Lab 4 Exercise - Fun with MLPs & MNIST

Jonathon Hare (jsh2@ecs.soton.ac.uk)

March 3, 2023

This is the exercise that you need to work through **on your own** after completing the forth lab session. You'll need to write up your results/answers/findings and submit this to ECS handin as a PDF document along with the other lab exercises near the end of the module (1 pdf document per lab).

You should use *no more* than one side of A4 to cover your responses to *this* exercise. This exercise is worth 5% of your overall module grade.

1 Wide MLPs on MNIST

In the lab exercise you built a `BaselineModel` which was a simple MLP with one hidden layer and trained it on MNIST. You're now going to explore this model further.

1.1 Wider MLPs (5 marks)

Consider an MLP with a single hidden layer. How wide does the network need to be (how many hidden units) before it overfits[a] on the MNIST training dataset? Provide rationale (1 mark) and experimental evidence for your findings (see below), and suggest reasons why they might be so (1 mark).

For practical purposes you're limited by available GPU memory; don't try training networks with more than 500,000 hidden units, which have almost 400 million learnable parameters! For experimental evidence you should include training curves (plots of loss and accuracy against epochs) with both the training and test data (which you are technically treating as validation data for this experiment) for a range of different sized models (3 marks).

---

[a]fails to generalise *reasonably* to the MNIST test set