In this homework we simulated a neural network system that uses back propagation to perform a parity judgment based off of random input units. The system is trained on 3 layers: 8 input units, 3 hidden units and one output unit. The system was trained with two matrices uniformly randomized in the range of -0.5 to 0.5, created to signify the input to hidden layer and the hidden to output layer. To train the system, the input, input-to-hidden and hidden-to-output matrices were passed through a loop of less than 1000 epochs until the sum of square errors were less than 0.01. An epoch was defined as one sweep through all the input patterns since the input patterns were pulled in one column at a time to minimize errors. As each column was pulled in, from the input to hidden a sigmoid is activated and from hidden to output, a sigmoid gradient is calculated and then applied as the system sweeps backwards changing the weights. An example of the outputs is seen below.

input =

1 1 1 0 1 0 1 1

0 0 0 1 1 1 0 0

0 0 0 0 0 0 1 1

1 1 1 0 0 0 1 0

1 1 0 0 1 0 1 1

1 0 0 0 1 1 0 0

0 1 1 0 1 0 0 1

0 0 0 0 1 1 1 1

**Seen above is a randomized matrix of the matrix vector.**

check\_one = 31; There are odd number of ones!

**The check\_one variable checked to see if the number of ones in the matrix is odd or even.**

Results:

epoch:10 SSE:0.066257

epoch:20 SSE:0.03624

epoch:30 SSE:0.024044

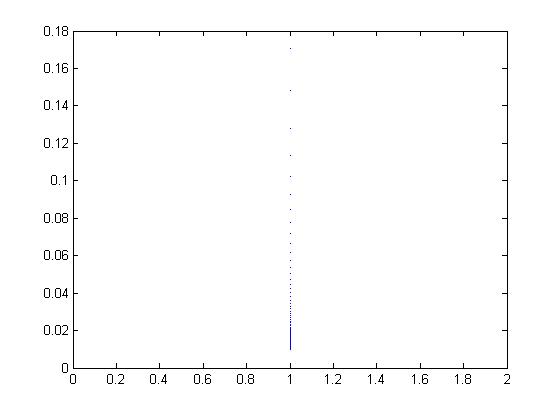
epoch:40 SSE:0.017706

epoch:50 SSE:0.013896

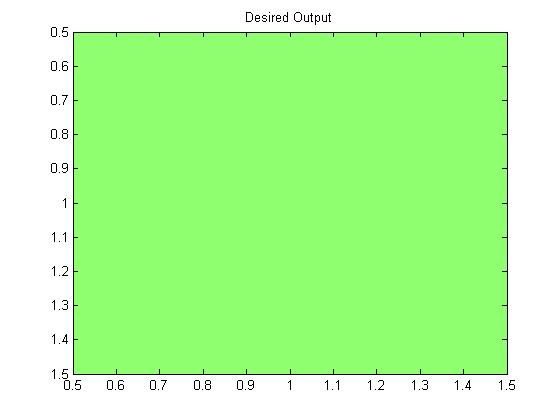
epoch:60 SSE:0.011379

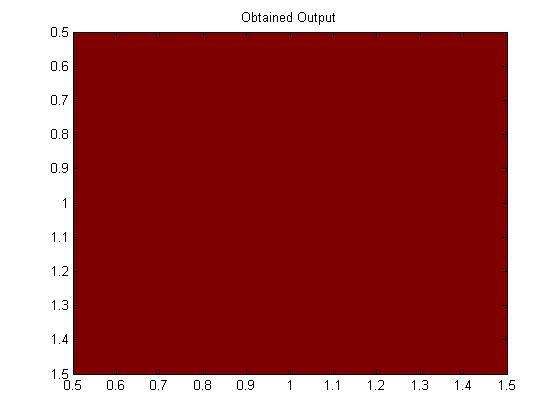
epoch:68 SSE:0.0099148

This was printed to see how many epochs were needed until the SSE was below 0.01. On average, it took 40-60 epochs.



**Seen above is a graph that demonstrates the decrease of the SSE as the epochs increased.**





After testing the system with a new set of patterns, you can distinguish how well the system learned. Without applying the weight changes, the new set of patterns had a SSE of 0.7286 on the 10th epoch iteration while the weights with the weight changes applied had an SSE of 0.066257. That’s nearly a 0.7 difference. The model does not generalize well to the new stimuli unless the weight changes are applied.