CS 578 HW1 Report

Linxiao Bai

Introduction:

After implemented the DNN as required, the performance of the model is tested on the test dataset. Performance is evaluated from aspects of:

1. Test Accuracy.
2. Training Time Cost.

While the controlled variables are:

1. Depth of the NN.
2. Width of the NN.

Methodology:

The experiment is carried out by reconstructing NNs with change of number of hidden layers from 1-10. For each layer. The width is also a ranged change from 10 to 100 nodes, with a step increment of 10 and a random standard deviation of 5.

The input layer and out layer is fixed to be 764, and 10 input/outputs.

For example, a valid sample of NN looks like:

4 hiden layers:

input: [10*24]
H1:[24*23]
H2:[23*20]
H3:[20*17]
H4:[17*23]
output: [23*10]

Notice that at each hidden layer the output is between 5 standard deviation of 20.

The idea is to enumerate combination of the width and depth of DNN as much as possible with a controlled variation. For each generated DNN, training time and accuracy is recorded individually. 3D plots of precision and time against width and depth is generated to visualize the result. The rest of the report will focus on inference of depth and width' s impact on the model.

Test Accuracy:

Figure 1 shows how test accuracy changes with width and depth. From the figure, at low depth and low width, accuracy reaches a high status. The best accuracy point is around the approximate region with depth=3, and width=30.

Also, the performance of the model decreases sharply once the width and depth cross a separation (As shown in the blue area). The high decrease rate of accuracy at the separation line indicates a strong overfitting potential once the complexity of the model reaches a threshold.

Figure 2 also shows that for this data, once the model-depth exceeds 6, and width exceeds 70. The performance of the model drops to a flat valley with an accuracy of around 0.1 which is essentially blind guessing a number from 10 outcomes.

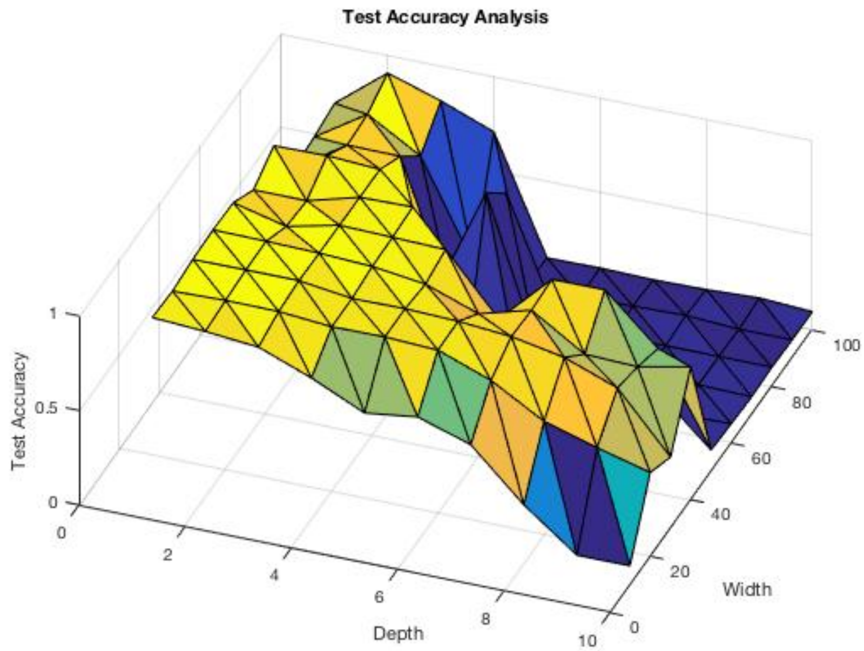

Figure. 1: Overall View of How Accuracy Changes with Depth and Width

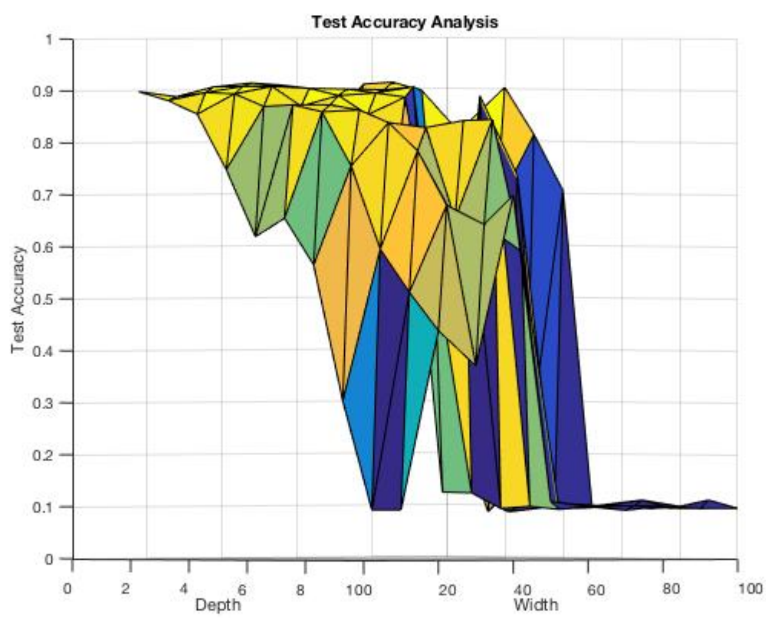

Figure. 2: The Valley of the DNN, Pay Attention to The Bottom Right Corner.

