COMP6248 Lab 4 Exercise – Fun with MLPs & MNIST

Wei Chien Teoh (Eugene)
wct1c16@soton.ac.uk

29 April 2021

Introduction

The results are seeded using `torch.manual_seed(0)` to provide reproducible results.

1 Wide MLPs on MNIST

1.1 Wider MLPs

Fig. 1 illustrates the learning curve for different width of the MLP. It can be observed that the validation curve starts to diverge from the minima with around 5000 neurons in the hidden layer. The divergence increases as the hidden size increases after that point.

By intuition, deep networks are better at generalising because it is able to learn features with increasing level of abstractions.[1] Shallow and wide networks tend to memorise the structure of the features, thus causing them to have bad generalisation.

As the width of the MLP increases, the model complexity increases, which allows the model to approximate functions of higher order. This is useful in learning the structure of the dataset, thus decreasing the training loss. However, increasing model complexity comes with the cause of learning the structure of that particular dataset, hence will be unable to have good performance when unseen data of similar structure is given. Such a phenomenon is seen in the divergence of test loss.

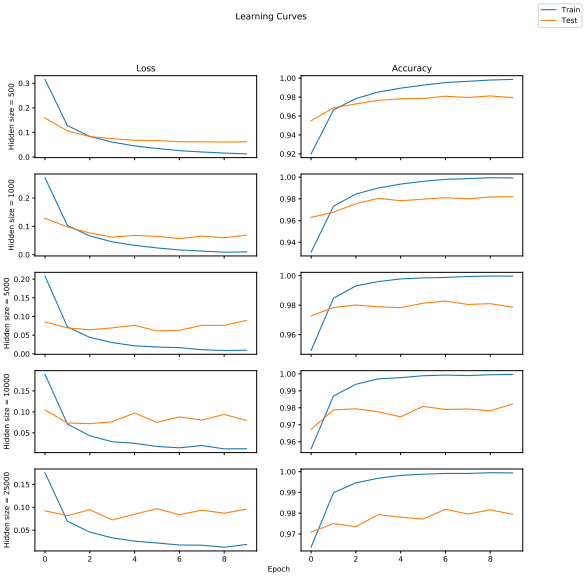

Figure 1: Learning curves for different number of hidden layer neurons. (Left) Loss Plots. (Right) Accuracy Plots.

[1] **eldanPowerDepthFeedforward2016**.