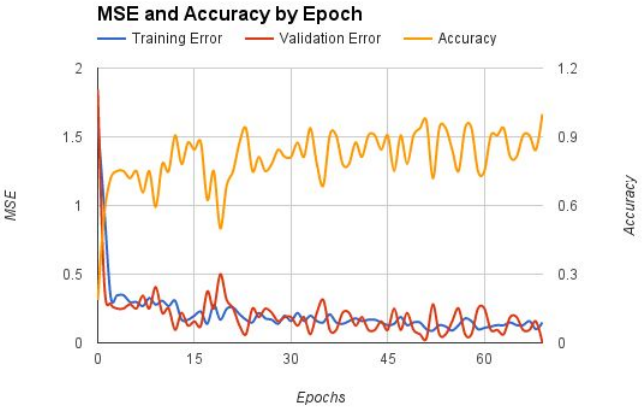
Isai Mercado Oliveros
February 13, 2016

Backpropagation Lab

Validation Set VS Test Set

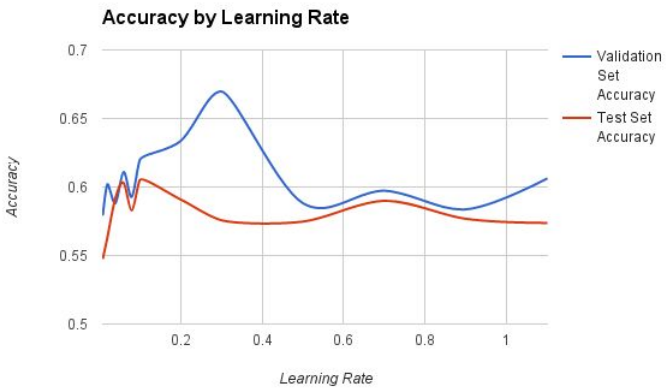
- Use iris dataset
- Use one layer of hidden nodes
- Number of hidden nodes being twice the number of inputs.
- Learning rate of .1.
- Create one graph with the MSE on the training set, and MSE on the VS.
- Classification accuracy of the VS on the y-axis, and number of epochs on the x-axis.



This chart shows how the mean square errors of the training set, and of the validation set go down while the accuracy of the test set increases. I use this behavior as the stopping criteria of the neural network. I stop when the error of the training set is lower than the error of the validation set which means that the model is probably overfitting, and the accuracy of the validation set decreases for 8 consecutive epochs which means that it is for sure overfitting.

Learning Rate

- Use the vowel data set
- Select features you should actually use.
- Use one layer of hidden nodes. Number of hidden nodes twice the number of inputs.
- Try some different learning rates (LR).
- For each LR find the best VS solution (in terms of VS MSE).
- Create one graph with MSE for the training set, validation set, and test set
- Classification accuracy on the validation set, and test set.
- Create another graph showing the number of epochs needed to get to the best VS solution on the y-axis with the different learning rates on the x-axis.

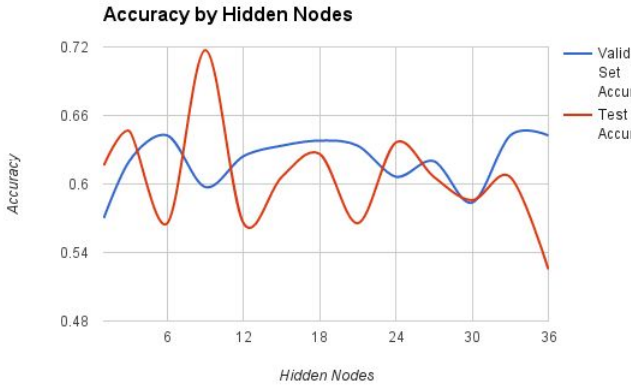


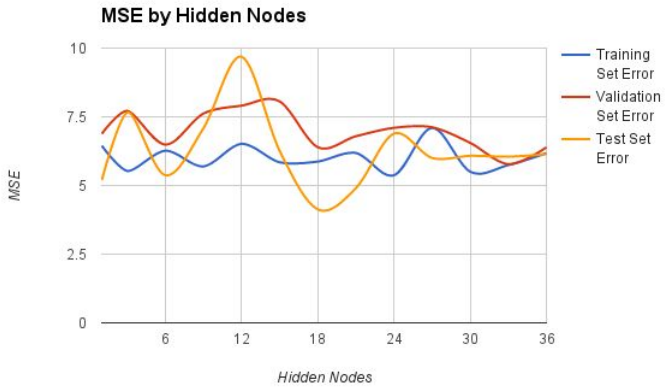
From the first chart we can see that the test set accuracy is higher at learning rates 0.1 and 0.7. In addition, from the second chart, we see that the error is lower at 0.1, so I chose 0.1 as the best learning rate to continue the experiments.

