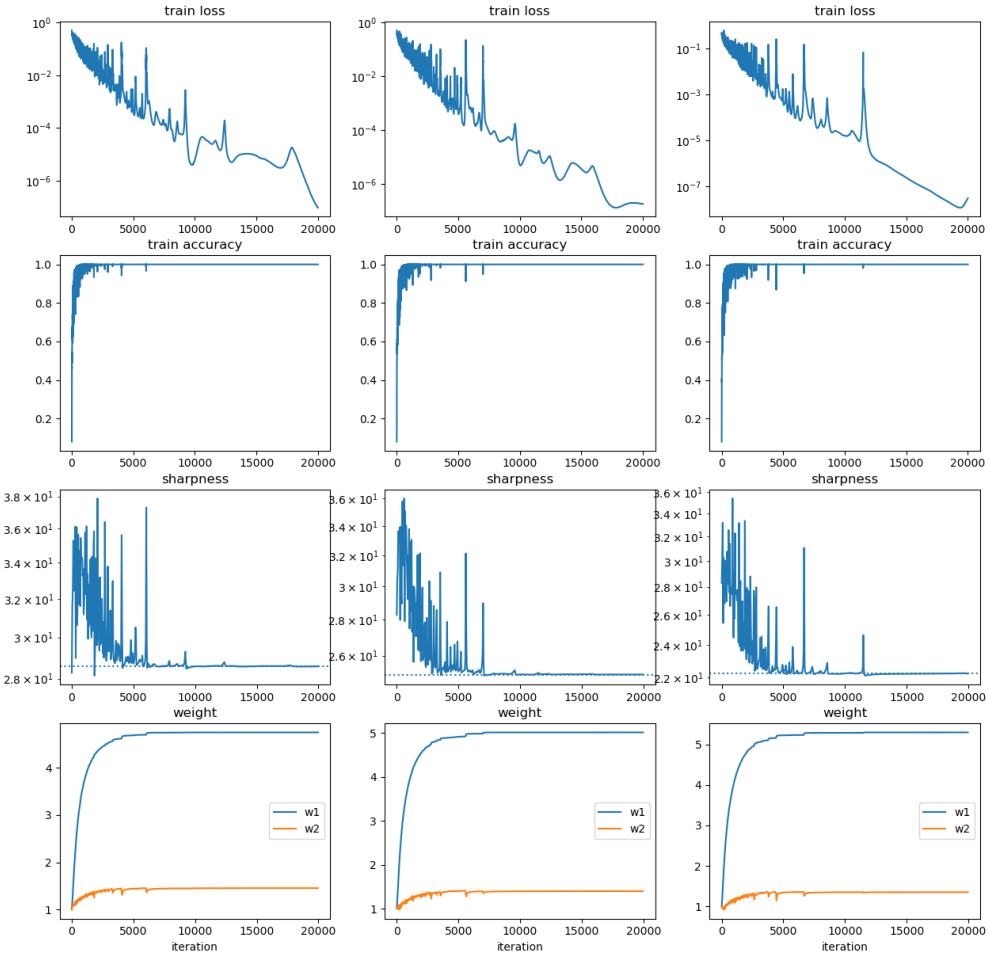
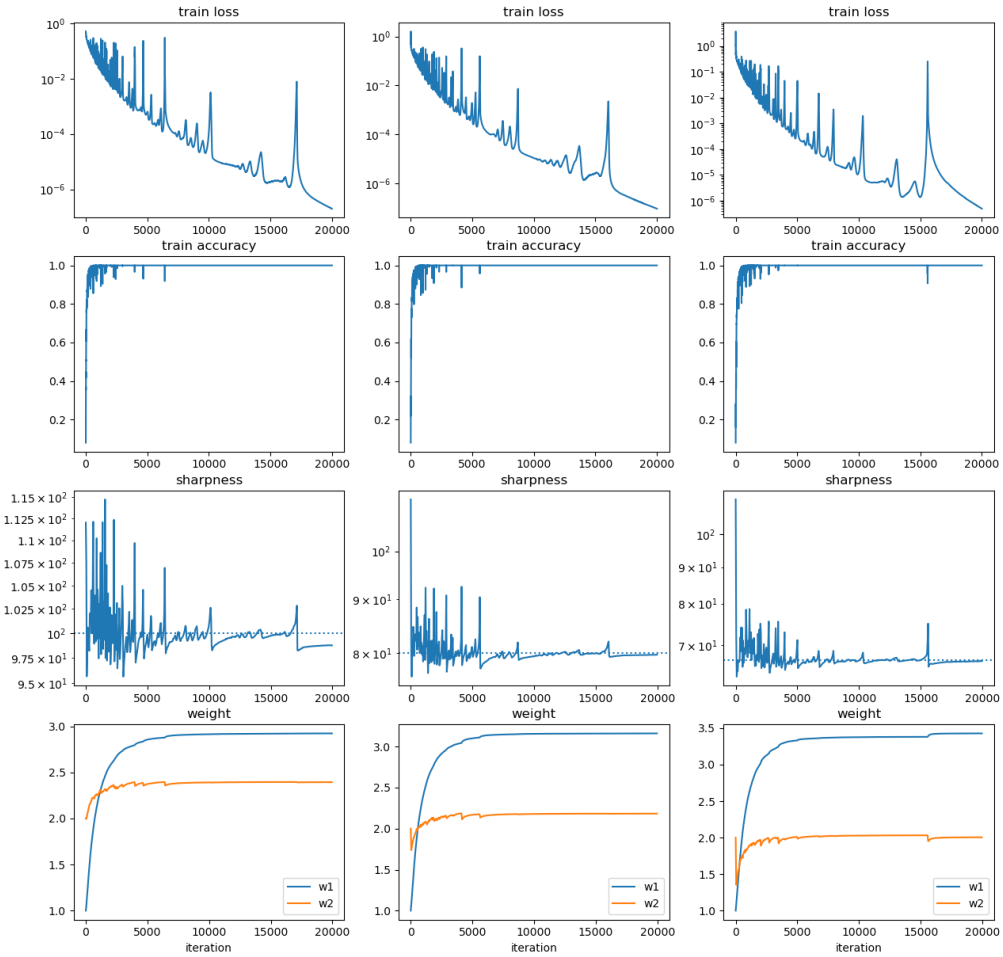
MSE+ReLU

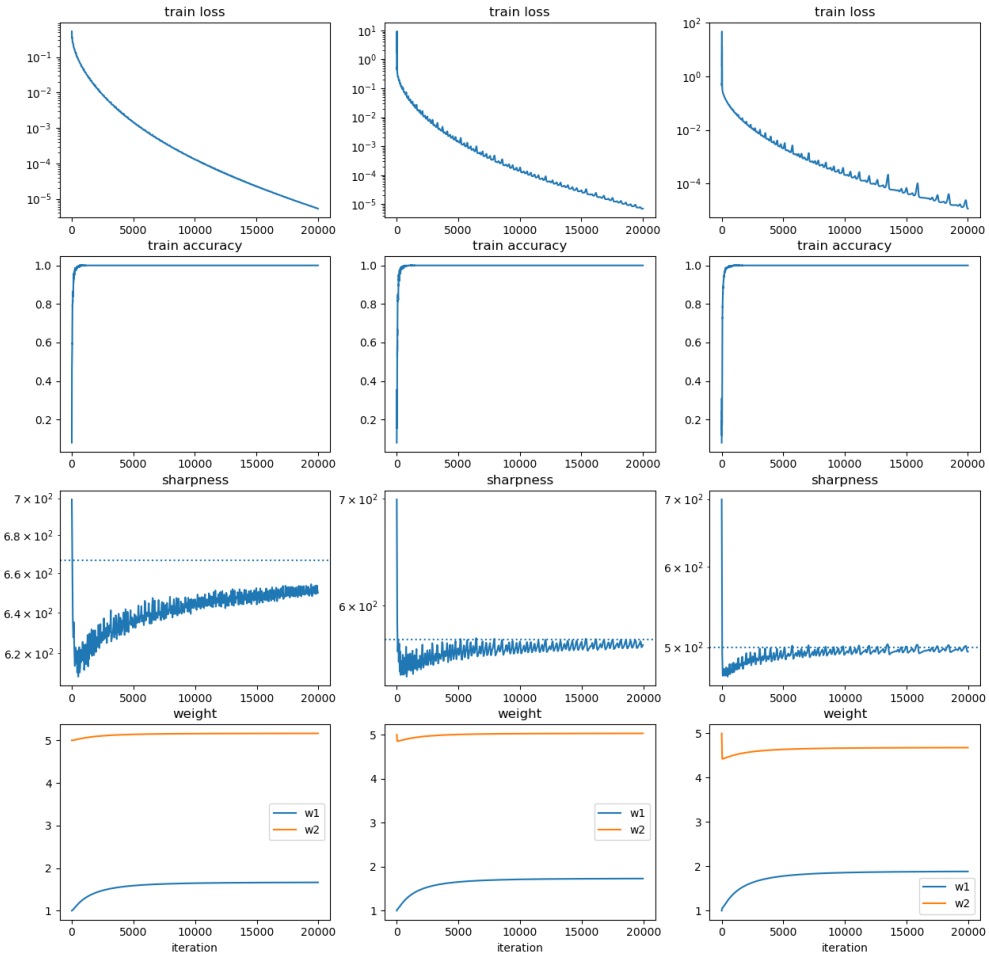
cifar10-1k/fc-relu-depth1/seed_0/mse/gd/w1_1_w2_1_lr_[0.07 0.08 0.09]



