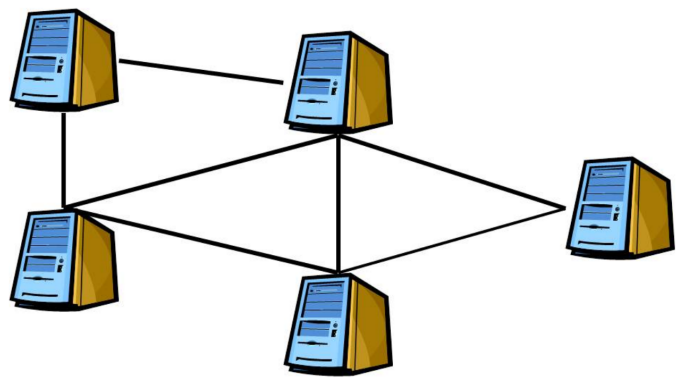
Data set formed by n samples (X_i, Y_i) $(i = 1 \dots n)$

- Y_i = variable to be explained
- X_i = explanatory features

Looking for a model.

$$\min_x \sum_{i=1}^{n} \ell(x^T X_i, Y_i) + r(x)$$

Distributed processing: the problem is separable



$$\min_x \sum_{v \in V} \sum_{i=1}^{n} \ell(x^T X_{i,v}, Y_{i,v}) + r(x)$$

[Boyd'11, Agarwal'11]

Formally

$$\min_x \sum_{v \in V} f_v(x)$$

- $\mathcal{G} = (V, E)$ is the *graph* modelling the network
- f_v is the cost function of agent v

Difficulty : $\sum_v f_v$ is nowhere observed.

Methods : from distributed gradient algorithms to advanced proximal methods

Common principle :

1. process local data
2. exchange information with neighbors
3. iterate.

Key issues

- Distribute cutting-edge optimization algorithms (eq. primal-dual methods, fast-admm, etc.)
- Include stochastic perturbations:
  - On-line algorithms
  - Asynchronism
- Investigate specific ML application, e.g. ranking