

Deep Neural Networks

Palash Chauhan

May 11, 2017

1 Assignment-2

1.1 Gradient Descent with Tensorflow

- Implemented with a subset of training data (100K) and varying number of iterations

SubsetSize: 5000 NumSteps: 500 TimeElapsed: 13.93 Test accuracy: 80.7%
SubsetSize: 5000 NumSteps: 1000 TimeElapsed: 27.49 Test accuracy: 81.5%
SubsetSize: 5000 NumSteps: 1500 TimeElapsed: 41.05 Test accuracy: 81.8%
SubsetSize: 5000 NumSteps: 2000 TimeElapsed: 54.59 Test accuracy: 81.8%
SubsetSize: 5000 NumSteps: 2500 TimeElapsed: 68.15 Test accuracy: 82.0%
SubsetSize: 5000 NumSteps: 3000 TimeElapsed: 82.06 Test accuracy: 82.0%
SubsetSize: 5000 NumSteps: 3500 TimeElapsed: 95.65 Test accuracy: 81.9%
SubsetSize: 5000 NumSteps: 4000 TimeElapsed: 109.39 Test accuracy: 82.0%
SubsetSize: 5000 NumSteps: 4500 TimeElapsed: 123.09 Test accuracy: 82.1%

SubsetSize: 10000 NumSteps: 500 TimeElapsed: 26.82 Test accuracy: 82.2%
SubsetSize: 10000 NumSteps: 1000 TimeElapsed: 53.17 Test accuracy: 83.0%
SubsetSize: 10000 NumSteps: 1500 TimeElapsed: 79.68 Test accuracy: 83.5%
SubsetSize: 10000 NumSteps: 2000 TimeElapsed: 105.96 Test accuracy: 83.9%
SubsetSize: 10000 NumSteps: 2500 TimeElapsed: 132.19 Test accuracy: 84.1%
SubsetSize: 10000 NumSteps: 3000 TimeElapsed: 158.45 Test accuracy: 84.3%
SubsetSize: 10000 NumSteps: 3500 TimeElapsed: 184.82 Test accuracy: 84.4%
SubsetSize: 10000 NumSteps: 4000 TimeElapsed: 211.10 Test accuracy: 84.4%
SubsetSize: 10000 NumSteps: 4500 TimeElapsed: 237.34 Test accuracy: 84.5%

SubsetSize: 15000 NumSteps: 500 TimeElapsed: 40.75 Test accuracy: 82.4%
SubsetSize: 15000 NumSteps: 1000 TimeElapsed: 80.88 Test accuracy: 83.3%
SubsetSize: 15000 NumSteps: 1500 TimeElapsed: 123.19 Test accuracy: 83.9%
SubsetSize: 15000 NumSteps: 2000 TimeElapsed: 163.38 Test accuracy: 84.1%
SubsetSize: 15000 NumSteps: 2500 TimeElapsed: 203.56 Test accuracy: 84.4%
SubsetSize: 15000 NumSteps: 3000 TimeElapsed: 243.67 Test accuracy: 84.7%
SubsetSize: 15000 NumSteps: 3500 TimeElapsed: 284.50 Test accuracy: 84.9%
SubsetSize: 15000 NumSteps: 4000 TimeElapsed: 324.80 Test accuracy: 85.2%
SubsetSize: 15000 NumSteps: 4500 TimeElapsed: 365.15 Test accuracy: 85.3%

SubsetSize: 20000 NumSteps: 500 TimeElapsed: 54.14 Test accuracy: 82.3%
SubsetSize: 20000 NumSteps: 1000 TimeElapsed: 107.58 Test accuracy: 83.2%
SubsetSize: 20000 NumSteps: 1500 TimeElapsed: 167.59 Test accuracy: 83.9%
SubsetSize: 20000 NumSteps: 2000 TimeElapsed: 221.55 Test accuracy: 84.2%
SubsetSize: 20000 NumSteps: 2500 TimeElapsed: 274.99 Test accuracy: 84.7%
SubsetSize: 20000 NumSteps: 3000 TimeElapsed: 328.04 Test accuracy: 84.9%
SubsetSize: 20000 NumSteps: 3500 TimeElapsed: 381.22 Test accuracy: 85.1%
SubsetSize: 20000 NumSteps: 4000 TimeElapsed: 434.55 Test accuracy: 85.3%
SubsetSize: 20000 NumSteps: 4500 TimeElapsed: 488.82 Test accuracy: 85.4%

SubsetSize: 25000 NumSteps: 500 TimeElapsed: 67.05 Test accuracy: 82.5%
SubsetSize: 25000 NumSteps: 1000 TimeElapsed: 133.22 Test accuracy: 83.5%
SubsetSize: 25000 NumSteps: 1500 TimeElapsed: 199.75 Test accuracy: 84.0%