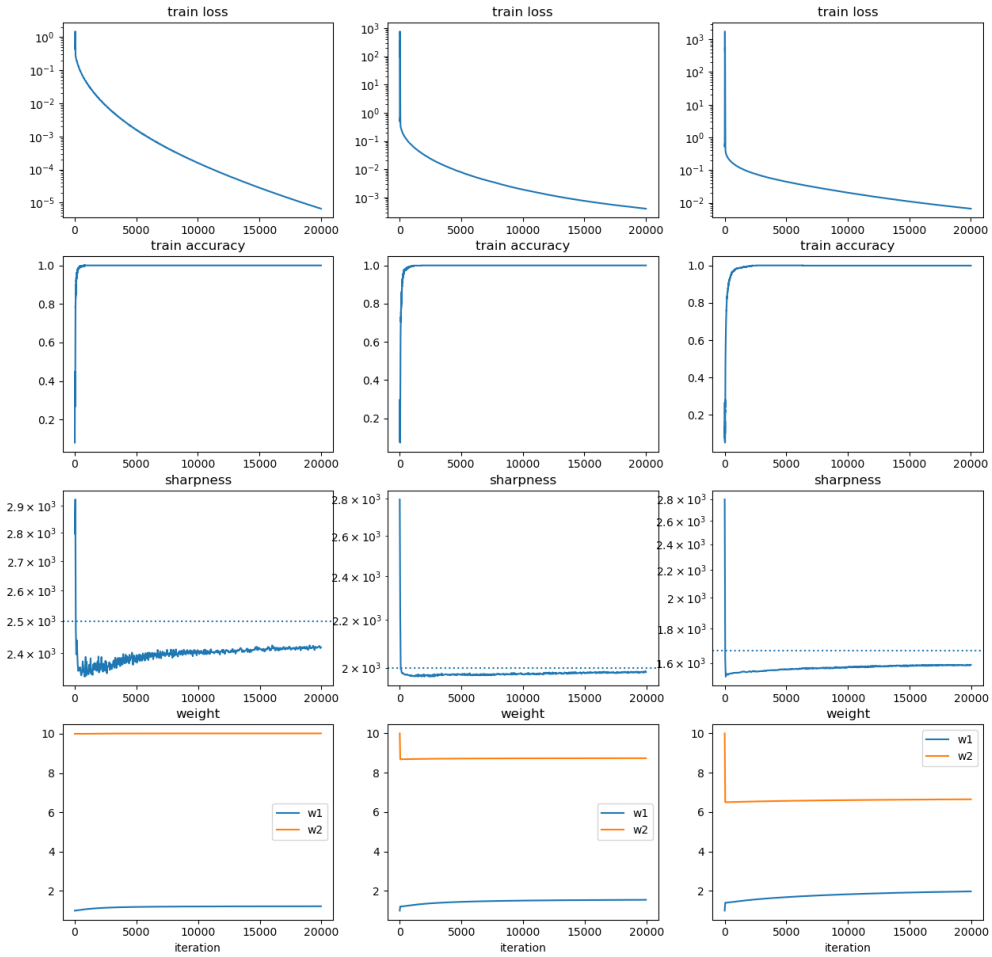
cifar10-1k/fc-relu-depth1/seed_0/mse/gd/w1_1_w2_2_lr_[0.02 0.025 0.03]



cifar10-1k/fc-relu-depth1/seed_0/mse/gd/w1_1_w2_5_lr_[0.003 0.0035 0.004]

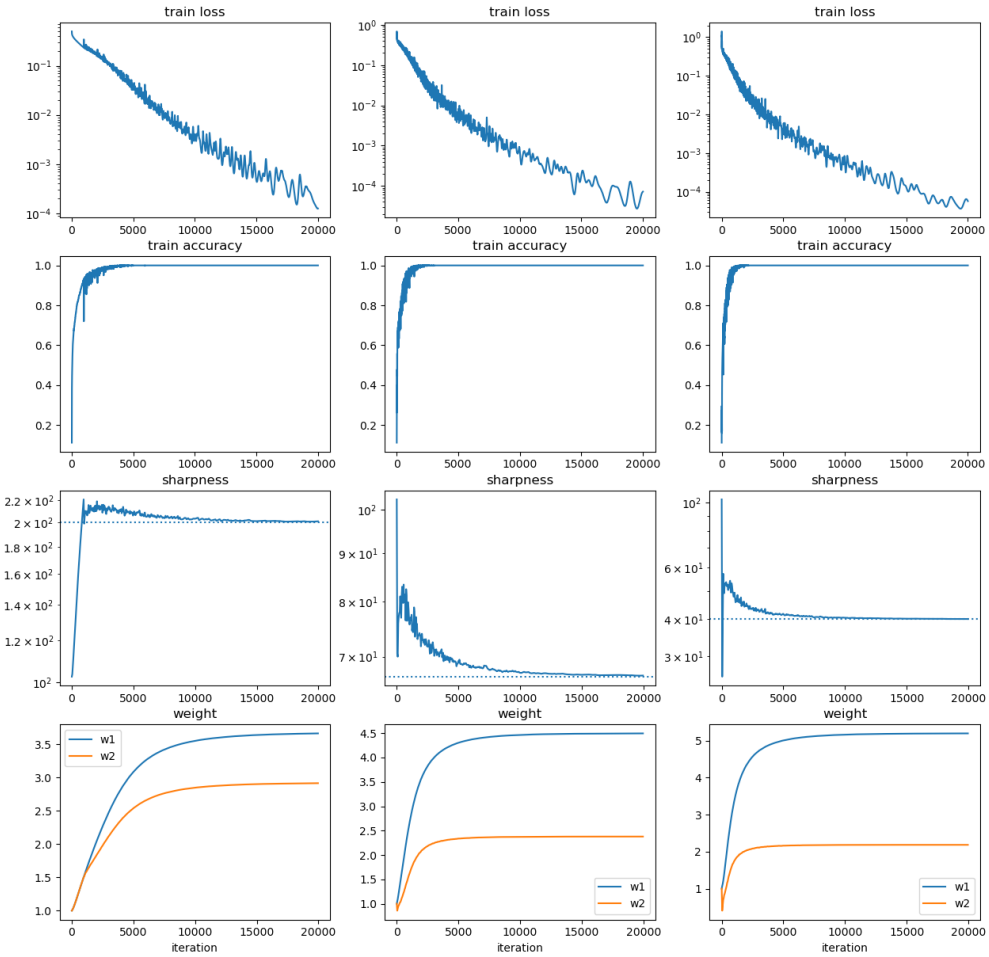


cifar10-1k/fc-relu-depth1/seed_0/mse/gd/w1_1_w2_10_lr_[0.0008 0.001 0.0012]

