DD2424 Lab 3

Artem Los (arteml@kth.se)

August 20, 2018

K Layer Networks with Batch Normalization

Gradient check

Instead of checking the gradients numerically, we used the assumption that if gradients are correct means we should be able to overfit the data and vice versa (this is influenced by the observations in the previous lab, i.e. when gradients were wrong, the accuracy was never higher than 10%). We also tested compared the results with assignment 2.

To test gradients, we set the goal to overfit the classifier on 100 samples using 100 epochs, a batch number of 100 and learning rate of 0.01. The final accuracy on training set was 0.95 for a two-layer network and 0.61 for a three layer network. Thus, the gradients should be correct. Later tests confirm this (once better hyper-paramters were found).

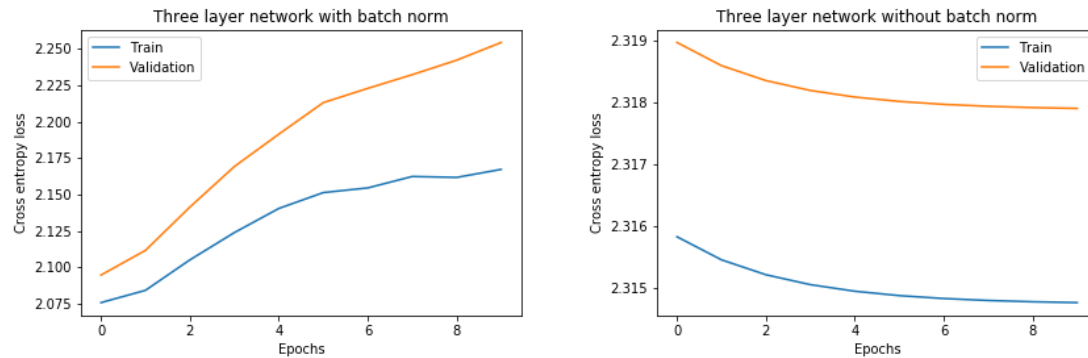### Three layer network with and without batch norm

Figure 1: Training and validation errors when a 3 layer network was trained with & without batch norm.

Best classifier

When performing fine search, we used 3 epochs and $\eta = [10^{-2}, 10^{1}], \lambda = [10^{-5}, 10^{-4}]$. As a result, the best classifier obtained $\eta = 0.128729289, \lambda = 1.55476073 \times 10^{-5}$ with 0.4252 test accuracy.

Two layer network with different learning rates

The plots below illustrate evolutions of the error for different learning rates. High, medium and small are 0.1, 0.01 and 0.01 respectively.
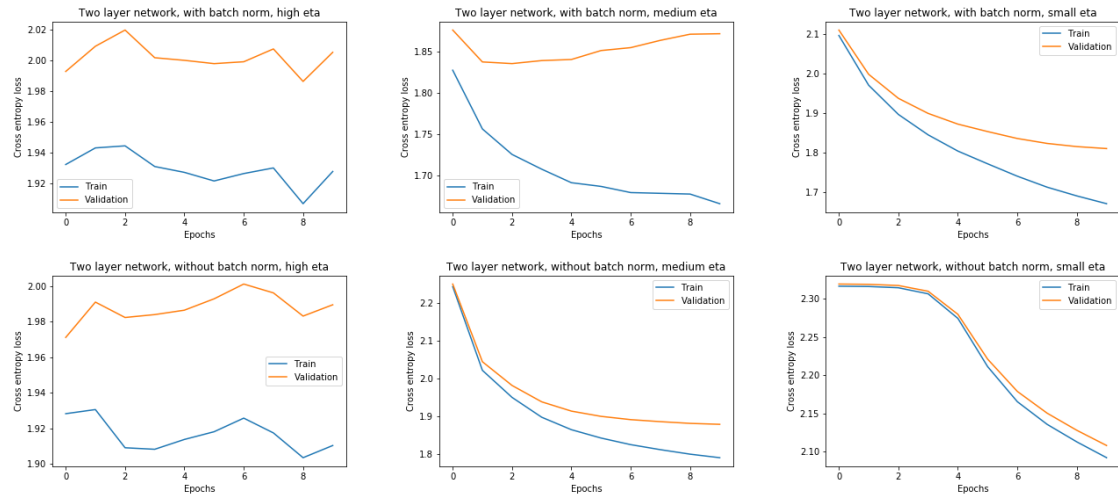
Figure 2: Evolution of two layer network for different learning rates and constant lambda. The top raw is with batch normalization and the bottom row is without.