From this chart we can see that learning rates from 0 to 0.1 can cause the epochs and accuracy to jump all over the place, and in general the accuracy decreases as shown in chart 1. In addition, we can see that with higher learning rates, the number of epochs needed decreases. I think that happens because the weights approach to the global minima faster, so the neural network needs less epochs. However, accuracy decreases because with high learning rates, weights change too much so they might unlearn what they learned in past iterations.

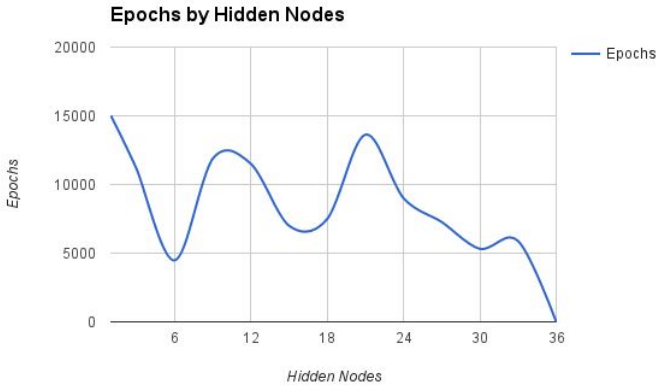
Hidden Nodes

- Using the best LR you discovered, experiment with different numbers of hidden nodes.
- For each number of hidden nodes find the best VS solution (in terms of VS MSE).
- Graph as in step 3 but with # of hidden nodes on the x-axis.





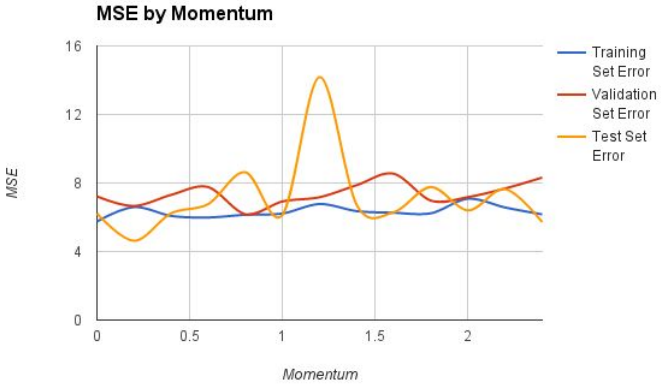
From the last experiment, I selected the learning rate 0.1 as the best learning rate, so this experiment continues using it. From the first chart, we can see that the accuracy of the test set increased dramatically when the number of hidden nodes, in a one hidden layer neural network, was 9. However, from the second chart we can also see that the error of the test set was bigger when the network had 9 hidden neurons. I actually do not know what that means, but I could not ignore that the accuracy of the test set jumped from 60% to 70%, so I selected 9 hidden neurons as the best number of hidden neurons.

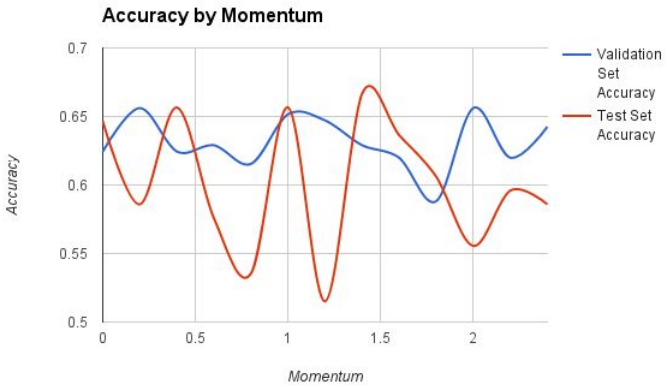


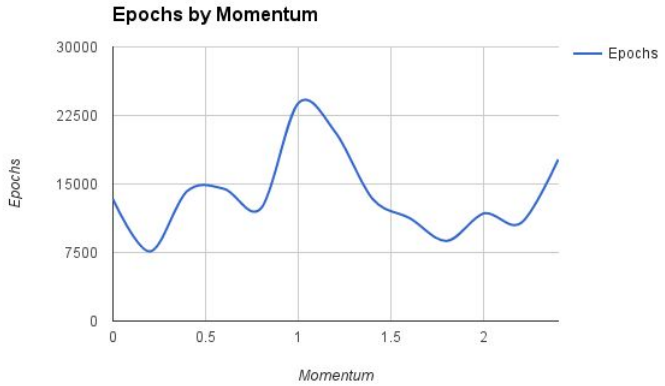
From this chart It seems like by adding hidden nodes, the need for epochs decreases. I think, I need to experiment with more data sets, learning rates, etc to confirm that this is the case, but in this experiment I observed less epochs with the increment of hidden nodes.

