

## Our new aggregation rule: Bulyan **Bulyan**: description

Bulyan is a "composite" aggregation rule.

Let: A a Byzantine-resilient aggregation rule *f* the # of Byzantine workers to support  $n_A$  the # of workers A needs to support f

Then: Bulyan(A) needs  $n \ge n_A + 2f$  workers

- The 1<sup>st</sup> step of Bulyan works as follow:
  - R = {received gradients} S = {} # selected gradients





For full model descriptions, please see the paper.

• Fading learning rate:  $\eta(epoch) = \eta_0 \frac{1}{epoch + r}$ 

This step recursively uses A to select a majority (i.e.  $\geq 2 f + 1$ ) of non-Byzantine gradients.

- The 2<sup>nd</sup> step builds the output gradient. The selected gradients *S* is a matrix of *d* rows and  $(n - n_A)$  columns. Let  $\beta = n - n_A - 2f$ .
- Then each coordinate *i* of the output gradient is equal to the *average* of the  $\beta$  closest values, in row *i* of *S*, to the **median** of these values.

**ICML 2018** 

 $10^{-4}$ • L2-regularization: • Bulyan used with: A = Krum

## References

- Blanchard, Peva, El Mhamdi, El Mahdi, Guerraoui, Rachid, and Stainer, Julien. Machine learning with adversaries: Byzantine tolerant gradient descent. In Advances in Neural Information Processing Systems 30, pp. 118–128. Curran Associates, Inc., 2017.
- Cohen, Michael B, Lee, Yin Tat, Miller, Gary, Pachocki, Jakub, and Sidford, Aaron. Geometric median in nearly linear time. In *Proceedings* of the forty-eighth annual ACM symposium on Theory of Computing, pp. 9-21. ACM, 2016.
- Rousseeuw, Peter J. Multivariate estimation with high breakdown [3] point. Mathematical statistics and applications, 8:283–297, 1985.

## July 10-15, Stockholm, Sweden