Lab 4 Report

Jack Hilton-Jones (jhj1g23@soton.ac.uk

In this lab a Baseline Model was built, this consisted of a simple MLP with one hidden layer, which was trained on the dataset MNIST. The MNIIST dataset is a large database of handwritten digits, making it useful for training models. In this lab, the investigation of the amount of hidden layers within an SGD model before the model overfits the data is conducted. Cross Entropy is used as the loss function and Stochastic Gradient Descent (SGD) is used as the optimiser with a learning rate of 0.01. The data is split into training data and testing data, it is trained on a model with 784 inputs with 10 outputs for the 10 numbers 0-9. The amount of hidden layers was altered and ranged from 500 to 100000, this provided a large range of data to conclude the optimum amount of hidden layers to train this model. The model was used to calculate the accuracy of the training model and the test model, which evaluates the number of correct decisions compared to the total number of data points. In addition, the training and test loss is calculated. The plot in Figure 1 compares the train and

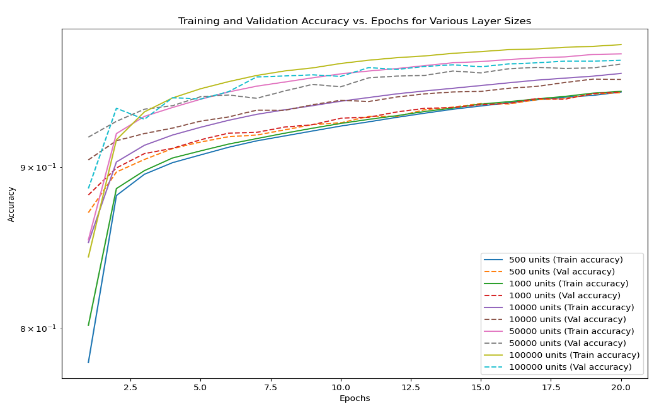

Fig. 1. Train and test accuracy against epochs

test accuracy when increasing the number of hidden layers in the Baseline model. Note this is plotted on a log scale. Each hidden layer size was run for 20 epochs. It is seen from the graph that the training accuracy continues to increase, eventually plateauing at a higher value. In comparison, the validation accuracy has a lower rate of increase and levels off at significantly lower values than the training accuracy. As the hidden layers increase there is evidence that the model is overfitting due to the divergence between the training and validation accuracy. The model can perform exceptionally well on the training data with an increase in the number of hidden layers but not generalising equally well on the validation set. In addition, the training and test loss against the epochs is plotted on a log scale to improve readability in Figure 2, this highlights the plateauing of the test loss in comparison to the constant training loss decrease. The model is overfitting as it is becoming more complex with the increase of hidden layers, it starts memorising the training data and fits to the noise. This reduces its ability to generalise, leading to a smaller reduction in validation loss compared to the training loss seen in the graphs. The layer size against the training and test loss and accuracy is plotted on a log scale in Fig. 3. It is observed that the points intersect at around 5000, this is the point where beyond this layer size the training loss and training accuracy begin to overfit as the train and test loss and accuracy begin to diverge. With an increase in layer size, the model becomes significantly more computationally expensive and longer to train with not much more value in test loss and accuracy.

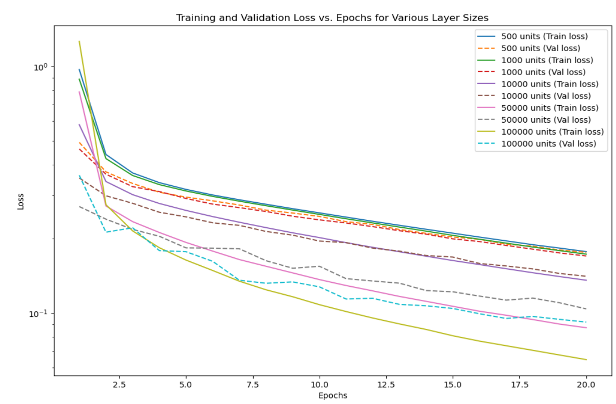

Fig. 2. Train and test accuracy against epochs

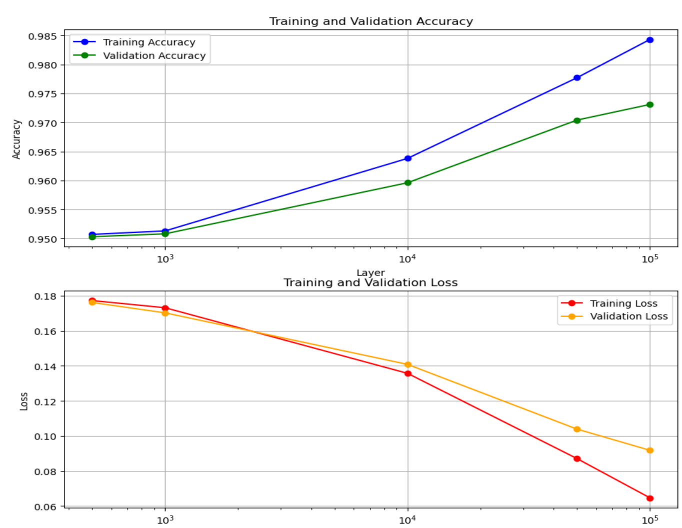

Fig. 3. Training and validation accuracy against the number of hidden layers