Training Time Analysis:

Similar to the Test Accuracy part, the training time trend is also plotted against width and depth. Figure 3 shows the general trend of changes. In general, the time complexity increases as width or depth increases.

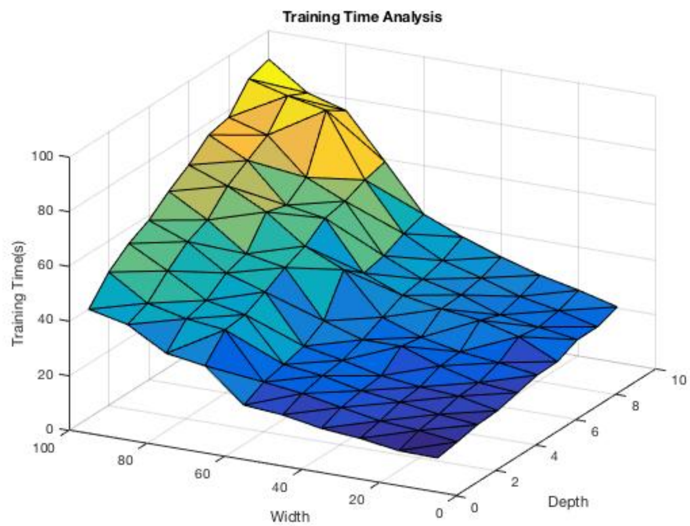

Figure. 3: Overall View of How Training Time Changes with Depth and Width

Also, Figure 4, and 5 shows the result after marginalize time against Width or Depth. The result shows that training time complexity is approximated to be a linearly increment of depth. However, it is a none linear increment of width. Figure 5 especially shows a singular point at width=60. The rate of time over width increases at this point. This suggests that to increase the width of a network greater than 60 by one unit, the time complexity cost is higher than ever.

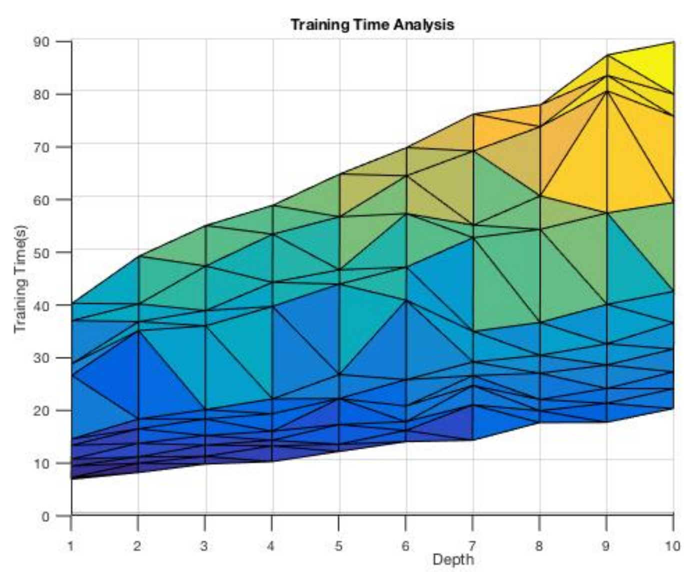

Figure. 4: Training Time Against Depth, A Linear Increment.

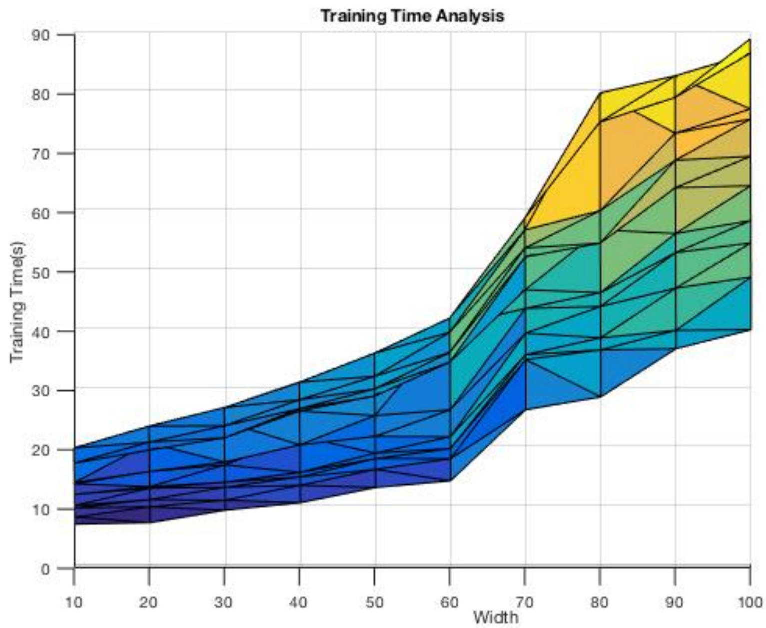

Figure. 5: Training Time Against Width, A Non-Linear Increment. Change Point at 60.

Conclusion:

This report analyzes the impact of DNN model complexity on model performance. Model complexity is measured from factor of width and depth, while performance is measured by test accuracy and training time.

The result shows that as depth and width increase, test accuracy drops and training time increases. That is, model complexity has bad impact on the model performance, at least in this problem/ data being studied.