Lab 4 - Fun with MLPs and MNIST

Abhilash Pulickal Scaria
aps1n21@soton.ac.uk

Introduction

An MLP with single hidden layer is constructed. The train and test accuracy for different hidden units size is noted.

1 Wide MLPs on MNIST

1.1 Wider MLPs

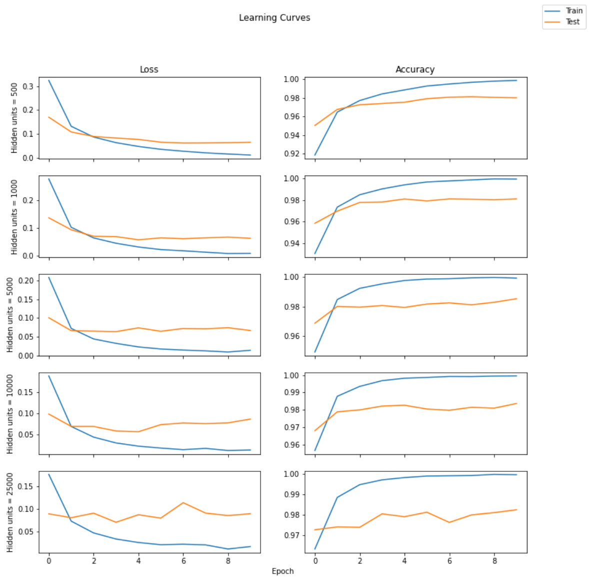

Figure 1: Plot of loss and accuracy on training and test data for different number of hidden units.

Increasing the number of hidden units increases the number of trainable parameters which can lead to overfitting of the model. We can observe from Figure 1 and table 1, the model start to lose its ability to generalise once the number of hidden units increase beyond 5000.
Initially the accuracy of both training and testing increases as we increase the number of hidden units. But as obseverd from the figure the training accuracy curve and the test accuracy curve start to diverge as number of hidden units reach 5000 and it continues to diverge. This is due to the model overfitting to the given data.

Table 1: Training and Test accuracies (After 10 epochs).

Hidden units	Training Accuracy	Test Accuracy
500	99.87	98.00
1000	99.94	98.10
5000	99.93	98.53
10000	99.95	98.35
25000	99.94	98.22

Deep Nueral networks have the ability to learn features from multiple levels of abstraction. The initial hidden layers of a DNN can learn simpler features and deeper layers can learn more complex features. This can help in genaralisation. Shallow and very wide networks are not very good at generalisation. They are very good at memorisation and ends up overfitting to the data. In practical applications we require models which can generalize and work with unseen data and shallow very wide networks which ends up overfiting to data are not preferred.