DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
UNIVERSITY AT BUFFALO

CSE 574 Programming Assignment 1
Handwritten Digits Classification

Group 100
Luigi Di Tacchio Zakieh Hashemifar Alexander Simeonov
March 5, 2015

1 METHODOLOGY

We conducted several simulations to determine how the network parameters (λ and number of hidden units) influence accuracy and training time.

We started from 4 hidden nodes and varied λ from 0 to 1 in increments of 0.1. Then we increased the number of hidden units to 8, 12, 16 and 20 while continuing to vary λ with the same pattern. We measured training, validation and test set accuracy and network training time for all the cases and plotted the results against the network parameters. Then, based on the results, we selected the parameters that give the best performance.

The next sections contain our observations, together with the most significant plots we derived from the data.

2 RELATIONSHIP BETWEEN λ AND PERFORMANCE OF THE NETWORK

In figure 2.1, we plotted the performance of the network against different values of lambda, for different numbers of hidden units.

While increasing the value of λ, we expected the accuracy on the training to decrease; at the same time, we expected the accuracy on the test set to increase. Whereas we observe some kind of the expected behavior for some number of hidden units (for example with 12 hidden nodes, like figure 2.1 shows), it's not always like that.

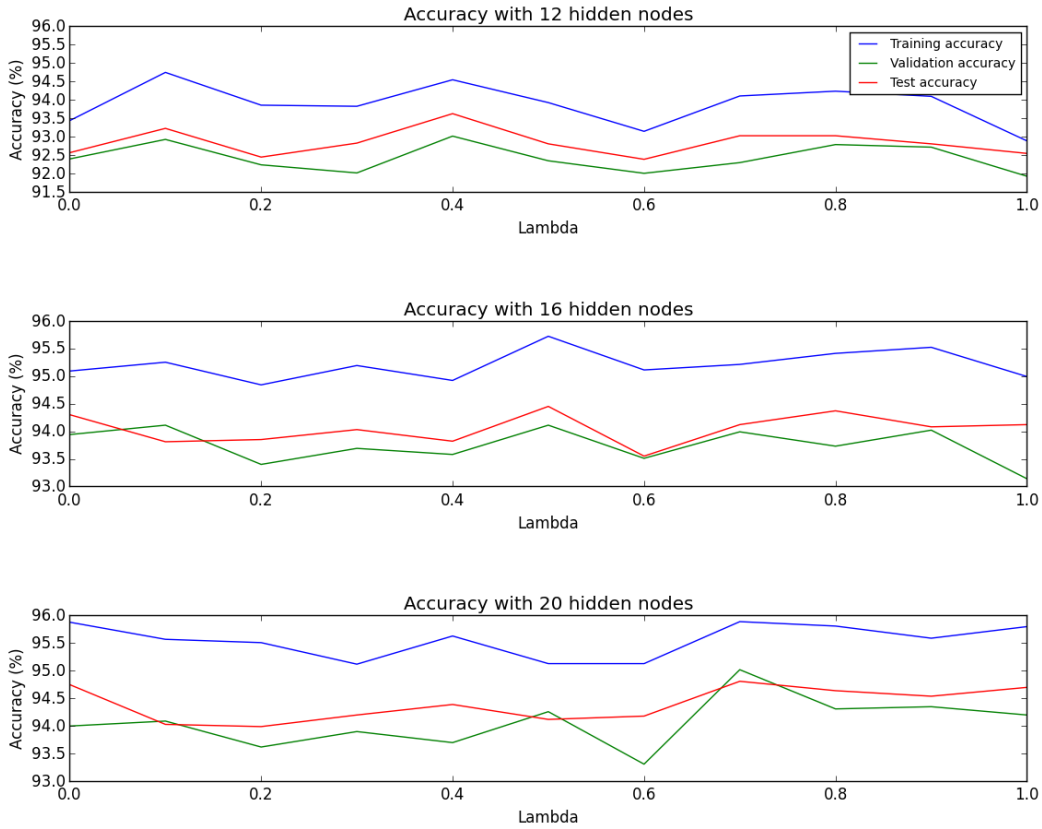

Figure 2.1: Performance of the network for different values of lambda

We were puzzled and tried to think about the reason of such outcome. We believe there may be two main causes: the similarity of the test set to the training data and the fact the we randomly chooses the training samples and the initial weights.

In particular, the similarity between training and test data may cause the performance on the test set not to be affected by overfitting on the training set, and vice versa.

Overall, we noticed that there are no significant changes in performance due to lambda variations when the number of hidden nodes is high.

3 Relationship between number of hidden units and training time

Figure 3.1 shows the relationship between training time and number of hidden units. We set the maximum number of iterations to 100.

As it can be seen in the plot, there is a direct relation between the number of hidden units