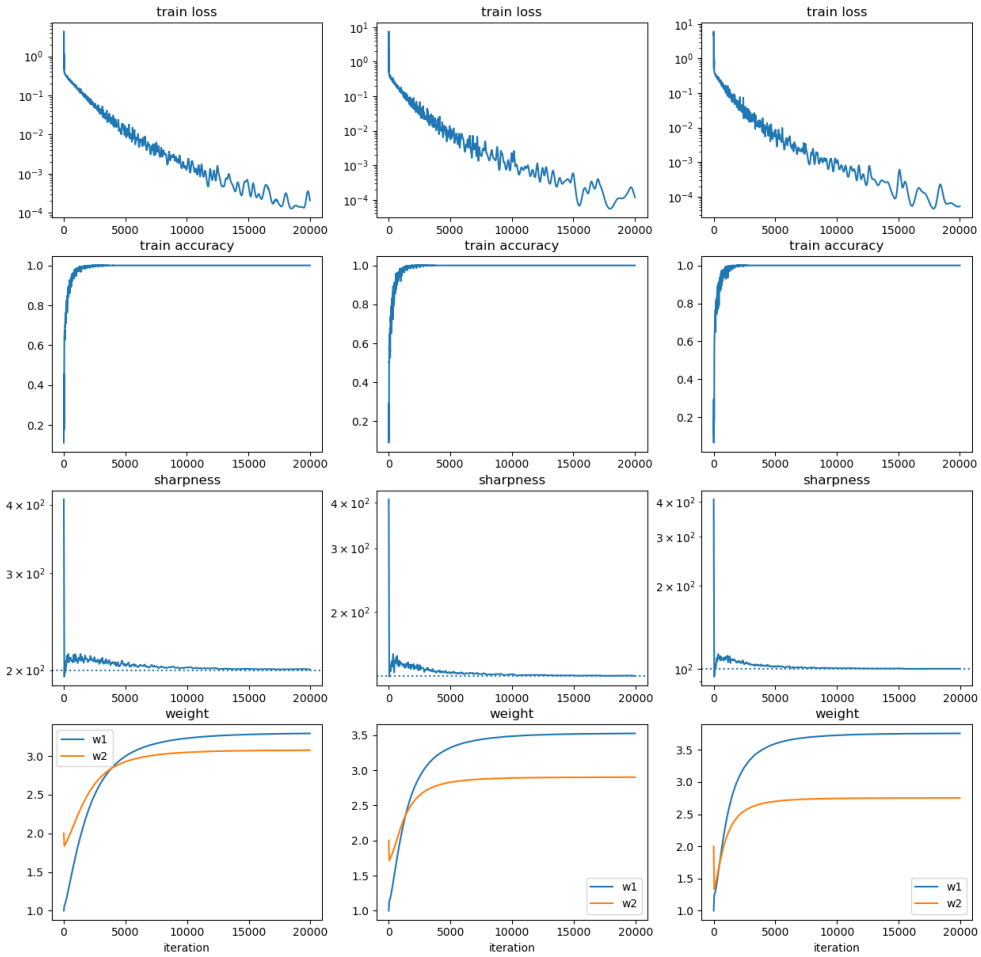


MSE+Tanh

cifar10-1k/fc-tanh-depth1/seed_0/mse/gd/w1_1_w2_1_lr_[0.01 0.03 0.05]

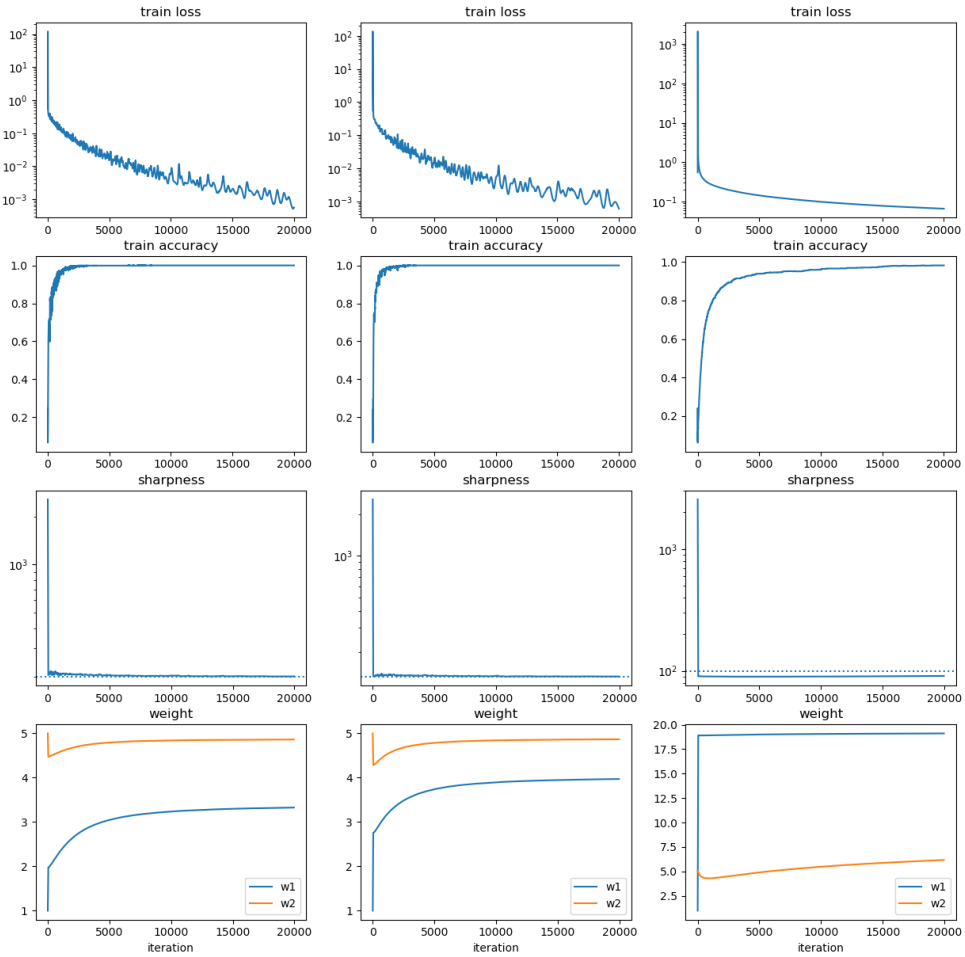


cifar10-1k/fc-tanh-depth1/seed_0/mse/gd/w1_1_w2_2_lr_[0.01 0.015 0.02]



