WORKSHEET FOR EXERCISES FROM CHAPTER 8

EXERCISE 8.1

1. **What is the total SSE for the network after training has finished?**

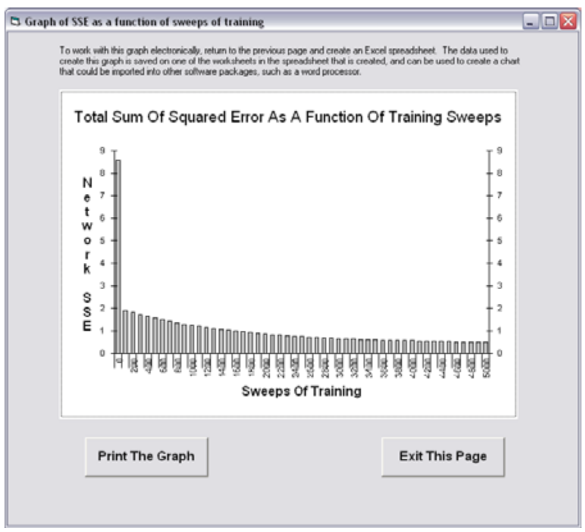
Total SSE was approximately 0.50 at the end of training.

2. **How many epochs of training occurred before the program stopped training the network**

The network was trained for a full 5000 epochs. As SSE was always decreasing (and had never reached 0), the program never really "stopped training" until the program itself was halted.

3. **Examine how SSE for this network changed over time. In general, what can be said about the performance of this network on this problem?**

There is a very large early decrease in error, followed by a much slower decrease in error. It would appear that this slower decrease is reaching an asymptote of about 0.50. This is shown in the graph below:



4. **Describe the kind of errors that the network made, if any. Is the network generating errors to a small number of problems, or are errors for all of the training patterns uniformly large? Relate the properties of any observed errors to what you know about the structure of the training set.**

The network is generating larger errors to some of the patterns. For instance, its performance is nearly perfect for all 8 outputs for patterns 3, 4, 5, 6, and 7. However it makes large errors for patterns 1, 2, and 8. This makes perfect sense given the structure of the training set, because when the linear dependence was built into the training set, it was done by creating vector 8 from vectors 1 and 2, and these vectors are all involved in this particular stimulus pairings.

5. **Rerun the network on the depend8.net problem, with a maximum number of sweeps set to 5000, training with the delta rule, and printing out information every 100 sweeps. Play with the learning rate a bit, and examine total SSE when the program stops training. Are you able to improve the performance of the network in any significant way? What are the implications of these observations? (To answer this question, you should provide some information about the settings that you used to run the study.)**

With a smaller learning rate (0.05), training for 5000 sweeps leads to a total SSE of 0.80. With a larger learning rate, training for the same amount of sweeps reduces error to 0.36. However, keeping this larger learning rate and training for 15,000 sweeps leads to a total SSE of 0.36. None of my simulations led to a total SSE of 0. This suggests that the linear dependence built into this training set prevents the network from perfectly learning the associations required of this training set.