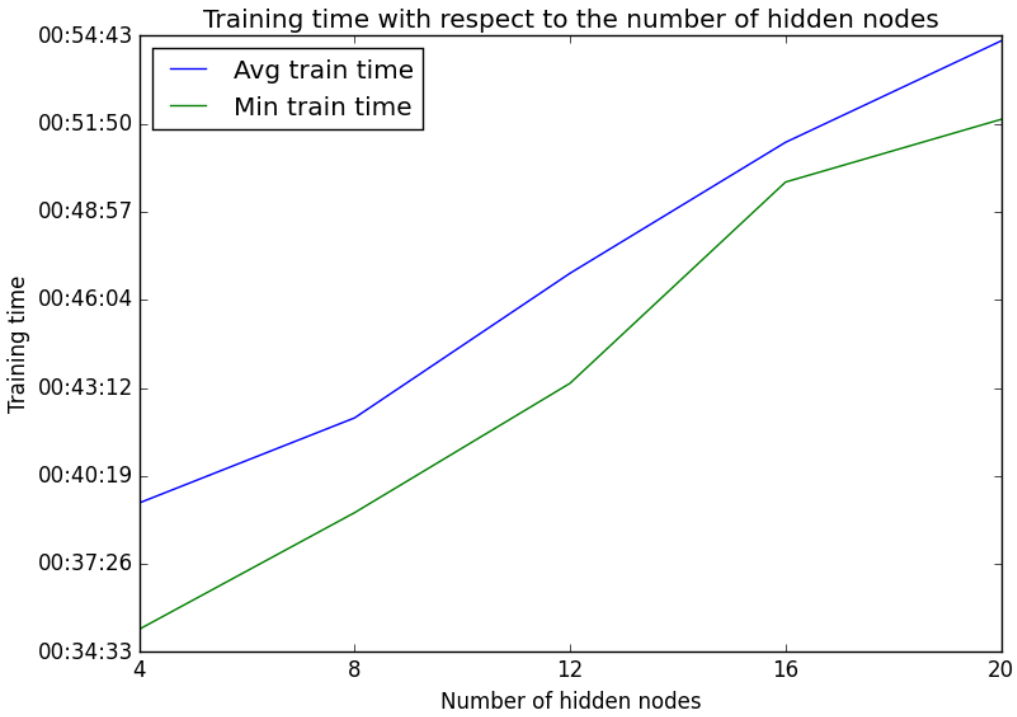

Figure 3.1: Training time for different number of hidden units

and the training time. As the number of hidden units increase, the training time grows too.

The reason is that more hidden units implies more weights and gradients to compute: this adds a fair amount of computational complexity to the network, and a longer time is needed to converge to the optimal solution.

4 Relationship between number of hidden units and performance of the network

In figure 4.1, we plotted the accuracy of the network for different numbers of hidden units.

As we increase the number of hidden nodes, we see that the performance of the neural network improves. This is to be expected as a larger number of hidden units provides for higher precision.

As it can be seen in our plots, when the number of hidden units is very low, the performance is fairly poor and we get mixed results. However, as the number increases, the accuracy improves and after a while gets more steady.

Performance seems to level off at around 95% once we reach 20 hidden units.

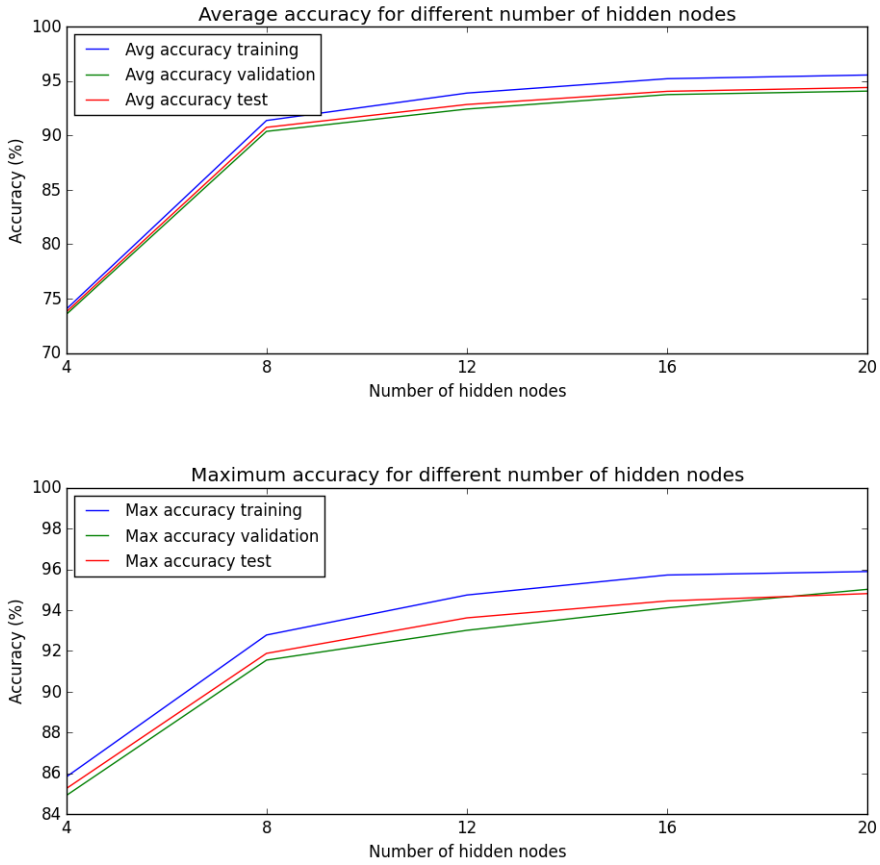

Figure 4.1: Accuracy for different number of hidden units

5 Choosing the optimal number of hidden units and value of λ

As shown in the plots, we tried different combinations of number of hidden nodes and values of λ in order to determine the optimal settings.

If time is not an issue, the best number of hidden nodes (among the ones we included in our plots) is 20. The accuracy will slightly improve increasing the number of hidden units, but it will saturate after a while. We tried up to 50 units in our experiments and we did not notice any major improvement. If time is instead an issue, then 8 to 12 hidden nodes seems a good compromise between accuracy and training speed.

The optimal value of λ seems to change when we increase the number of nodes, although we are not able to say how it changes.

On a 20 hidden nodes network, we believe by inspecting the plot that the optimal value is around 0.7, where the accuracy on training, validation and test set are all over 94.5%.