

Deep Learning Lab Assignment 1

Setting

MLP with three hidden layers:

[hidden layer 1] 100 units, 'relu' function

[hidden layer 2] 100 units, 'relu' function

[hidden layer 3] 10 units, linear function

[output layer] linear output

Hyperparameter:

using SGD, with batch_size 64, learning rate 0.1, train 20 epochs

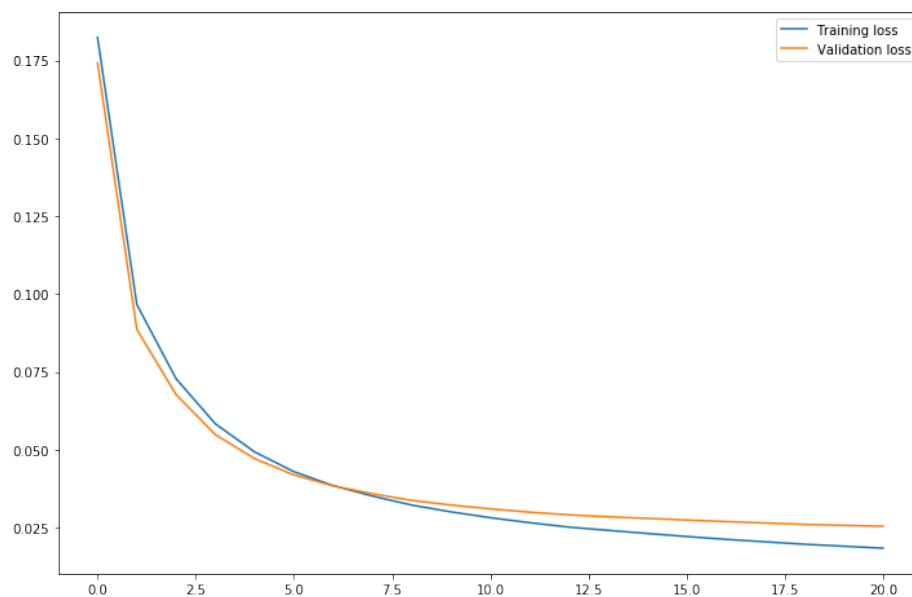
Result

Training Time: 72.8s

Training Error: 0.98 %

Validation Error: 2.19 %

Test Error: 2.15 %



Discussion

1. Comparing to 'relu', 'sigmoid' or 'tanh' produce a much higher loss. So 'relu' is chosen.
2. Using 'gd' produces a much higher loss than using 'sgd', it may need more epochs to converge. For timing reason, 'sgd' is chosen.
3. Using 'softmax output', the validation error remain high comparing to training error, while using 'linear output' both errors go down together. So 'linear output' is chosen.
4. Other tunings (like changing the number of layers, the number of units, or the size of batch) doesn't make a big difference in loss so the original settings are preserved.