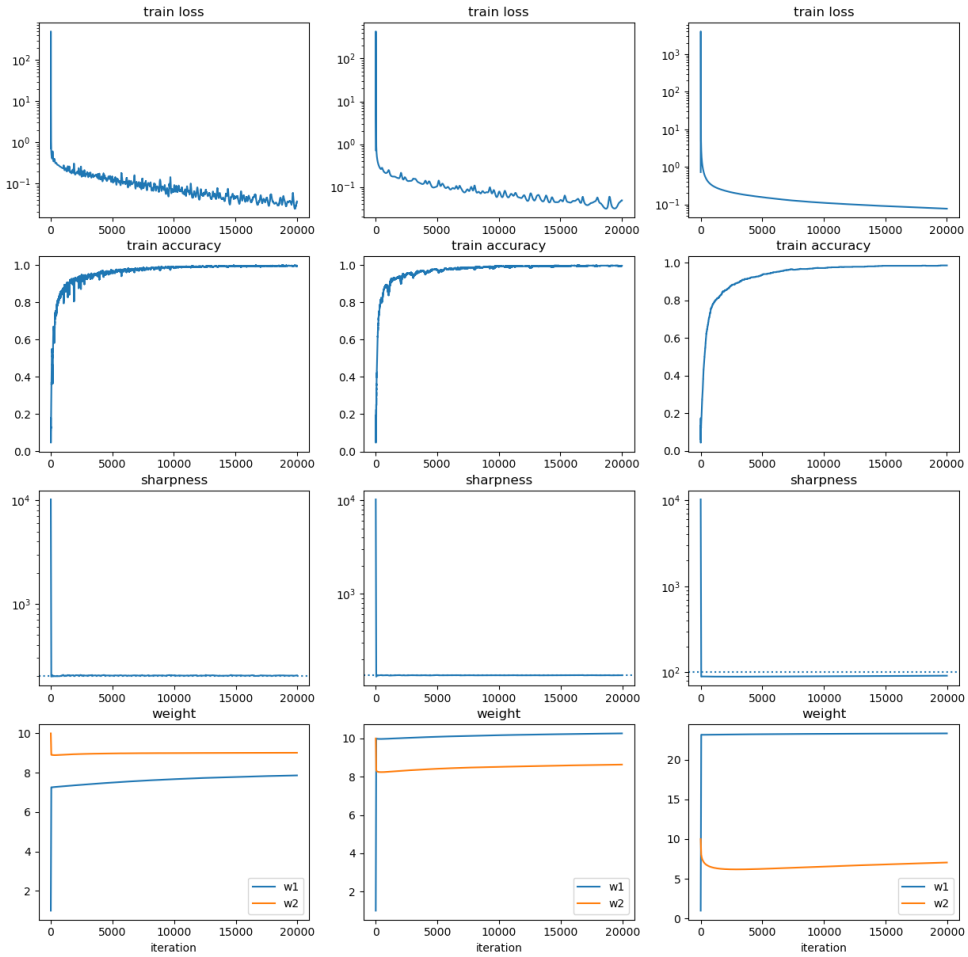
(under train):

cifar10-1k/fc-tanh-depth1/seed_0/mse/gd/w1_1_w2_5_lr_[0.01 0.015 0.02]



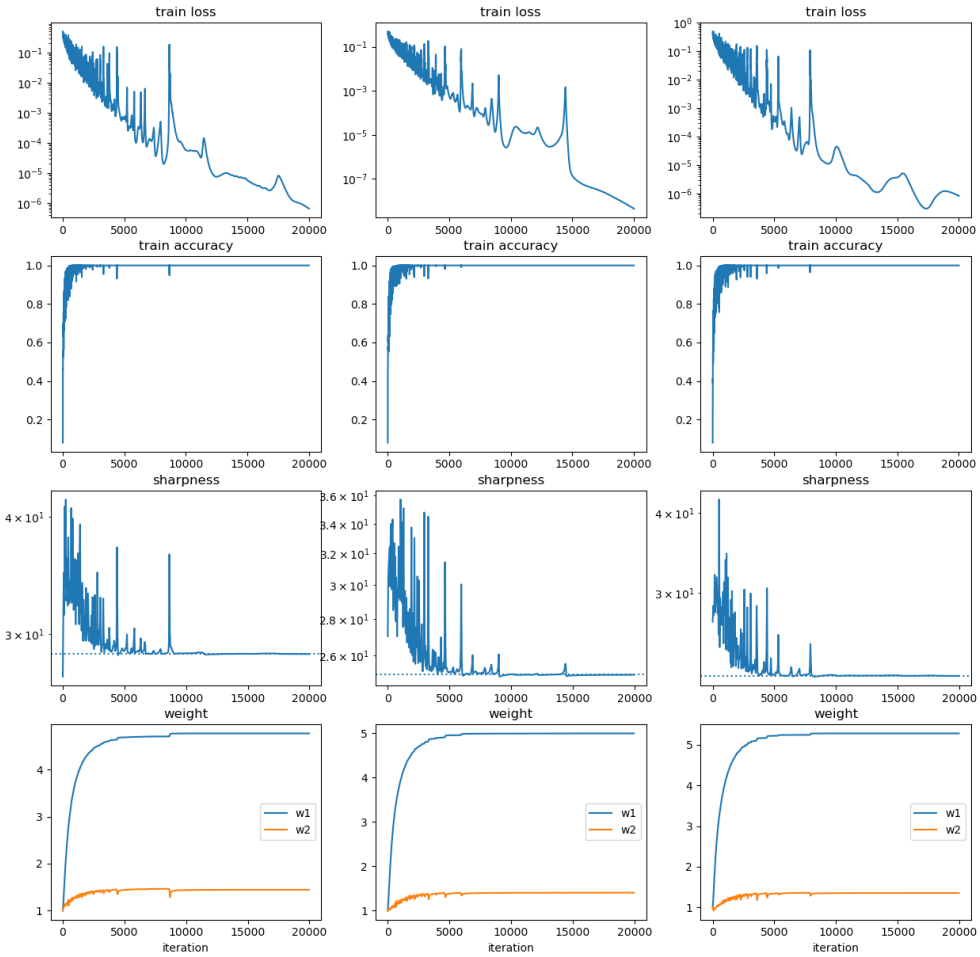
(under train):

cifar10-1k/fc-tanh-depth1/seed_0/mse/gd/w1_1_w2_10_lr_[0.01 0.015 0.02]

