

## SUPPLEMENTARY MATERIAL FOR SYSML 2020 REBUTTAL

### PROGRAM CHAIR’S APPROVAL

Dimitris Papailiopoulos

11/17/19

DP

Re: [Systems and ML 2020] Author Feedback Period  
Extended to Nov 20

[Hide](#)

To: [REDACTED]

Cc: Systems and ML 2020 Program Chairs

Dear [REDACTED],

Yes you may use anonymous links and content as long as your identity is hidden. However, please note that the reviewers are not expected to use material from external links during their assessment.

Best,  
Dimitris

### APPENDIX 1: CIFAR10 WITH VGG11

We adopt a VGG11 architecture (Simonyan & Zisserman, 2014) and simulate the training of CIFAR10 (Krizhevsky et al., 2009) dataset using time measurements on Raspberry Pi 4’s. The data are randomly partitioned to 5 clients. Each of the clients updates the model with a mini batch of 20 data, and 5 such mini batches in a “federation”. The samples are 1% (500) data randomly drawn from the training set. We use an SGD (LR = 0.35) optimizer for this simulation. The results are presented in Figure A.

Since VGG11 is primarily convolutional, and we are not able to find implementation of sparse convolutional kernels in any existing libraries, the time reduction mainly comes from communication, which is less evident than the reduction in LeNet-300-100 architecture. Nevertheless, we still find that Figure A(a) agrees with Figure 10, and Figure A(b) agrees with Figure 11 to a large extent.

### APPENDIX 2: FEMNIST WITH CONV-FEMNIST AND SAMPLED CLIENTS

The experimental setting is the same as in Section 5.2 (see Table 1), except that, instead of sampling all clients, here we train with the data sampled from 10 out of 193 clients. The results are presented in Figure B.

Because we sample from 10 out of 193 clients, the learning curve is slowed down. Nevertheless, the results are as expected: Figure B(a)(b) are in accord with Figure 10 and 11.

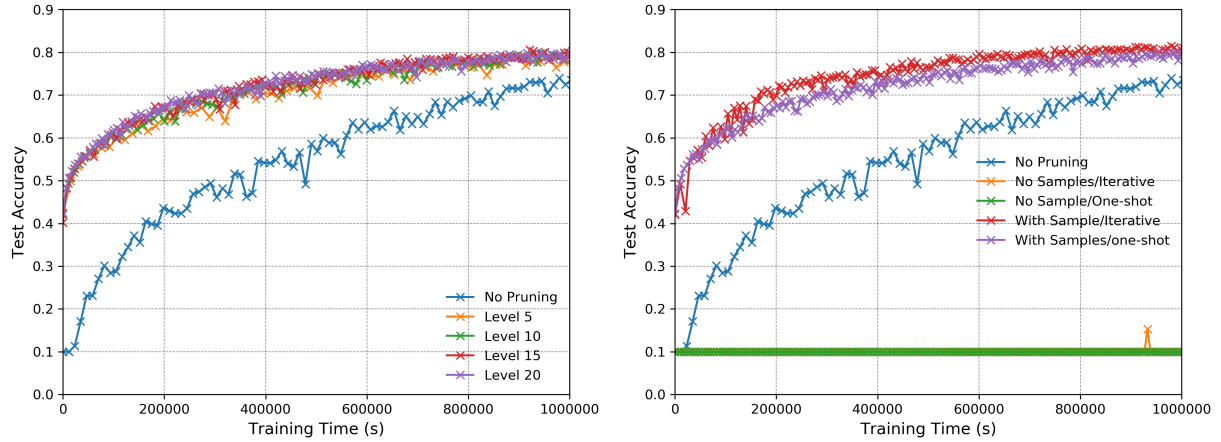
### APPENDIX 3: FEMNIST WITH LENET-300-100

The experimental setting is the same as in Section 5.2 (see Table 1), except that we use a much smaller LeNet-300-100 (see Table 1 for details) for training. The results are presented in Figure C.

Since we consider the LeNet-300-100 architecture underparameterized for the FEMNIST dataset. Figure C proved that even with an underparameterized model, our claims still hold. The only difference is the lower upper bound accuracy (the accuracy of the converged original model) compared with an overparameterized model.

