Multilayer MNIST CNN [Time 203.543s, GPUs 4] (Ir = 0.000100, batch size = 5000) Loss Accuracy 1.00 train val 0.98 0.96 - 10^{-1} 0.94 0.92 0.90 -10-2 0.88 train 0.86 val 40 20 20 60 40 60