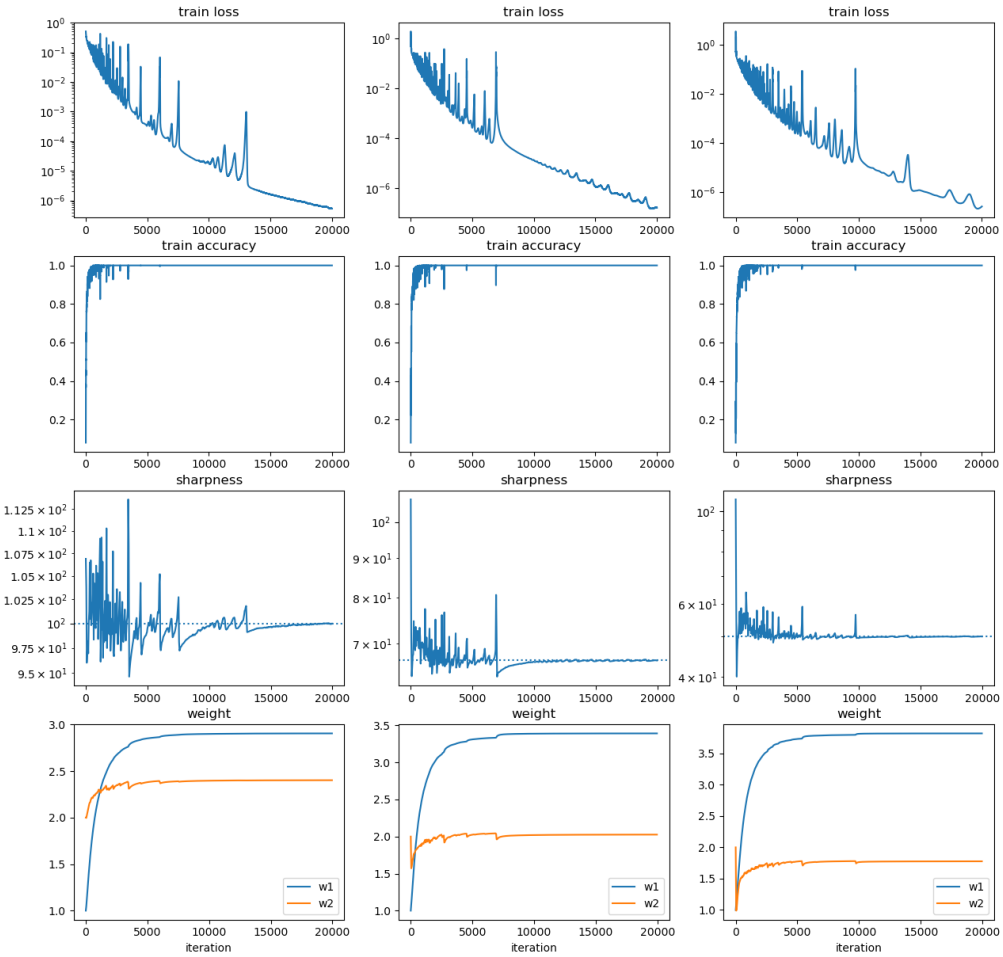


Huber+ReLU

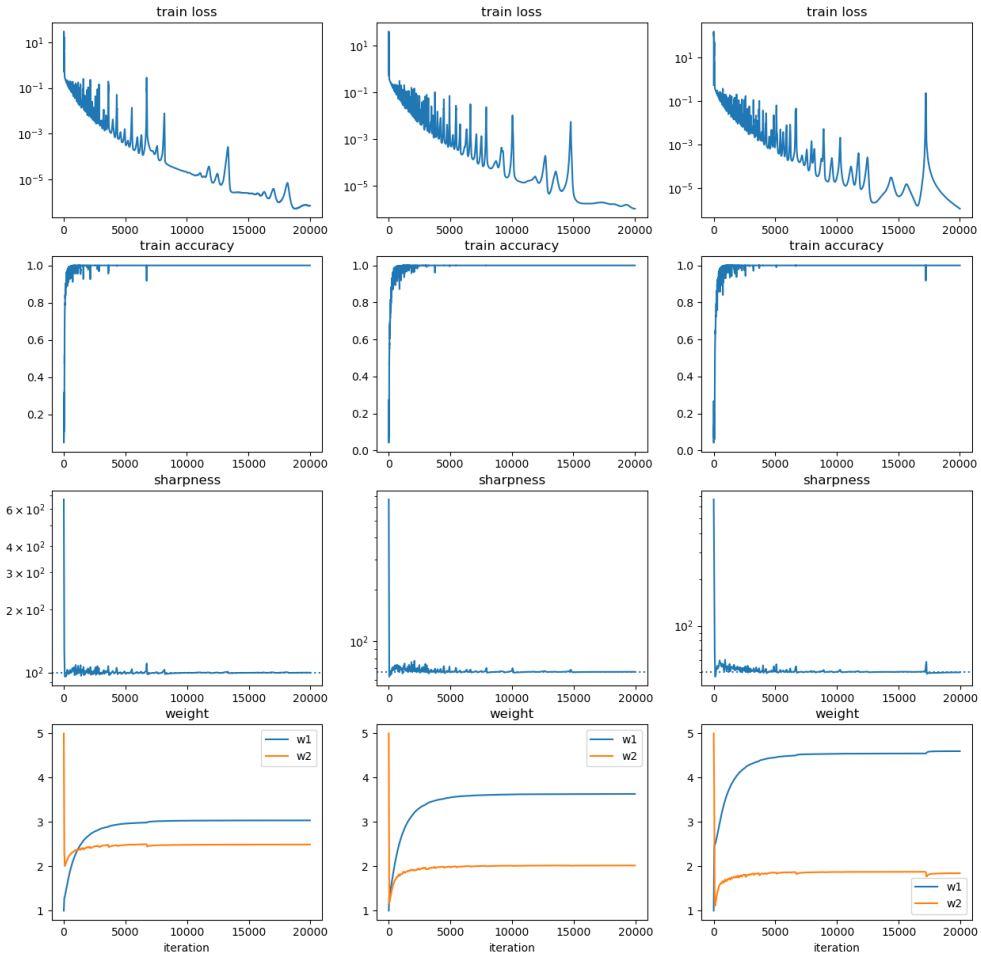
cifar10-1k/fc-relu-depth1/seed_0/huber/gd/w1_1_w2_1_lr_[0.07 0.08 0.09]



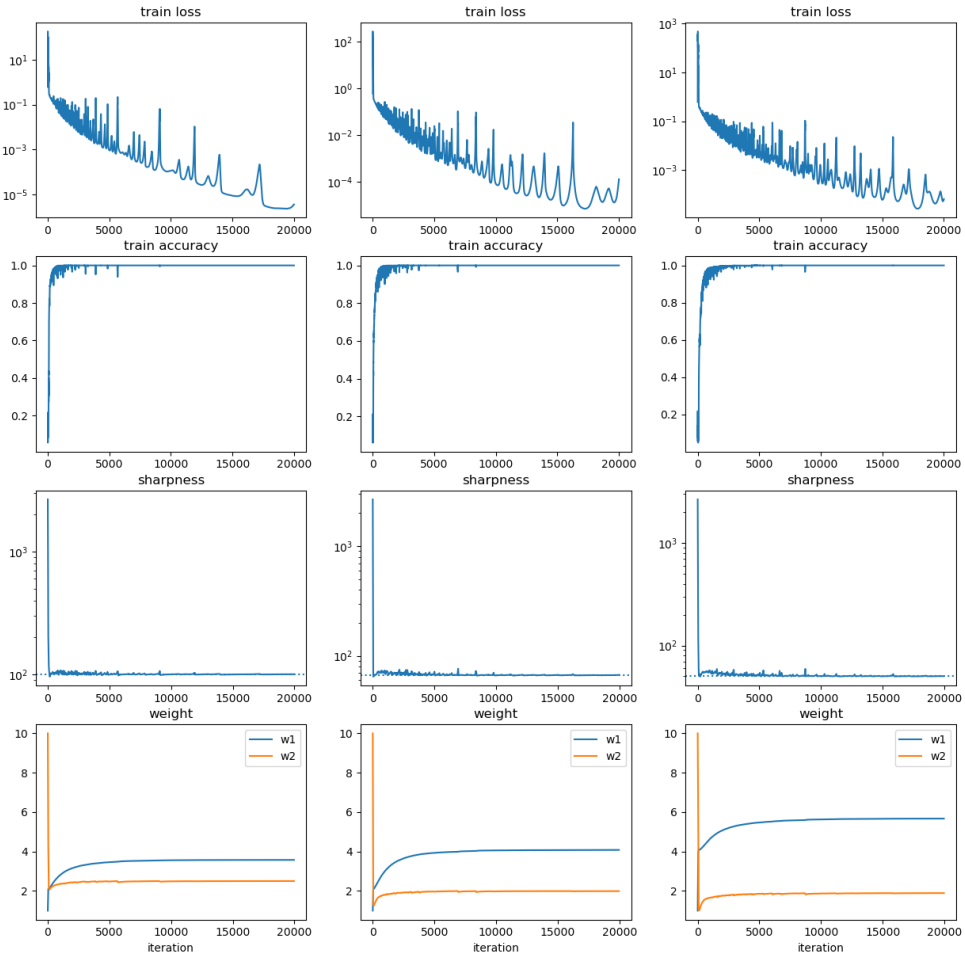
cifar10-1k/fc-relu-depth1/seed_0/huber/gd/w1_1_w2_2_lr_[0.02 0.03 0.04]



cifar10-1k/fc-relu-depth1/seed_0/huber/gd/w1_1_w2_5_lr_[0.02 0.03 0.04]

