PA2 ASSIGNMENT REPORT

JUSTIN CAPPIELLO
JORDAN CHAU
TANISHA DEBASISH

CSE 474
GROUP 29

<u>Evaluation of hyper parameters for neural network</u>:

For regularization term λ values from 0 to 60 were used with a step size of 10. The test and train accuracy were then plotted against λ giving the results below:

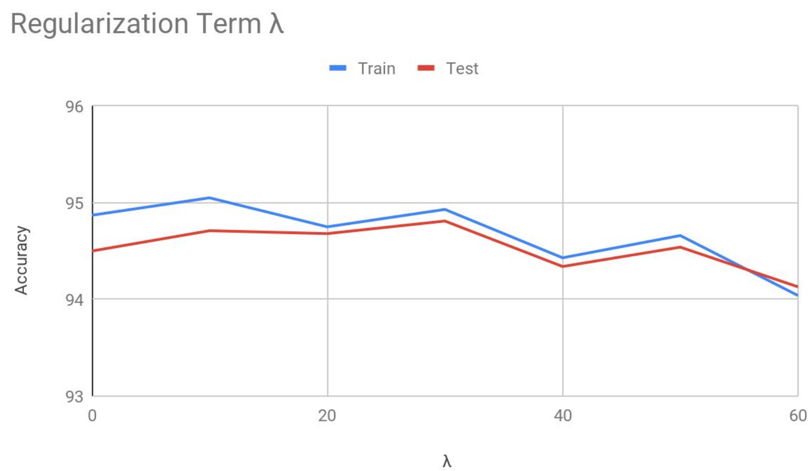

Fig 1.

It can be seen that the optimal value appears to be 30 as that is where the train accuracy reaches a maximum. The use of λ is to avoid overfitting, when the value of lambda is high, larger weights get penalized and have less influence. When it is too high, in this case >30, the data gets under fit and large weights do not have a large enough influence to maintain accuracy.

For the number of hidden units, the default lambda value of 50 was used and the range was 0 to 100 with a varying step size, initially small but increased as data was collected.

PA2 ASSIGNMENT REPORT

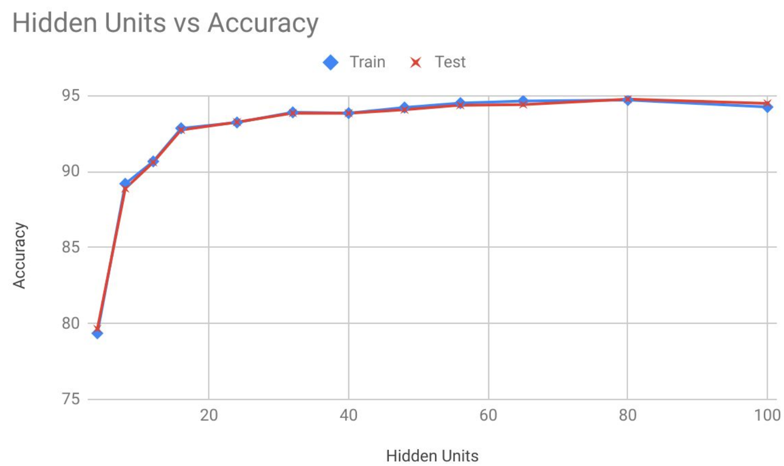

Fig 2

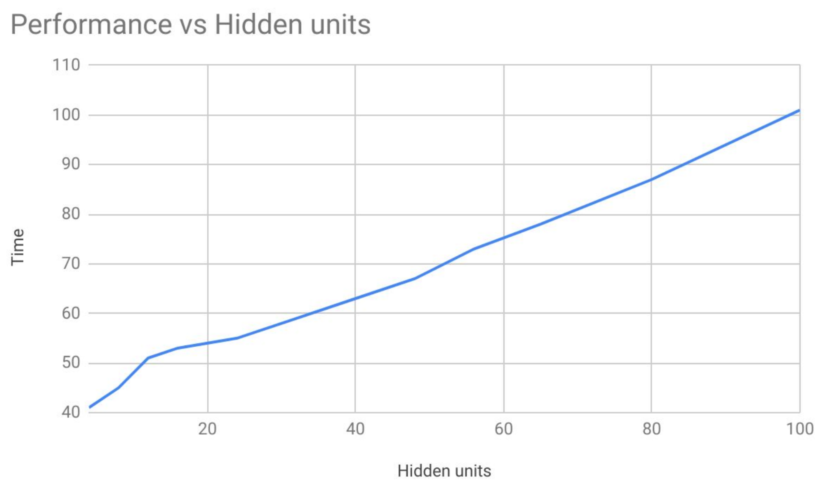

Fig 3

Fig 2 displays the relationship between train/test accuracy and the number of hidden units. It can be seen there is a sharp increase in accuracy initially however, at around 50 it levels out leaving little need for additional units. Fig 3 has the number of units compared to total runtime. As expected there is a linear relationship meaning as we add more units we will always add runtime. Overall 50 appears to be the best value as it is around the point where accuracy stops increasing and runtime is not significantly high. From this experiment we can conclude that overall the best way to chose hyper-parameters is to control other variables and analyze the effect changing one has on the result. After determining how important performance vs accuracy is to the end product, one can then use similar methods to those in this experiment to maximize the potential of the neural net and avoid under or over fitting the data.

