DD 2424 - Assignment 3 Bonus

Jia Fu (jiafu@kth.se)

June 3, 2021

Find a better network architecture

In this part, 45000 data points were used for training and 5000 data points for validation. Other hyper-parameter settings: `n_batch = 100`, `eta_min = 1e-5`, `eta_max = 1e-1`, `lambda = 0.005`, and `n_s = 2250`. Two cycles of training and Xavier initialization were applied. `alpha` was set as 0.7 for batch normalization (BN) which is smaller than the typical number 0.9 since the training is shorter than usual. The results are shown in table 1.

Number of Layers	Hidden Nodes	Test Accuracy
3	[50, 50]	53.22%
3	[500, 100]	57.50%
6	[50, 40, 30, 20, 10]	53.13%
6	[500, 250, 100, 50, 20]	58.48%
9	[50, 30, 20, 20, 10, 10, 10, 10]	51.73%
9	[500, 250, 100, 50, 40, 30, 20, 10]	57.84%

Table 1: Test accuracy of models with different number of layers and hidden nodes

Thus, we can draw the conclusion that making the network deeper won't improve the performance, but increasing the number of hidden nodes will, as it can approximate some more complicated functions.

Apply BN to the scores after the non-linear activation function

It has been empirically reported in several works that better performance can be obtained by applying BN after the non-linear activation function has been applied. Hereby I adjusted my forward function `EvaluateClassifier` and backward pass `ComputeGradients` accordingly. Other parameters were kept unchanged as last section. Table 2 shows the results.

Network Architecture	BN before ReLU	BN after ReLU
3 layers with hidden nodes [50, 50]	53.22%	54.31%
6 layers with hidden nodes [50, 40, 30, 20, 10]	53.13%	53.65%
9 layers with hidden nodes [50, 30, 20, 20, 10, 10, 10, 10]	51.73%	52.19%

Table 2: Test accuracy of models with different order of BN and ReLU

The results indicate that this approach indeed brings some improvements where it can raise test accuracy by around $0.5\% \sim 1.0\%$ in our case.

Final result

Finally, a 6-layer network with hidden nodes [1000, 500, 250, 100, 30] was tested by applying BN after ReLU. Tune the amount of regularization to `lambda = 0.003819` and keep other parameters the same as before, test accuracy can reach **59.42%** after three cycles' training.