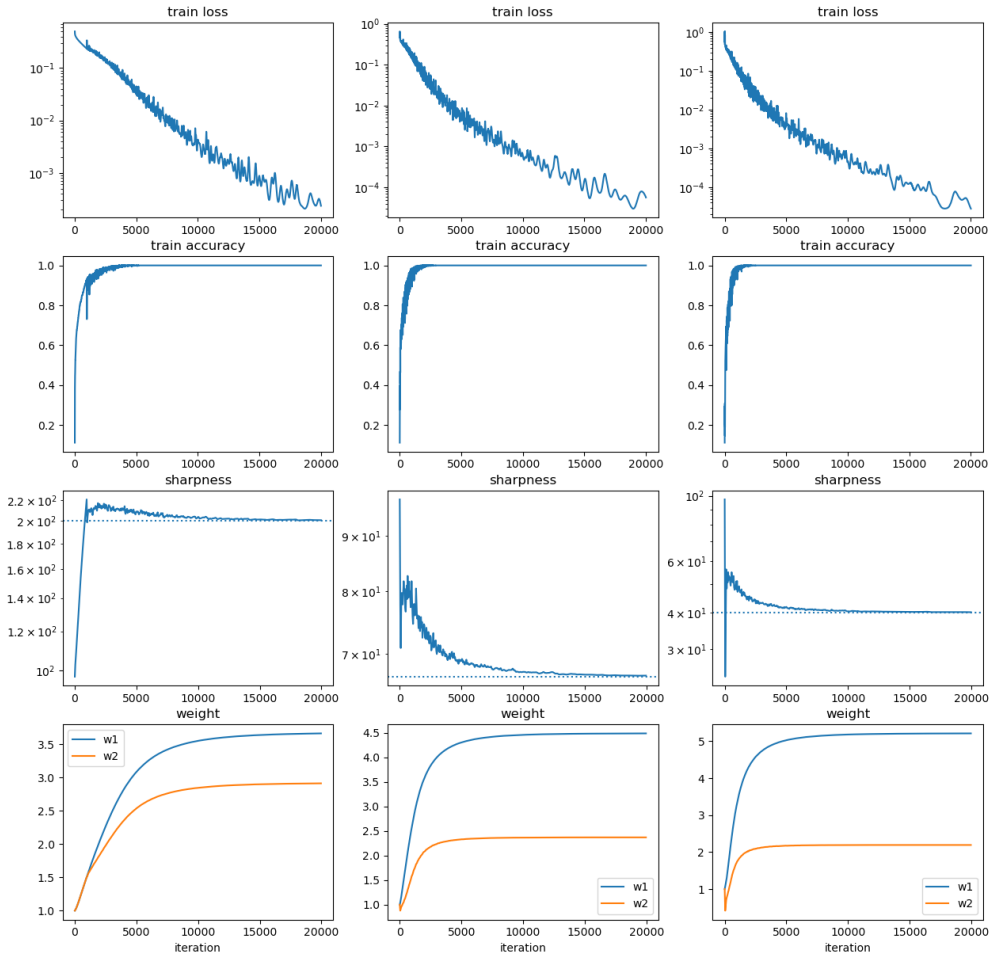


cifar10-1k/fc-relu-depth1/seed_0/huber/gd/w1_1_w2_10_lr_[0.02 0.03 0.04]

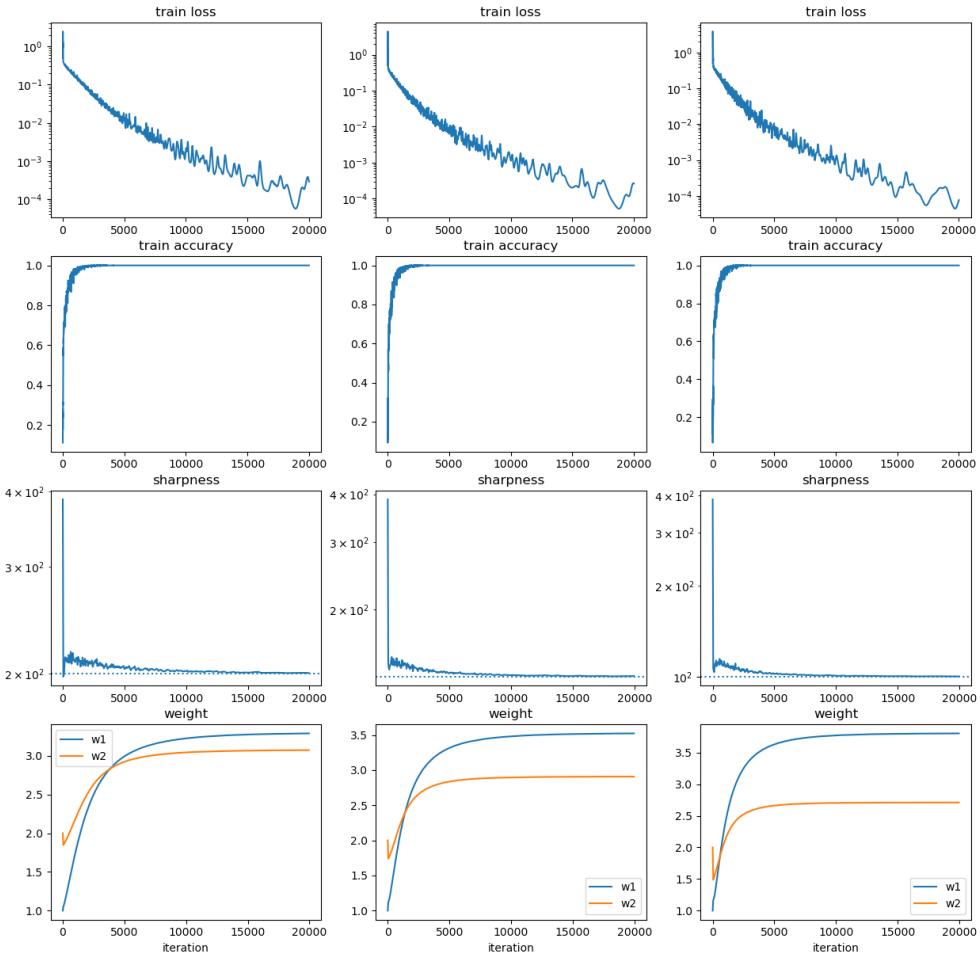


Huber+Tanh

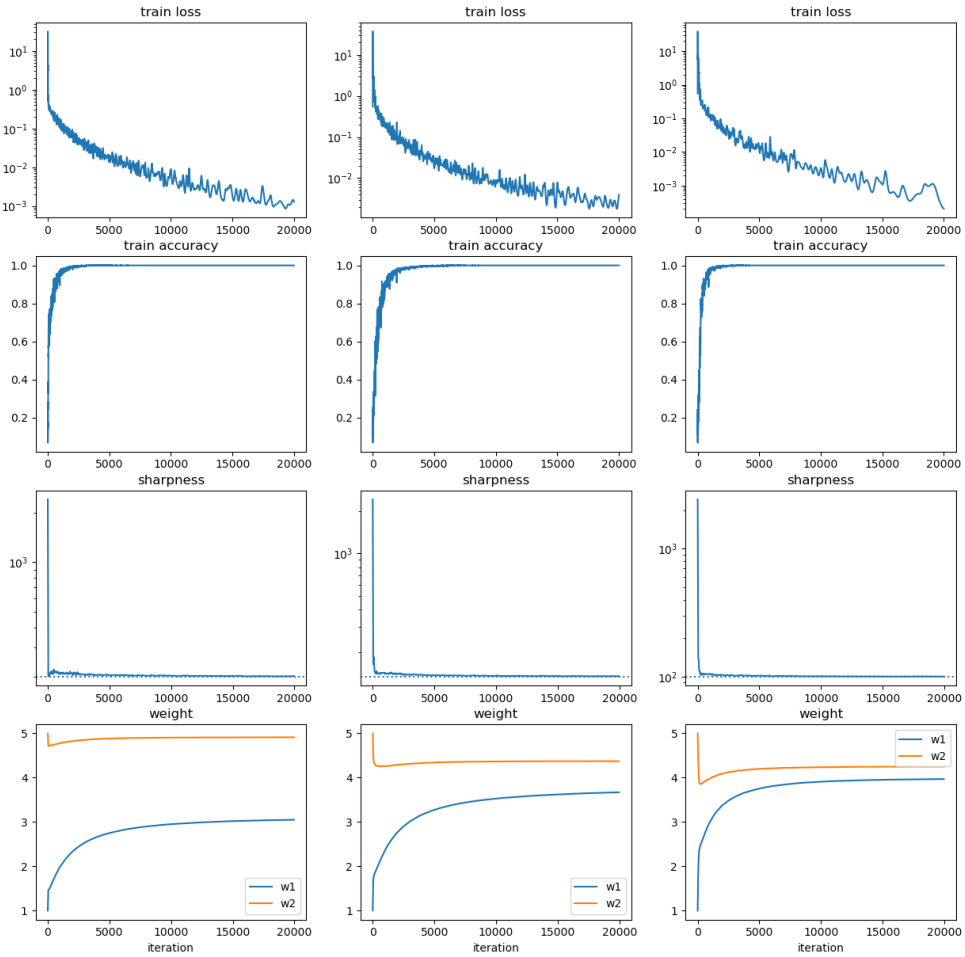
cifar10-1k/fc-tanh-depth1/seed_0/huber/gd/w1_1_w2_1_lr_[0.01 0.03 0.05]

