

Brennan McFarland

10/30/18

Problem Set 4

### Validating Computation

The program, run on beats.dat with 4 neurons at iteration 1000, yields the following output to indicate that the derivatives are correct relative to numerical estimates:

$F_{wji} =$

0	0	0	0
0	0	0	0
-0.0915	0.1026	0.0187	0.1591
0.0509	0.0615	-0.1691	0.0133

$\text{approx\_}F_{wji} =$

0	0	0	0
0	0	0	0
-0.0915	0.1026	0.0187	0.1591
0.0509	0.0615	-0.1691	0.0133

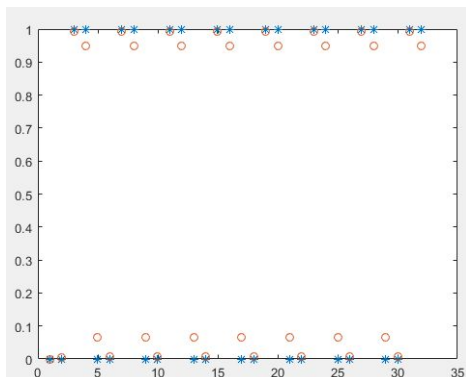
In all cases the derivatives are accurate to the numerical estimates to the number of decimal places shown.

### Minimum Interneuron Count

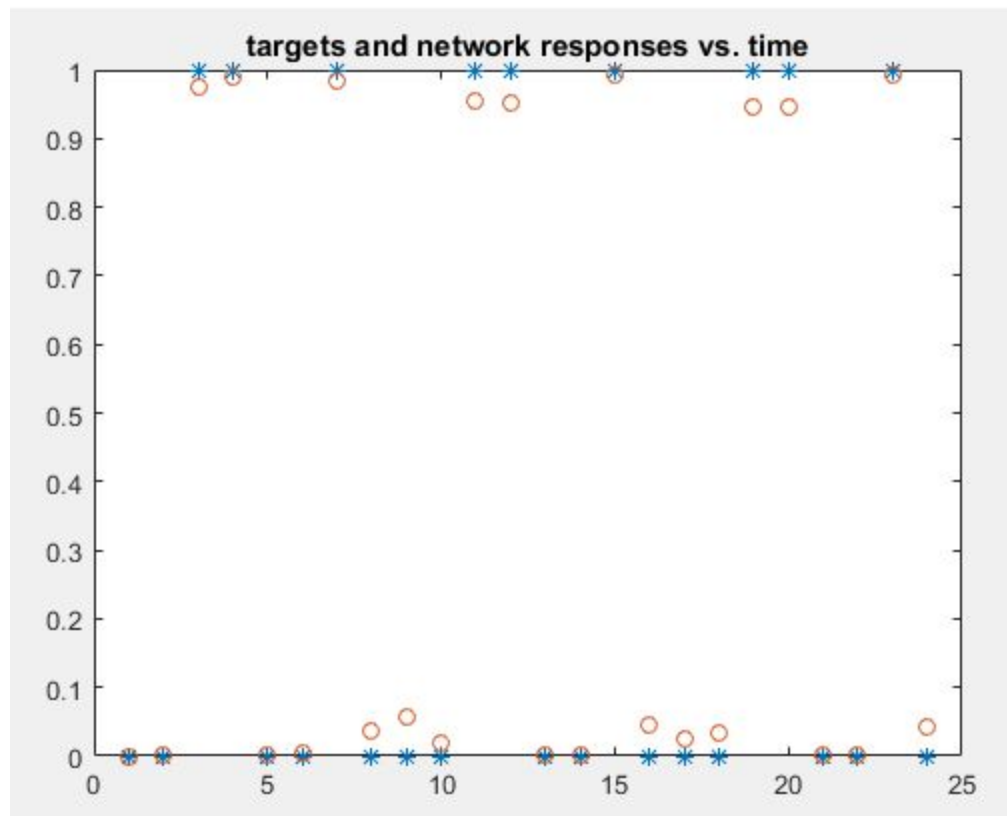
Beats.dat can be trained with only 1 interneuron with rms error under .03 in under 7000 iterations.

iiter = 7000

rms\_err = 0.0287



Beats2.dat, however, cannot. After much experimentation I found it difficult to get beats2 to converge in a reasonable time with only a single hidden layer, but was able to reach a state near convergence ( $\text{rms\_err}=.0210$ ) with 20 neurons.



### Learning Factor

In performing training, I adjusted the static learning factor based on convergence rate, and found a value of  $\text{ETA}=.1$  to work best to avoid oscillating while still reaching the solution in as close to optimal time as possible for beats1.dat. This led to convergence for beats.dat in approximately 7000 iterations. For beats2.dat, I tuned the learning factor to  $\text{ETA}=.01$  and got rough convergence in approximately 10000 iterations. This would indicate that the more complex character of the target function in beats2 resulted in a more complex decision boundary thus requiring a more finely-tuned learning rate to reach convergence. In the future, I would have liked to have tried an adaptive learning factor as it would be better suited to adjusting based on the topology of the hidden decision surface of different datasets (beats vs beats2), but I unfortunately did not have time to implement it.