

The Lottery Ticket Hypothesis

Jonathon Frankle and Michael Carbin

(Presented by Logan Engstrom)

Central Claim

Consider an initialized neural network

The Lottery Ticket Hypothesis

Claim: It contains a *subnetwork* that, when trained from initialization:

- Matches or exceeds the test accuracy of the original network
- Requires fewer training iterations than the original network

Note that this subnetwork has significantly fewer parameters than the original!

Finding Subnetworks

Consider a trained neural network $f(x; \theta)$

How do we find a subnetwork?

Answer: threshold weights by absolute value

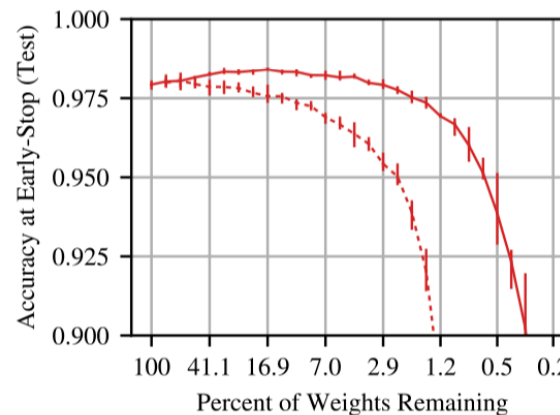
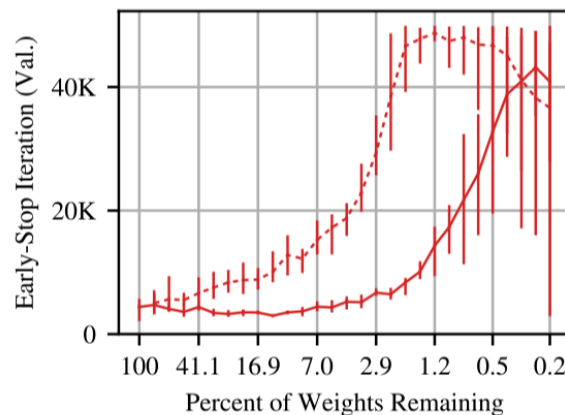
Constructed subnetwork: $f(x; \theta \odot m)$

(m : 1 if in top p % of weights, 0 otherwise)

(Optimization: repeat k times, zeroing $(1-p)^{1/k}$ % each time)

Lottery Tickets Exist

What happens in practice?

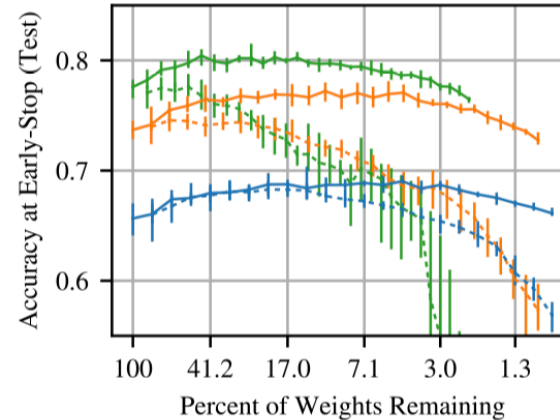
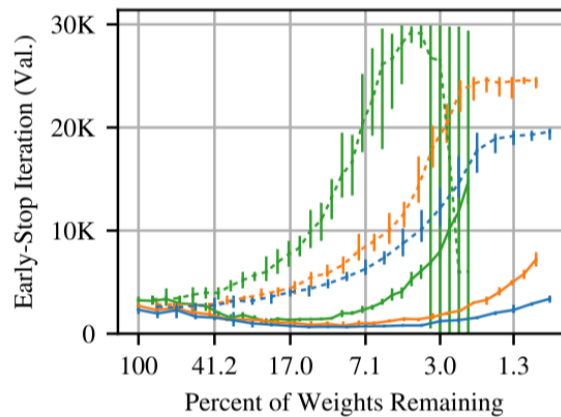


Dataset: MNIST, Model: LeNet (Fully Connected)

Wonders: Faster training, higher accuracy for sparser networks!

Comment: accuracy is far off from SOTA

Lottery Tickets Exist



—+— Conv-6 -.-+.-.- random —+— Conv-4 -.-+.-.- random —+— Conv-2 -.-+.-.- random

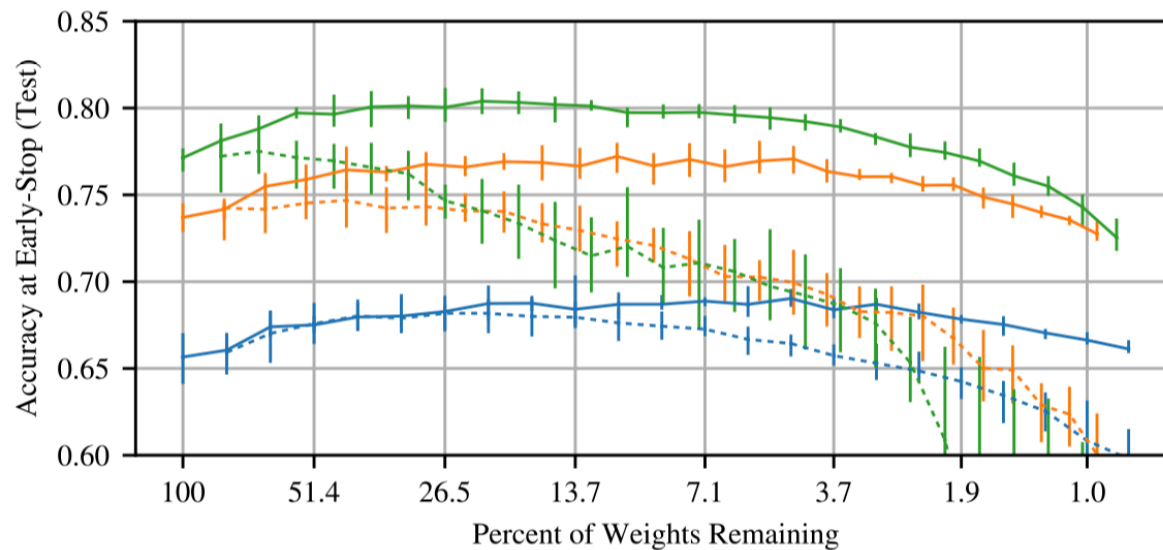
Dataset: CIFAR10, Model: VGG

Wonders: Faster training, higher accuracy for sparser networks!

Comment: accuracy is far off from SOTA

Key Properties of Lottery Tickets

Importance of structure vs initialization?



—+— Conv-2 - - -+ - - Conv-2 reinit —+— Conv-4 - - -+ - - Conv-4 reinit —+— Conv-6 - - -+ - - Conv-6 reinit

Dataset: CIFAR, Model: VGG

Caveats

Large models use global pruning, small models use layerwise

ResNet18, VGG19 require new LR schedule

No models hit SOTA

No ImageNet

Some questions

What are the actual implications of lottery tickets?

How does capacity relate to resulting lottery tickets?

Do lottery tickets transfer across datasets?

Are there losing tickets? Can we use these to find adv inits?

Can we characterize good subnetworks?