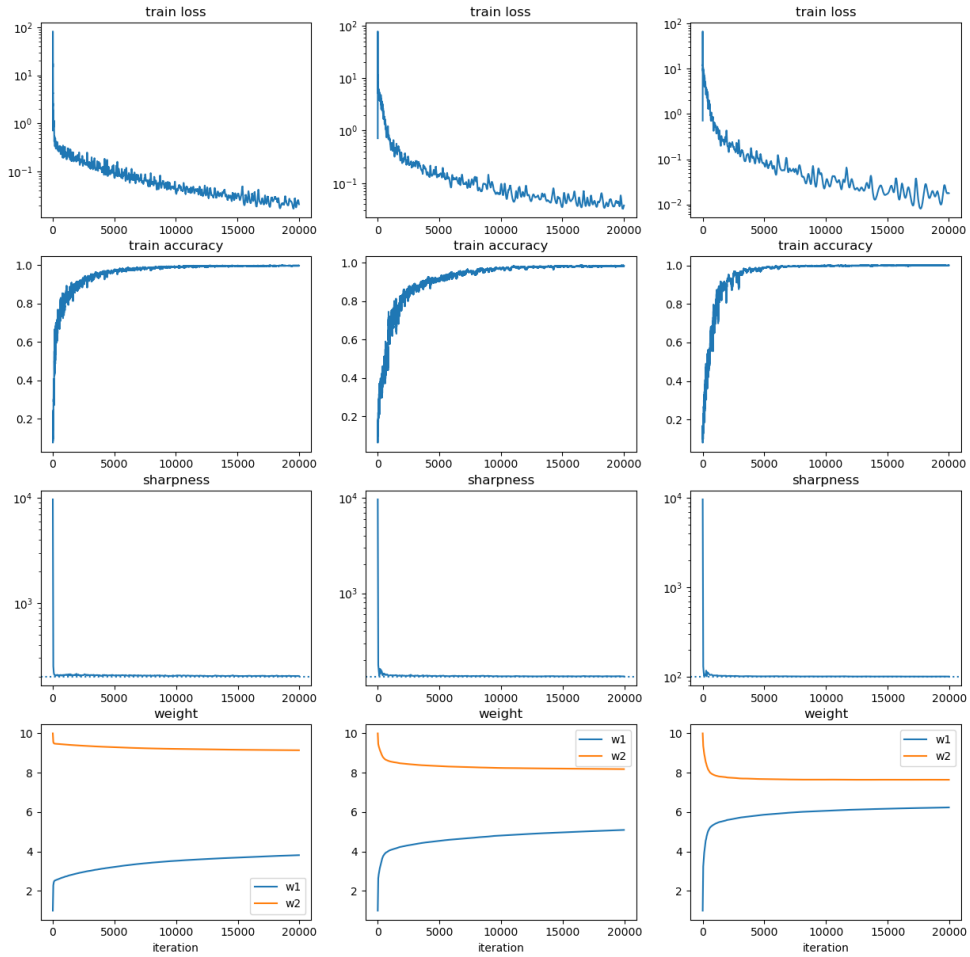


cifar10-1k/fc-tanh-depth1/seed_0/huber/gd/w1_1_w2_2_lr_[0.01 0.015 0.02]



cifar10-1k/fc-tanh-depth1/seed_0/huber/gd/w1_1_w2_5_lr_[0.01 0.015 0.02]

