EE6367: Topics in Data Storage and Communications **2023**

Lecture 16: 16 October 2023

Instructor: Shashank Vatedka *Scribe: Gautam Singh*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

16.1 Randomized Correlated Quantization

Suppose that each of m users has k bits to transmit, that is, $Y_i \in \{0,1\}^k$. Let $K \triangleq 2^k$. We first generate $c_1 \sim \text{Unif}\left[-\frac{1}{k}, 0\right]$, and define

$$c_i = c_1 + (i-1)\,\beta, \tag{16.1}$$

where $\beta = \frac{K+1}{K(K-1)}$. Note that from (16.1),

$$c_K = c_1 + 1 + \frac{1}{K} \sim \text{Unif}\left[1, 1 + \frac{1}{K}\right]. \tag{16.2}$$

Again, let π be a random permutation as before. Now, c_1 and π constitute the shared randomness among the users. At each user U_i, define

$$z_i \triangleq \frac{x_i}{\beta} \tag{16.3}$$

$$c_i' \triangleq \max_{c_j < z_i} c_j \tag{16.4}$$

$$\hat{x_i} = Q_i\left(x_i\right) \triangleq c_i' + \beta \mathbb{1}_{\left\{\frac{\pi}{m} + \gamma_i < z_i\right\}} \tag{16.5}$$

The estimate of the empirical mean at the server is

$$\hat{\bar{x}} = \frac{1}{m} \sum_{i=1}^{m} \hat{x_i} \tag{16.6}$$

For the multidimensional case $\mathbf{x} \in \mathbb{R}^d$, then we first apply the random rotation followed by the above scheme. The amount of shared randomness is $\mathcal{O}\left(dm \log m\right) + \mathcal{O}\left(d\right) + \mathcal{O}\left(dm\right) = \mathcal{O}\left(dm \log m\right)$ bits, assuming that a structured random rotation matrix is used. It is an open problem to reduce the amount of shared randomness while keeping the same MSE (asymptotically).

16.2 Learning in a Distributed Setting

Parameter	**Datacenter Learning**	**Cross-Silo Federated Learning**	**Cross-Device Federated Learning**
Dataset	Not private, iid Same amount per node	Private, not iid Different amount per user	Private, not iid Different amount per silo
Number of Users	10-100	10-100	Millions
Communication Constraints	Yes, but not at the cost of communication.	Yes	Yes
Client Availability	Yes, at all times	Yes, at all times	Not always, only subsets
Client Reliability	Reliable	Reliable	Unreliable, dropouts common, susceptible to attacks

Table 16.1: Methods of distributed learning