SubsetSize: 25000 NumSteps: 2000 TimeElapsed: 266.00 Test accuracy: 84.3%
 SubsetSize: 25000 NumSteps: 2500 TimeElapsed: 332.50 Test accuracy: 84.6%
 SubsetSize: 25000 NumSteps: 3000 TimeElapsed: 401.35 Test accuracy: 85.0%
 SubsetSize: 25000 NumSteps: 3500 TimeElapsed: 471.72 Test accuracy: 85.3%
 SubsetSize: 25000 NumSteps: 4000 TimeElapsed: 548.04 Test accuracy: 85.5%
 SubsetSize: 25000 NumSteps: 4500 TimeElapsed: 621.31 Test accuracy: 85.7%

SubsetSize: 30000 NumSteps: 500 TimeElapsed: 81.26 Test accuracy: 82.7%
 SubsetSize: 30000 NumSteps: 1000 TimeElapsed: 168.11 Test accuracy: 83.7%
 SubsetSize: 30000 NumSteps: 1500 TimeElapsed: 248.38 Test accuracy: 84.5%
 SubsetSize: 30000 NumSteps: 2000 TimeElapsed: 328.91 Test accuracy: 85.1%
 SubsetSize: 30000 NumSteps: 2500 TimeElapsed: 416.95 Test accuracy: 85.4%
 SubsetSize: 30000 NumSteps: 3000 TimeElapsed: 500.80 Test accuracy: 85.7%
 SubsetSize: 30000 NumSteps: 3500 TimeElapsed: 585.57 Test accuracy: 85.9%
 SubsetSize: 30000 NumSteps: 4000 TimeElapsed: 666.61 Test accuracy: 86.2%
 SubsetSize: 30000 NumSteps: 4500 TimeElapsed: 747.29 Test accuracy: 86.4%

- On an average it takes 450s to get a test accuracy of 84.9%

1.2 Stochastic Gradient Descent with TensorFlow

- Implemented with varying batch size

BatchSize: 64 NumSteps: 500 TimeElapsed: 0.81 Test accuracy: 82.2%
 BatchSize: 64 NumSteps: 1000 TimeElapsed: 1.33 Test accuracy: 83.6%
 BatchSize: 64 NumSteps: 1500 TimeElapsed: 1.84 Test accuracy: 83.8%
 BatchSize: 64 NumSteps: 2000 TimeElapsed: 2.37 Test accuracy: 83.5%
 BatchSize: 64 NumSteps: 2500 TimeElapsed: 2.89 Test accuracy: 84.8%
 BatchSize: 64 NumSteps: 3000 TimeElapsed: 3.41 Test accuracy: 83.7%
 BatchSize: 64 NumSteps: 3500 TimeElapsed: 3.92 Test accuracy: 85.0%
 BatchSize: 64 NumSteps: 4000 TimeElapsed: 4.45 Test accuracy: 84.8%
 BatchSize: 64 NumSteps: 4500 TimeElapsed: 4.96 Test accuracy: 86.1%

BatchSize: 128 NumSteps: 500 TimeElapsed: 1.01 Test accuracy: 82.9%
 BatchSize: 128 NumSteps: 1000 TimeElapsed: 1.80 Test accuracy: 83.5%
 BatchSize: 128 NumSteps: 1500 TimeElapsed: 2.59 Test accuracy: 84.4%
 BatchSize: 128 NumSteps: 2000 TimeElapsed: 3.38 Test accuracy: 84.0%
 BatchSize: 128 NumSteps: 2500 TimeElapsed: 4.17 Test accuracy: 85.6%
 BatchSize: 128 NumSteps: 3000 TimeElapsed: 4.97 Test accuracy: 85.0%
 BatchSize: 128 NumSteps: 3500 TimeElapsed: 5.75 Test accuracy: 86.2%
 BatchSize: 128 NumSteps: 4000 TimeElapsed: 6.55 Test accuracy: 86.3%
 BatchSize: 128 NumSteps: 4500 TimeElapsed: 7.35 Test accuracy: 86.6%

BatchSize: 256 NumSteps: 500 TimeElapsed: 1.53 Test accuracy: 82.7%
 BatchSize: 256 NumSteps: 1000 TimeElapsed: 2.82 Test accuracy: 83.8%
 BatchSize: 256 NumSteps: 1500 TimeElapsed: 4.11 Test accuracy: 84.5%
 BatchSize: 256 NumSteps: 2000 TimeElapsed: 5.40 Test accuracy: 85.2%
 BatchSize: 256 NumSteps: 2500 TimeElapsed: 6.69 Test accuracy: 85.7%
 BatchSize: 256 NumSteps: 3000 TimeElapsed: 7.98 Test accuracy: 86.1%
 BatchSize: 256 NumSteps: 3500 TimeElapsed: 9.27 Test accuracy: 86.3%
 BatchSize: 256 NumSteps: 4000 TimeElapsed: 10.56 Test accuracy: 86.2%
 BatchSize: 256 NumSteps: 4500 TimeElapsed: 11.84 Test accuracy: 86.5%

