Brandon Death 101017130
Matt Walker 100980836

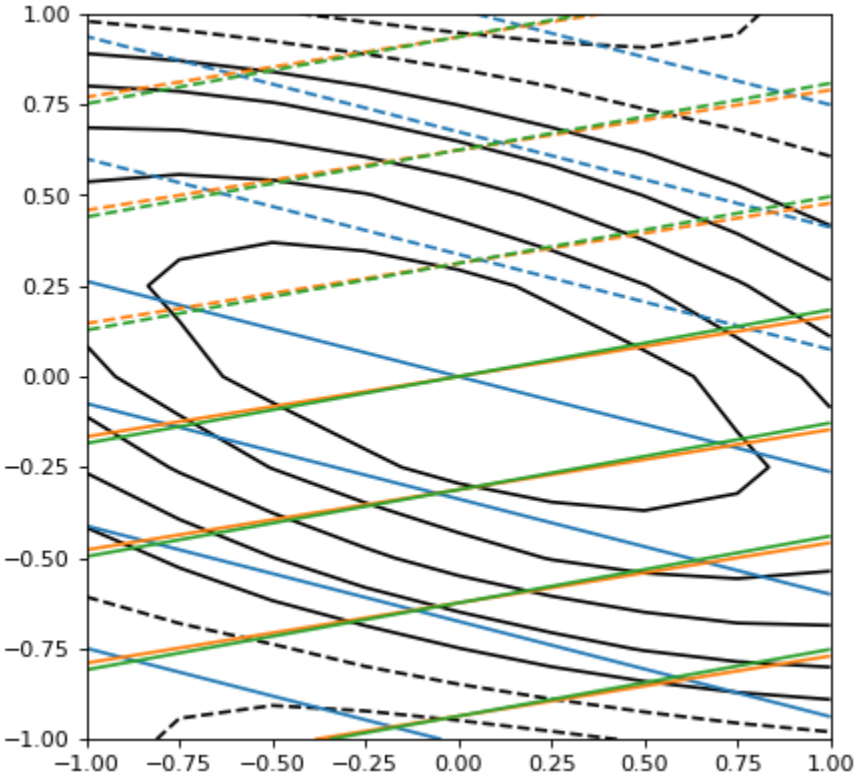
Oct. 28, 2018.

Assignment #2 Answers

Question 1

a) We had a difficult time getting the MSE to converge lower than 0.29790599. We used a learning rate of 0.1 because anything higher would cause the MSE to diverge, and any amount lower and the MSE would start at the minimum and wouldn't decrease at all. Larger batch sizes also cause the MSE to start at the minimum. The model, inputs, and test outputs seem correct, so we're not quite sure why the weights aren't improving. This also causes the contour lines to be far off from the target, as seen below. We also tried using the mean squared error and the L1 norm as cost functions, but it made no difference.

2.



3.

Number of Hidden Neurons	Final MSE	Number of Epochs
2	0.29790599	7
8	0.29790599	17
50	0.29790599	39

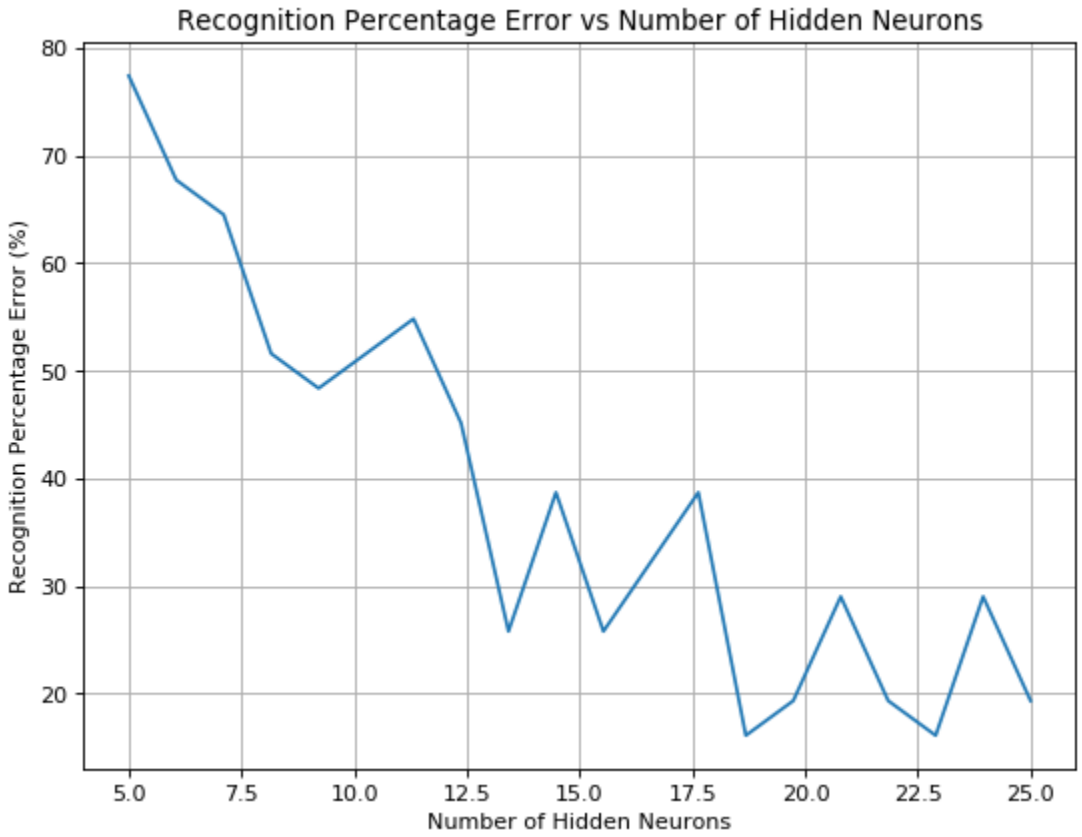
Brandon Death 101017130
Matt Walker 100980836
Oct. 28, 2018.

