Датасет с разными фичами:

$$D = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1k} \\ \vdots & \vdots & \dots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nk} \end{pmatrix}, D \in \mathbb{R}^{n \times k}$$

Таргет:

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, b \in \mathbb{R}^n$$

На каждом клиенте считается

$$f_i = d_i x_i$$

Где

- $d_i$ - фичи, которые доступны на клиентском сервере
- $x_i$ - параметры модели клиента

На основном сервере ответы с клиентов агрегируются и мы получаем

$$f = DX - b$$

Функция потерь, которую мы хотим оптимизировать:

$$L(f) = \|DX - b\|_2^2$$

$$\frac{\partial L(f)}{\partial x_i} = 2d_i^T(DX - b)$$

Аггрегирование ответов с клиентов происходит следубщим образом:

$$DX = \sum_{i=1}^{k} d_i x_i$$

$$\left.\begin{matrix} d_i x_i \in \mathbb{R}^n \\ d_i \in \mathbb{R}^n \end{matrix}\right| \Rightarrow x_i \in \mathbb{R}$$

После вычисления $(DX - b)$ рассылается ко всем клиентам. Для оптимизиции отправки результата аггрегированных вычислений может быть применен алгоритм квантизации (Например RandK). И тогда результат, который отправляется:

$$q(DX - b) = Q(\sum_{i=1}^{k} d_i x_i - b)$$

Либо

$$q(DX - b) = Q(\sum_{i=1}^{k} Q(d_i x_i) - b)$$

При получении $q(DX - b)$ Градиент на на клиентских серверах вычитывается как:

$$\frac{\partial L(f)}{\partial x_i} = 2d_i^T q(DX - b)$$

И веса всех клиентских моделей обновляются в соответствии с этим градиентом:

$$x_i^{k+1} = x_i - 2\gamma d_i^T q(DX - b)$$