BatchSize: 512 NumSteps: 500 TimeElapsed: 2.51 Test accuracy: 82.5%
 BatchSize: 512 NumSteps: 1000 TimeElapsed: 4.74 Test accuracy: 83.7%
 BatchSize: 512 NumSteps: 1500 TimeElapsed: 6.98 Test accuracy: 84.4%
 BatchSize: 512 NumSteps: 2000 TimeElapsed: 9.22 Test accuracy: 84.9%
 BatchSize: 512 NumSteps: 2500 TimeElapsed: 11.45 Test accuracy: 85.3%

BatchSize: 512 NumSteps: 3000 TimeElapsed: 13.70 Test accuracy: 86.0%
BatchSize: 512 NumSteps: 3500 TimeElapsed: 15.93 Test accuracy: 86.2%
BatchSize: 512 NumSteps: 4000 TimeElapsed: 18.17 Test accuracy: 86.4%
BatchSize: 512 NumSteps: 4500 TimeElapsed: 20.40 Test accuracy: 86.4%

BatchSize: 1024 NumSteps: 500 TimeElapsed: 4.30 Test accuracy: 82.5%
BatchSize: 1024 NumSteps: 1000 TimeElapsed: 8.35 Test accuracy: 83.8%
BatchSize: 1024 NumSteps: 1500 TimeElapsed: 12.39 Test accuracy: 84.4%
BatchSize: 1024 NumSteps: 2000 TimeElapsed: 16.44 Test accuracy: 85.2%
BatchSize: 1024 NumSteps: 2500 TimeElapsed: 20.48 Test accuracy: 85.6%
BatchSize: 1024 NumSteps: 3000 TimeElapsed: 25.14 Test accuracy: 86.2%
BatchSize: 1024 NumSteps: 3500 TimeElapsed: 29.67 Test accuracy: 86.4%
BatchSize: 1024 NumSteps: 4000 TimeElapsed: 34.03 Test accuracy: 86.7%
BatchSize: 1024 NumSteps: 4500 TimeElapsed: 38.60 Test accuracy: 86.9%

BatchSize: 2048 NumSteps: 500 TimeElapsed: 8.14 Test accuracy: 82.9%
BatchSize: 2048 NumSteps: 1000 TimeElapsed: 16.21 Test accuracy: 83.9%
BatchSize: 2048 NumSteps: 1500 TimeElapsed: 24.39 Test accuracy: 84.7%
BatchSize: 2048 NumSteps: 2000 TimeElapsed: 32.00 Test accuracy: 85.3%
BatchSize: 2048 NumSteps: 2500 TimeElapsed: 39.43 Test accuracy: 85.8%
BatchSize: 2048 NumSteps: 3000 TimeElapsed: 46.82 Test accuracy: 86.1%
BatchSize: 2048 NumSteps: 3500 TimeElapsed: 54.28 Test accuracy: 86.5%
BatchSize: 2048 NumSteps: 4000 TimeElapsed: 61.88 Test accuracy: 86.8%
BatchSize: 2048 NumSteps: 4500 TimeElapsed: 69.37 Test accuracy: 87.0%

- **Optimizing using SGD can achieve 86%+ accuracy using a batch size of 128 in just 7.35s**
- SGD is way faster than batch GD.

1.3 Neural Network

- 1 hidden layer, 1024 nodes, 128 batch size
- Achieves 88.3% test accuracy in around 50s

	Without Regularization	L2-loss	dropout	Learning Rate Decay	10K steps
LR with SGD	83.2%	85.8%	–	87.7%	89.1%
1 layer NN with SGD	88.1%	89%	89.7%	90%	90.7%
Overfitting NN (50 batches)	83.8%	84.7%	85%	84.7%	–

Table 1: Effect of regularization

2 Regularization

- **Early Termination:** Stop when performance does not improve on validation set
- **L2 loss:** Add l_2 norm of the weights to penalize large weights along with a constant(hyperparamter)
- **Dropout:** Randomly squash some of the activations. Use a bigger network if dropout is not working. During evaluation we don't need this randomness, we need the average of the activations. While training, squash the dropouts but also scale the rest appropriately. While evaluationg, skip this dropout and scaling.

2.1 Assignment 3

128 batch size. 2000 steps. Refer to Table 1. We can see increase in test accuracy with the addition of more regularization.