4. The hyperbolic tangent activation function was used, since our target function can have negative values and tanh squashes the values from -1 to 1, whereas the sigmoid function squashes them from 0 to 1.

b) & c) We ran into the same convergence issue that we encountered for part a), which doesn't allow us to obtain usable data to produce the tables and charts. Our error never reaches the 0.02 threshold, and we can't seem to adjust the weights enough to improve our output.

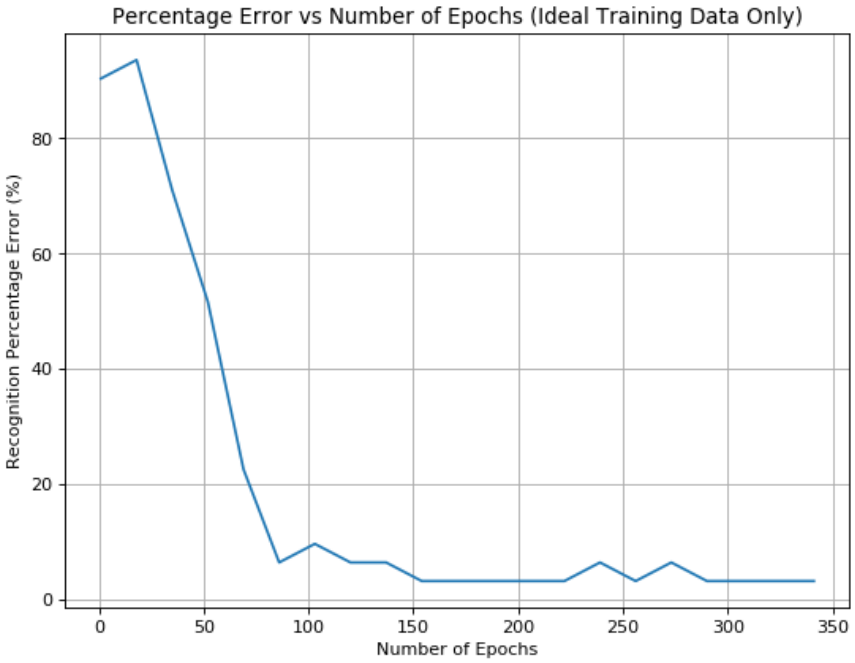
Question 2

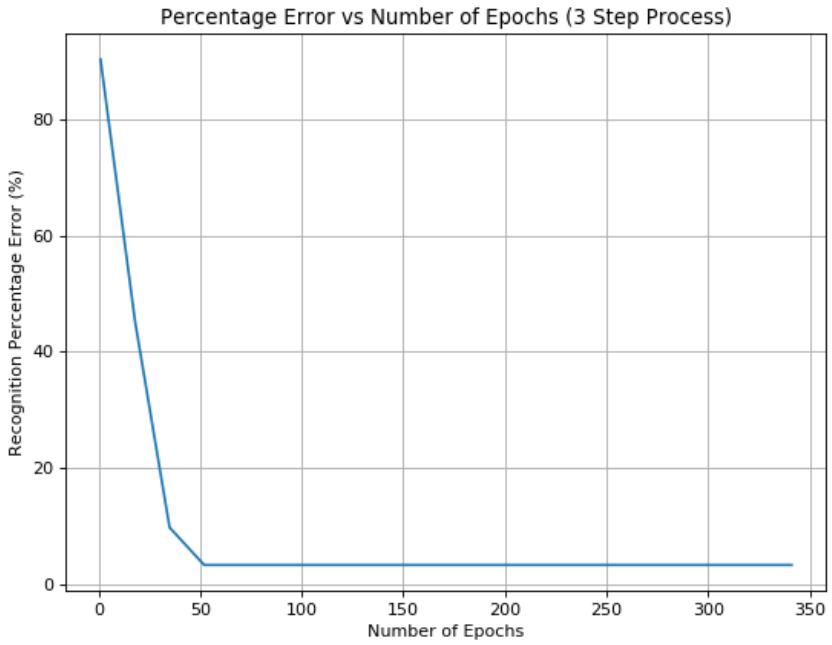
a) Plot used with noise level of 3. The model was trained with a learning rate of 0.05 for 250 epochs with batches of size 3:



