Lab 4 Exercise - Fun with MLPs & MNIST

Amishapriya Singh
Student id: 33333904
as6u21@soton.ac.uk

1. Wide MLPs on MNIST

1.1. Wider MLPs

To overfit the training data, our model needs to memorise the training examples and not be capable of generalising. A neural network that is shallow will only learn low-level discriminating features. As the depth of the network increases, the more the ability of the network to form a hierarchy of high-level concepts that can learn these discriminating patterns. If we reduce the depth of the network, we essentially cripple the network of its ability to learn high-level features at various levels of abstraction. With a single hidden layer neural network, to overfit the MLP, we could over-parameterise it by making it unnecessarily wide and fit a high order complex function through all the points in the dataset.
A wide network will have millions of parameters that can be trained to memorise the dataset exactly, reaching near 100% training set accuracy which can be validated by looking at the training curves in Fig. 1. We can see that as the size of the hidden layers is increased, the loss and accuracy on training and test set begin to diverge indicating overfitting of the model to train data.

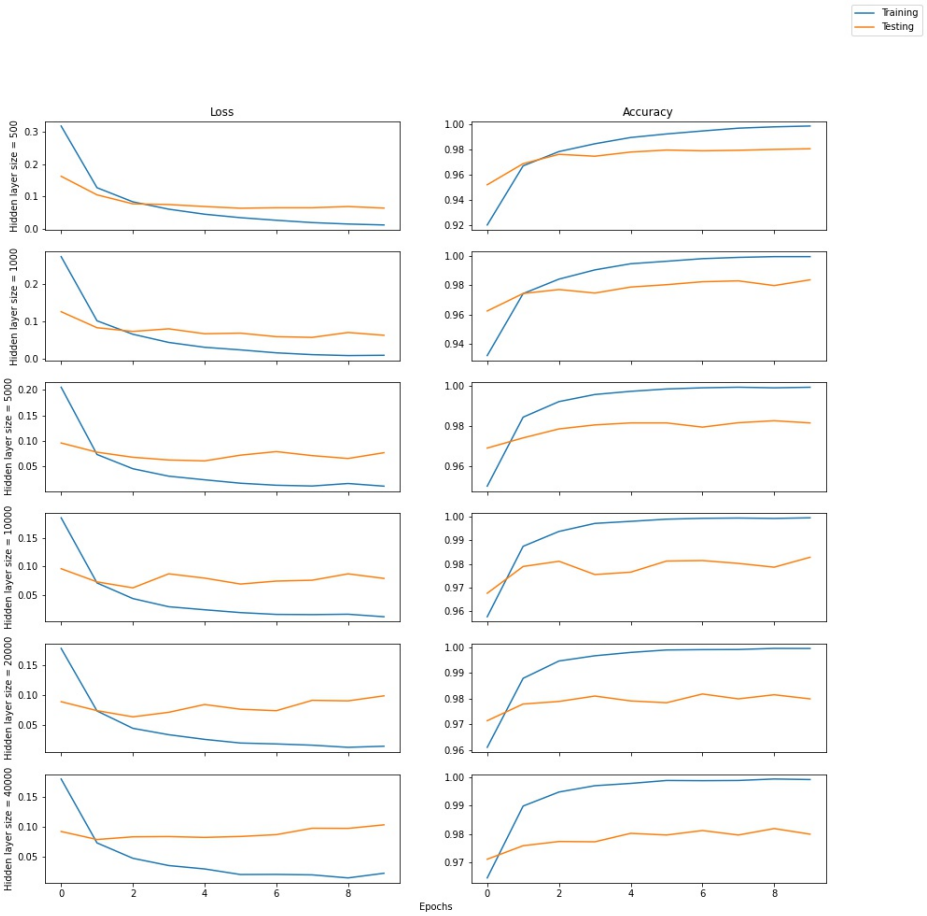

Figure 1. Training curve