Deep Learning Lab Assignment 1

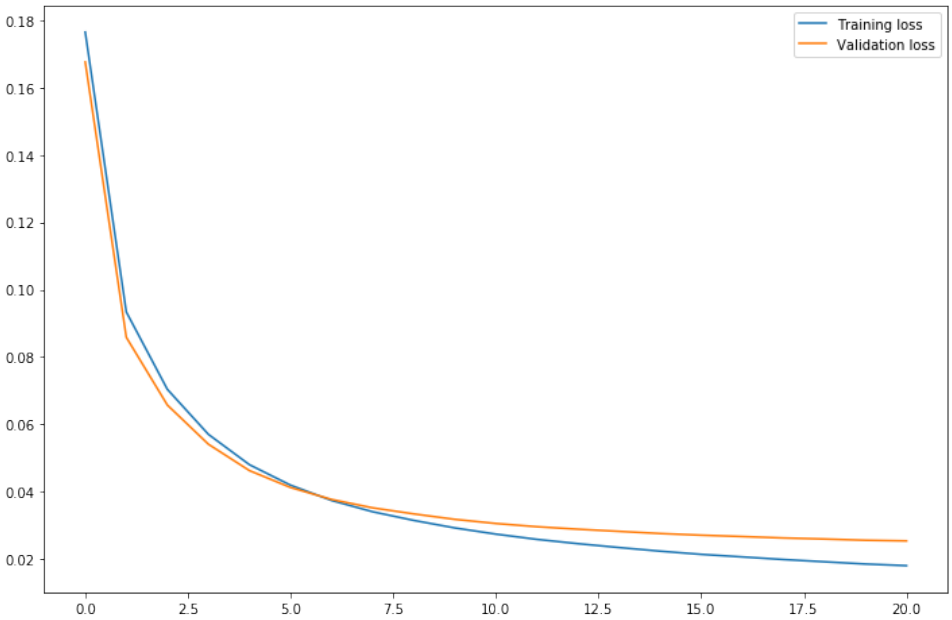
Setting

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:	74.1s
Training Error:	0.98 %
Validation Error:	2.14 %
Test Error:	2.28 %



Discussion

1. Comparing to ‘relu’, ‘sigmoid’ of ‘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. Cannot figure out the derivative of ‘softmax’, so only linear output layer has been tried.
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.