Brandon Death 101017130
Matt Walker 100980836
Oct. 28, 2018.

b)

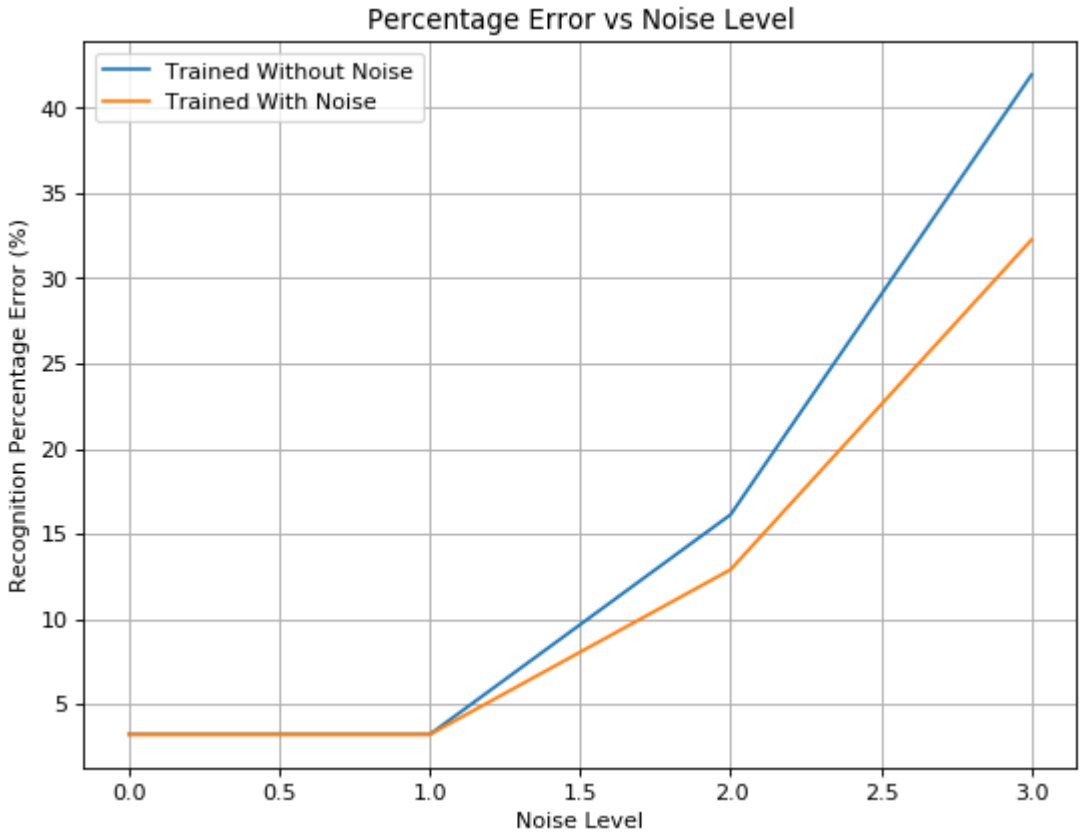




As seen from the graphs above, the 3 step method converges to a minimum error much quicker than the one step process trained on ideal input only. This isn't quite the same as shown in Figure 13 on the assignment because the error starts high, but it makes sense that it's high in the beginning because the weights need enough training to be adjusted.

Brandon Death 101017130
Matt Walker 100980836
Oct. 28, 2018.

c)



This graph confirms that the model trained with noise can outperform the model trained without noise, when tested with distorted data.