Deep Learning - Lab 4 Exercise

Han Zhang hz5g21@soton.ac.uk

1 Exercise 1

Exercise 1.1:Wide MLPs on MNIST

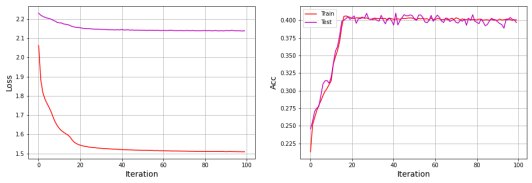

Figure 1: hidden units: 1,

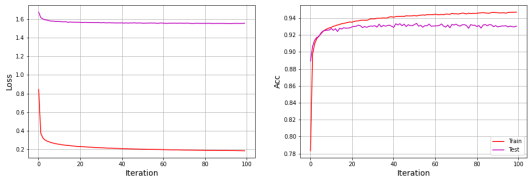

Figure 2: hidden units is 9

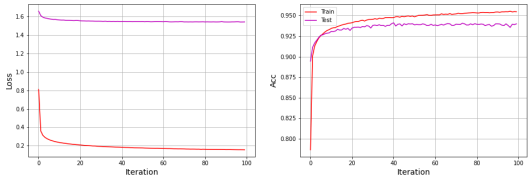

Figure 3: hidden units is 10

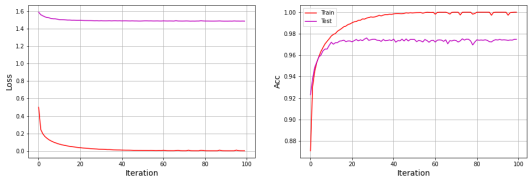

Figure 4: hidden units is 50

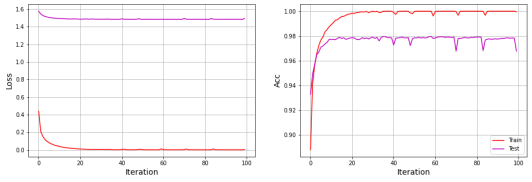

Figure 5: hidden units is 100

Generally speaking, over-fitting means that the gap between the training error, while under-fitting means that the model cannot obtain a sufficiently low error on the training set and the model's performance is bad. Their relation is shown in Figure 6.

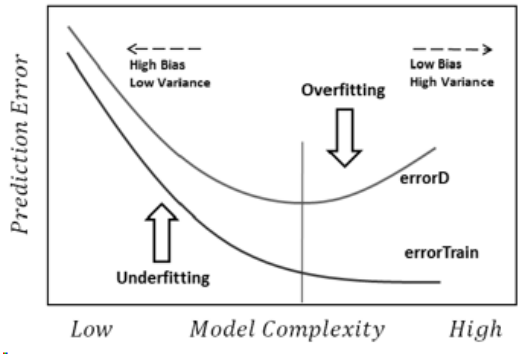

Figure 6: Over-fitting and under-fitting

As shown in Figures 1,2,3,4,5, the loss function of a different number of hidden units and the accuracy are demonstrated. Table 1 summarizes those models.

Units	1	9	10	50	100
Train_loss	1.5	0.2	0.19	0.0	0.0
Test_loss	2.1	1.6	1.58	1.5	1.48
Train_acc	0.4	0.94	0.951	1.0	1.0
Test_acc	0.4	0.93	0.930	0.975	0.98

From the results, the hidden units are around 9 and 10, the model is perfect and does not fall into over-fitting or under-fitting. It is possible because the MNIST is largely linearly separable. In single-layer MLP, a unit represents a linear discriminant function, and 10 classes can be linearly divided by 9 functions. So the classifiable samples of the dataset may be an important criterion for choosing hidden units.

It can be observed that as the hidden units become larger, the Train loss becomes smaller until it becomes zero, which means that the model becomes a memory machine and generalization ability is bad. However, in this case, the loss of the test set does not increase as Figure 6 expected. one of the reasons could be that the data in the test set is highly similar to the training set.