PA2 ASSIGNMENT REPORT

Evaluation of deep neural network:

The deep neural net uses hidden layers to help improve performance and we can show this effect by running the net on the same data set multiple times, each with a different number of layers. This neural net data set took significantly longer to train and test then the previous one, which is seen below.

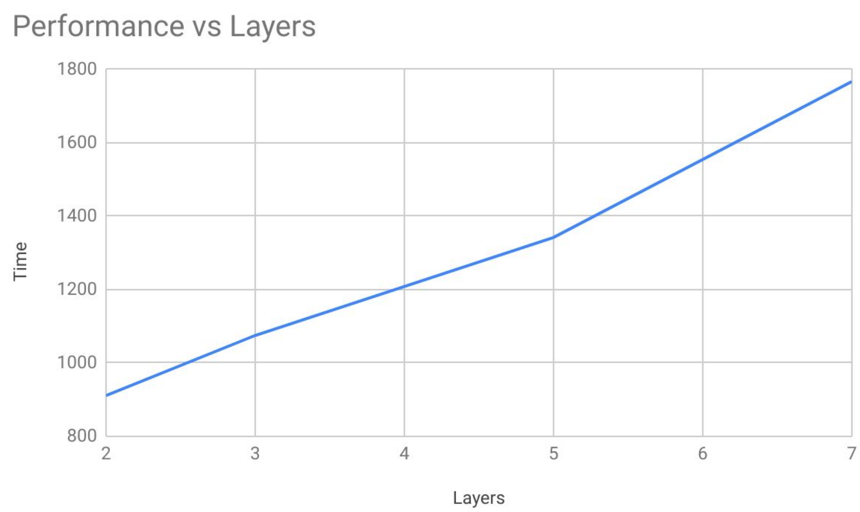

Fig 4

As the number of layers increases, the amount of time to process the data increases at a almost linear rate. This is due to the fact that additional layers require many more computations of weight for the additional node as well as their effect on other nodes.

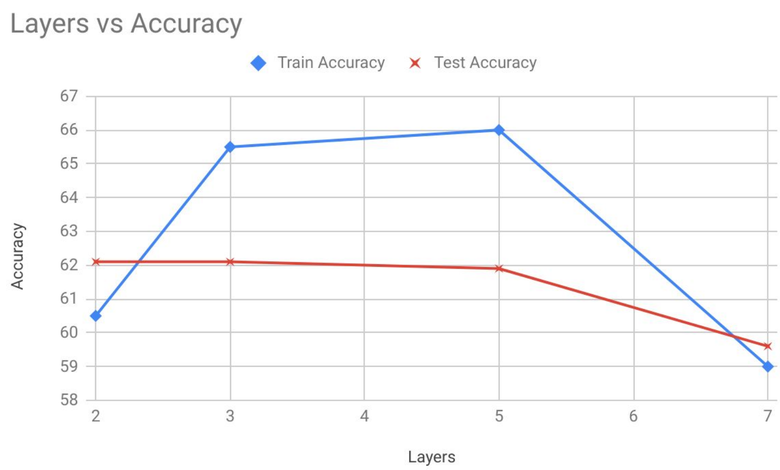

Fig 5

This next chart displays the effect additional layers has on both test and train accuracy. The train accuracy increases until it reaches 5 layers where it drops significantly. Test accuracy is relatively stable however, it also decreases after 5 layers. These results may demonstrate overfitting of data above 5 layers. The neural net places to much importance on the train data which causes a loss of accuracy when the same method is applied to the test data. Overall taking into account performance and accuracy 5 layers appears to be the best fit for this data. From this experiment we can conclude that it is vital to test multiple layers before deciding which to use. In addition other factors such as desired performance may have a trade-off with overall accuracy as many layers will dramatically increase the train and test time.

PA2 ASSIGNMENT REPORT

Data

Fig 1 data			
lambda	Train	Test	time
0	94.87	94.5	72
10	95.05	94.71	70
20	94.75	94.68	69
30	94.93	94.81	75
40	94.43	94.34	75
50	94.66	94.54	76
60	94.04	94.13	73

fig 2/3 data			
Hidden units	Train	Test	time
4	79.35	79.64	41
8	89.2	88.88	45
12	90.67	90.61	51
16	92.85	92.74	53
24	93.23	93.26	55
32	93.9	93.85	59
40	93.85	93.84	63
48	94.22	94.07	67
56	94.51	94.37	73
65	94.65	94.41	78
80	94.71	94.77	87
100	94.25	94.49	101

fig 4/5 data					
layers	Train Accuracy	Train Loss	Test Accuracy	Test Loss	time
2	60.5	1.15	62.1	1.15	911
3	65.5	1.04	62.1	1.41	1075
5	66	1.15	61.9	1.15	1342
7	59	1.24	59.6	1.2	1767

PA2 ASSIGNMENT REPORT