Deep Learning: Lab 4

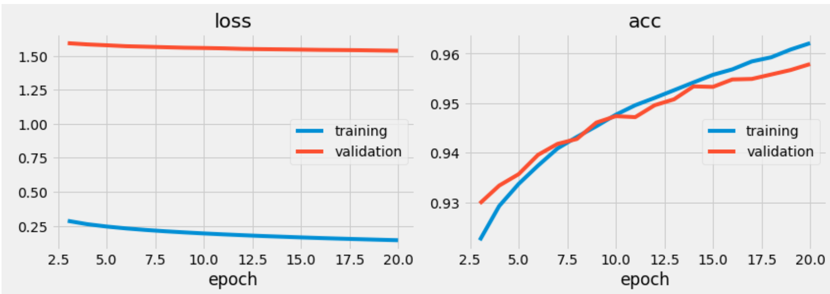
Johnny Joyce (jvjj1u19@soton.ac.uk)

1 Exploring optimisation of analytic functions

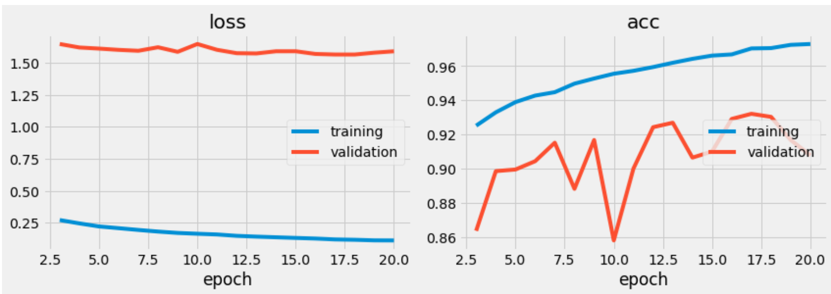
The MNIST dataset has 70,000 entries. If we were to use 60,000 of these entries for training, then a fully overfitted solution is guaranteed to exist. In such a solution, the MLP could achieve 100% accuracy by having exactly one neuron in the hidden layer activate for each unique training instance. Therefore, when seeing a new and unseen instance, the MLP will be unable to make any generalisations based on its training data.

To test this, an MLP was created based on *BaselineModel* using stochastic gradient descent (batch size=128, learning rate=0.01) with 60,000 neurons in the hidden layer, and a graph of the results across each training epoch can be seen in Figure 1(a). Although this does not create a situation as explained in the previous paragraph, it does demonstrate overfitting. The validation accuracy begins higher than the training accuracy, but the training accuracy improves at a faster rate than the validation accuracy and eventually overtakes it. Furthermore, we can see that in early epochs, significant reductions are made to the loss function across training data, but these do not translate to large reductions in the loss across the validation data.

Figure 1(b) shows the result if we use 10,000 hidden units and swap the training and validation data - that is, we use 10,000 data points for training and 60,000 for validation (note that the number of hidden units is still equal to the number training instances). This demonstrates overfitting in a much more noticeable manner since certain epochs slightly increase the training accuracy but significantly decrease the validation accuracy. This may be partially due to the reduced size of the training set, meaning that there is less information for the MLP to make generalisations. It may also be partially due to the increased size of the testing set, meaning that the validation accuracy gives a more accurate representation of the MLP's performance on unseen data. The validation accuracy on the trained network after 20 epochs was 90.7%.



(a) Results when the MLP has 60,000 units.



(b) Results when the MLP has 10,000 hidden units. Here, the training and testing sets were swapped, giving 10,000 training instances and 60,000 testing instances

Figure 1: Graphs of the loss and accuracy over time for the training and testing set various situations.