RNG 42, 1 neuron, 2 layers:

Stuck in minima! No movement

With momentum: works fine

When we let it run without convergence crit, it keeps improving

1 neuron, max acc #4788, Err: 0.16301, Test-Err: 0.17856, Predict: 96.3837, Time: 1s

, if SGD converges by T iterations, the mini-batch training with batch size b may need more than T /b iterations