### REFERENCES

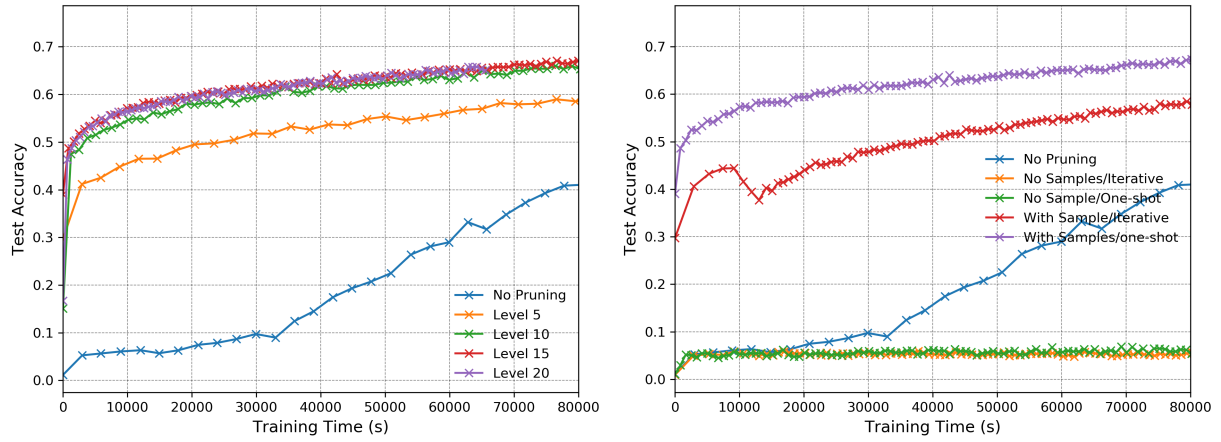
- Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.
- Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.



(a) The 4 Possible Pruning Cases

(b) Comparing Sample-based, One-shot Approach at Pruning Level 5, 10, 15, 20

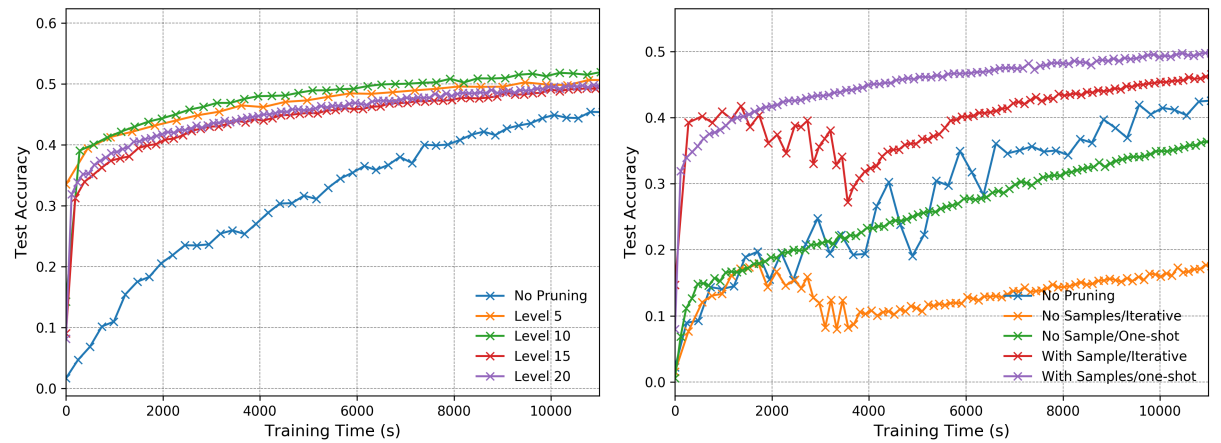
Figure A. CIFAR10 Dataset (i.i.d Partitioned) with VGG11 Architecture



(a) The 4 Possible Pruning Cases

(b) Comparing Sample-based, One-shot Approach at Pruning Level 5, 10, 15, 20

Figure B. FEMNIST Dataset (Non-i.i.d Partitioned) with Conv-FEMNIST Architecture and Sampled Clients



(a) The 4 Possible Pruning Cases

(b) Comparing Sample-based, One-shot Approach at Pruning Level 5, 10, 15, 20

Figure C. FEMNIST Dataset (Non-i.i.d Partitioned) with LeNet-300-100 Architecture