Zeno: distributed synchronous SGD that

- tolerates an arbitrary number of malicious workers
- provides convergence guarantees for non-convex problems

Goal: converge under attacks/failures, regardless of false negative

<table>
<thead>
<tr>
<th>Feature</th>
<th>Prev.</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tolerates a majority of malicious workers</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Considers the progress of optimization</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Tolerates stealth adversary (empirically)</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Byzantine-tolerant SGD

$m$ workers, distributed SGD:

1: Pull

2: Gradient Computation

3: Push

4: Aggregation

Byzantine gradient

Correct gradient

Server

Honest Worker

Byzantine Worker
Main Idea & Results

★ Sort $g_i(x), i \in [m]$ by the **Stochastic descent score**:

**Definition**

Stochastic descent score of any update $u$:

$$Score_{\gamma, \rho}(u, x) = f_r(x) - f_r(x - \gamma u) - \rho \|u\|^2,$$

$f_r(x)$: unbiased estimator of the loss $F(x)$, for validation.

★ Zeno: filter the $b$ “worst” gradients $\frac{1}{m-b} \sum_{i=1}^{m-b} \tilde{v}(i), b > q$.

★ Convergence after $T$ iterations:

$$\sum_{t=0}^{T-1} \frac{\mathbb{E}\|\nabla F(x^t)\|^2}{T} \leq O\left(\frac{1}{\sqrt{T}}\right) + O\left(\frac{(b-q+1)(m-q)}{(m-b)^2}\right).$$