Lab 4 Exercise - Fun with MLPs & MNIST

1. Introduction

This exercise is to study how large does the hidden layer of an MLP could be ahead of overfitting on the MNIST dataset.

To determine a threshold over which MLP overfits, my diagnostic approach is through observing the change of curves from different model. Figure 1 shows the validation error on a well-trained model is a U-shaped curve, vise versa for the accuracy. The bottom of the test sample curve is where a model begins overfitting. Our task is to find this by comparing different models.

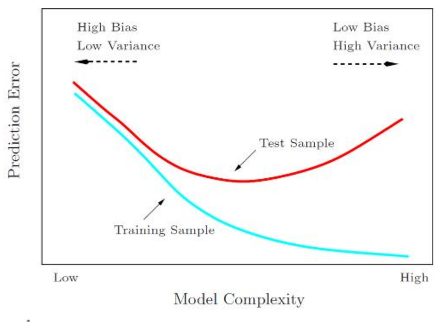


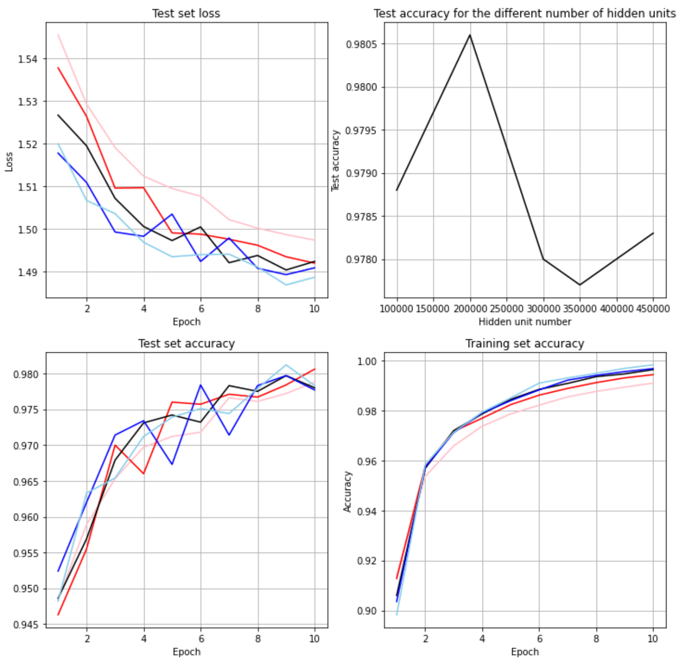
Figure 1: The Relationship between Model Complexity and Prediction Error

2. Experiment

The following graphs display different models' performance on MNIST dataset, with pink line being the 100,000 units model, red one 200,000, black one 300,000, sky-blue one 350,000 and blue one 450,000. For the training set accuracy on the bottom right graph, it can be seen that a model's final training accuracy (by 10th epoch) goes up as its unit number is augmented. This phenomenon corresponds to the theorem mentioned. Consequently, now we only need to find the point where test accuracy begins to fall and that will be the initiation of overfitting.

For the test accuracy graph on the bottom left, these models perform in a different way. The red one (200,000) becomes the highest one with over 98% accuracy, overtaking the black one (300,000), sky-blue one (350,000) and blue one (450,000). For better viewing, each curve's terminal is redrawn and plotted in the upper right graph.

At 200,000, clearly there is an overfitting signal since either toward the left and right, the test accuracy goes down.



3. Analysis

So, what happen after model complexity becomes too high? The training accuracy of a model increases with its complexity since its complexity allows the model to learn more information from the training set. Arguably, this information includes both random noise and the data pattern behind training data. When complexity does not exceed the overfitting threshold, most information a model learns will be the data pattern which contributes the most of the loss. And under the model complexity that just begins to overfit, a model is able to learn the most pattern in data and a rational level of random noise. However, once this threshold is exceeded, a model can and will learn random noise too much so that its test error will rise again, that is the case above.