1 Wide MLPs on MNIST

Case 1: Number of hidden units = input size. This shows significant overfitting as final training accuracy is 99.5% and validation accuracy is 97.9%.

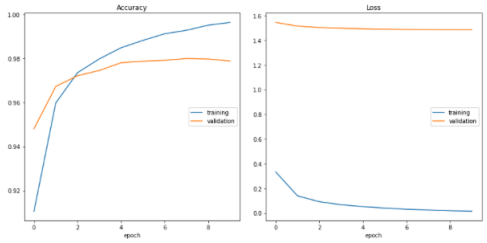

Figure 1: 784 hidden units showing model overfitting.

Case 2: Number of hidden units = $\frac{1}{16}\times$ input size. Test accuracy = 97.3%, validation accuracy = 96.8%. However this comes with the cost of lower accuracy.

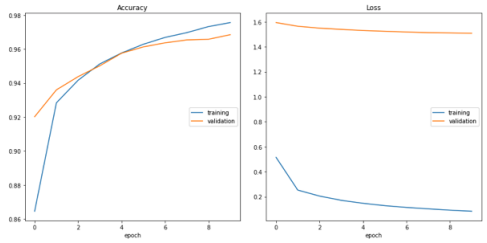

Figure 2: 49 hidden units showing marginal overfitting.

Case 3: Similar to Case 1. Number of hidden units = $2\times$ input size.

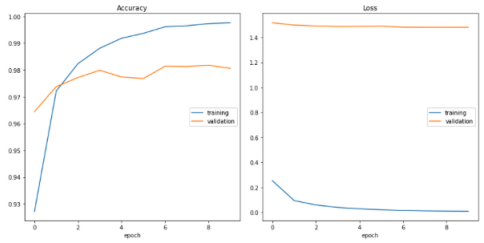

Figure 3: 1568 hidden units showing overfitting.

Increasing the number of hidden units leads to overfitting because the network can learn a more complex function that perfectly separates training set data but this form of memorisation means it performs poorly on unseen data in the validation set.