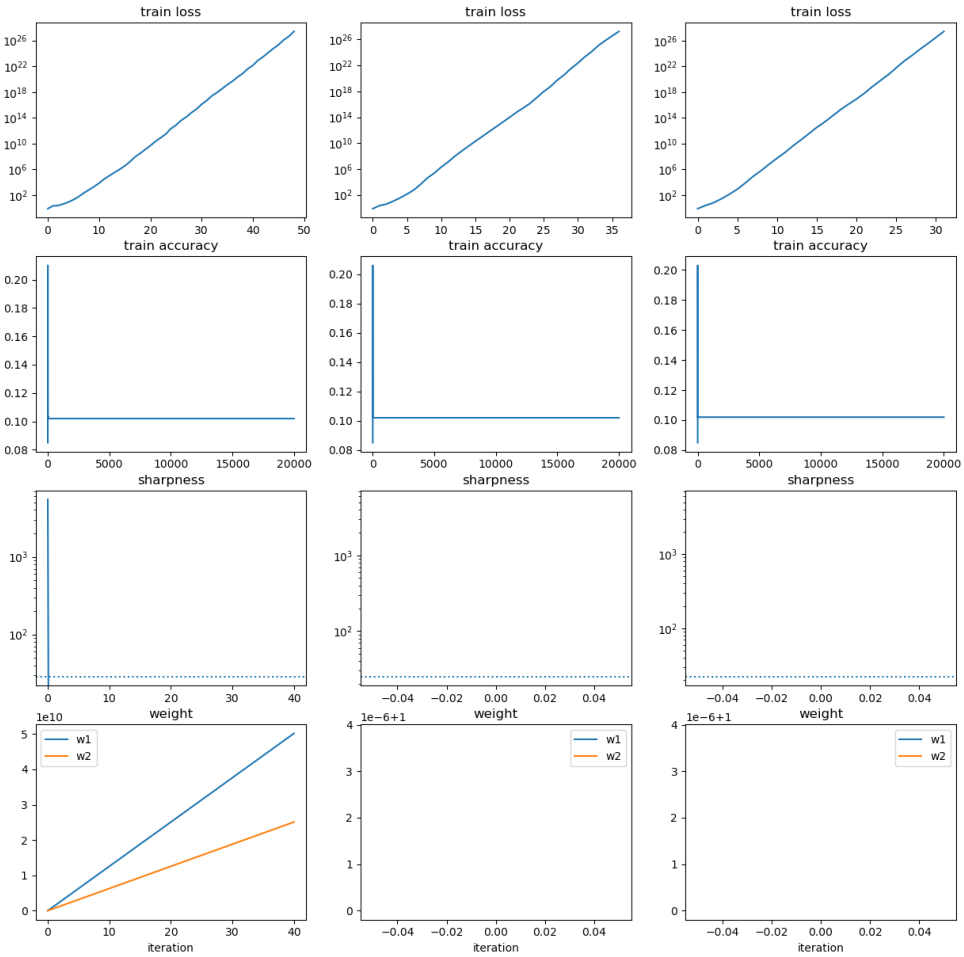




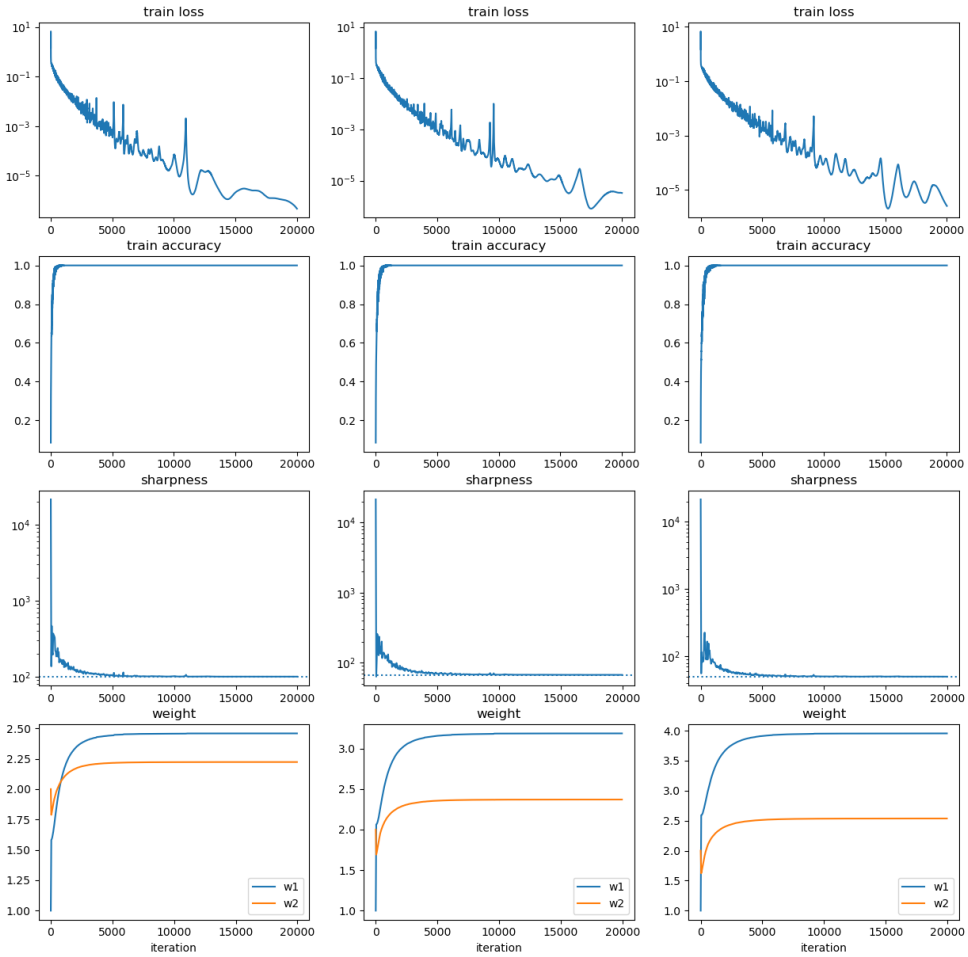
MSE+ReLU+BN

(smaller lr??):

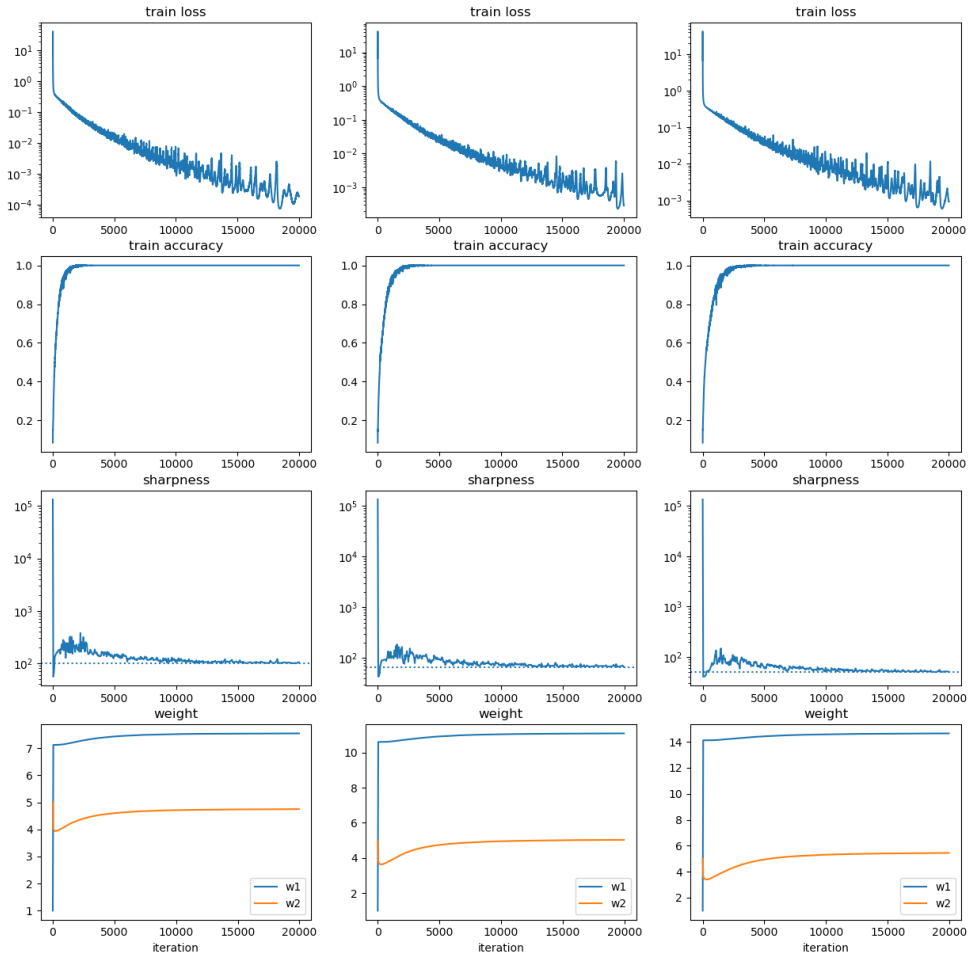
cifar10-1k/fc-relu-depth1_BN/seed_0/mse/gd/w1_1_w2_1_lr_[0.07 0.08 0.09]



cifar10-1k/fc-relu-depth1_BN/seed_0/mse/gd/w1_1_w2_2_lr_[0.02 0.03 0.04]

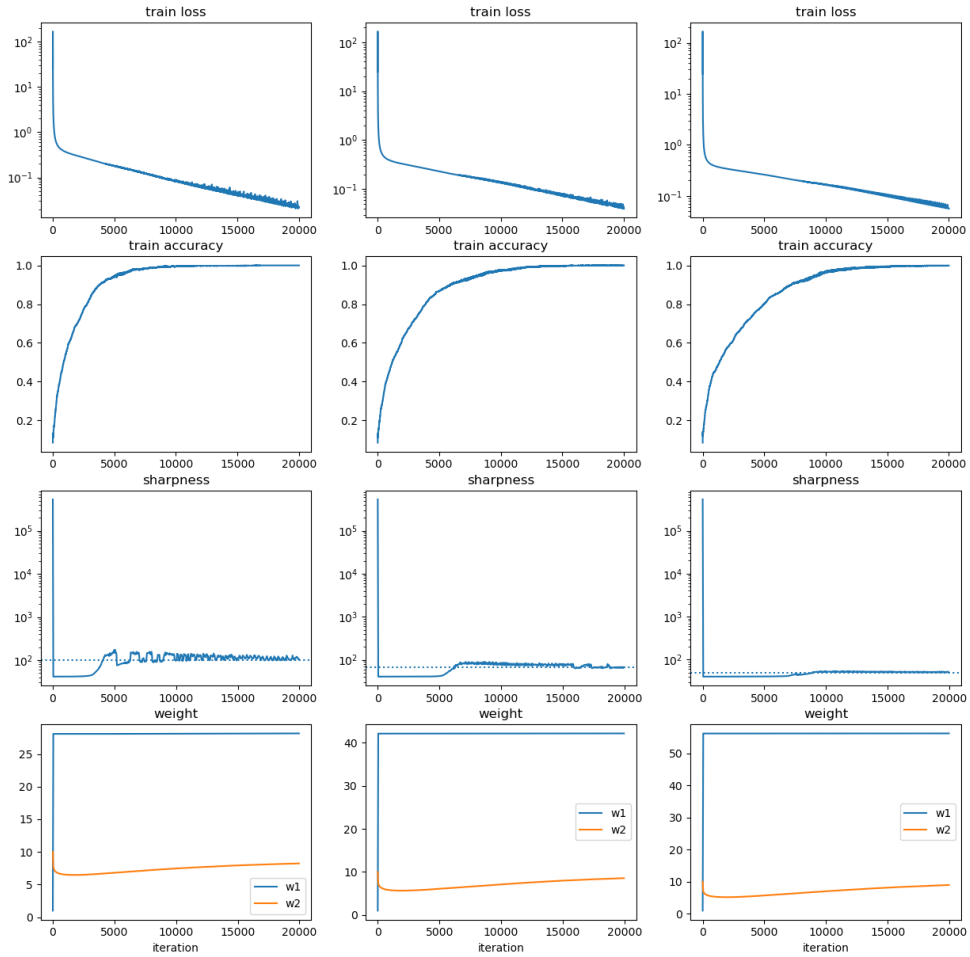


cifar10-1k/fc-relu-depth1_BN/seed_0/mse/gd/w1_1_w2_5_lr_[0.02 0.03 0.04]



(under train):

cifar10-1k/fc-relu-depth1_BN/seed_0/mse/gd/w1_1_w2_10_lr_[0.02 0.03 0.04]



- large lr rerun.
- norm difference square -> plot.
- dimension normalized weight.
- activation: relu^3.