Momentum

- Try some different momentum terms in the learning equation using the best number of hidden nodes and LR from your earlier experiments.
- Graph as in step 3 but with momentum on the x-axis.



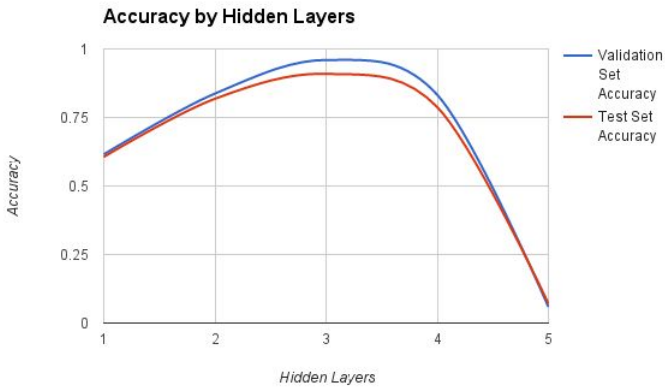


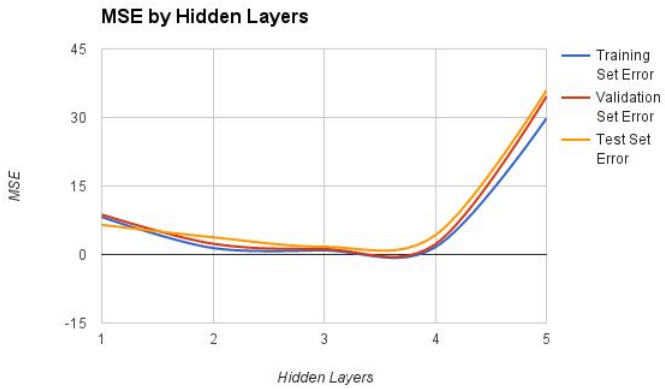


I know that the goal of the momentum is to make the neural network converge faster. However, from the Epochs by Momentum chart, we can see that the number of epochs increased from momentum 0 to 1, and then it decreased. I cannot explain that since I was expecting the number of epochs to go down from 0 to 1, and then to go up from 1 to infinity since the connections' weights in the neural network would be jumping more while trying to converge. However, the chart shows that from 1 to 1.7, the number of epochs decreases, so I chose a momentum rate of 1.4 because from chart 2, we can see that at that rate, the test set accuracy is higher, and it does not have a lot of epochs, so I chose that momentum rate.

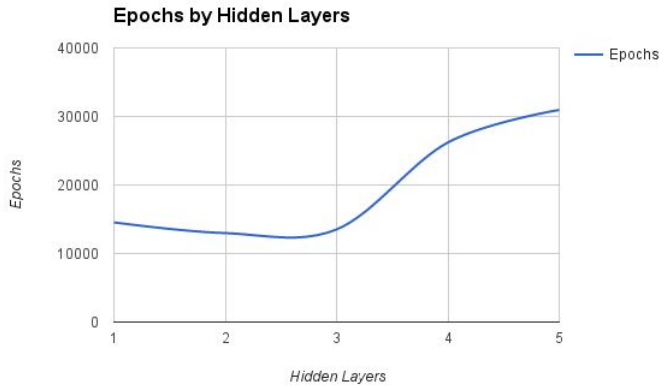
Hidden Layers

- Do an experiment of your own. (I had LR = 0.1, HN = 9, Momentum = 1.4)





I tested adding hidden layers, and I found that with 3 hidden layers, the accuracy of the test set can get as high as 0.90 which is a big improvement over the 0.70 that I was getting from having 9 hidden neurons



This Epochs by Hidden Layers chart shows that after 3 hidden layers the number of epochs increases significantly. With 4 hidden layers, it takes longer to converge to a solution, but the solution is still as high as 0.90; however, with 5 hidden layers, the number of epochs increases and the accuracy of the test set decreases to 0.70