WORKSHEET FOR EXERCISES FROM CHAPTER 19

EXERCISE 19.1

1. **How many sweeps of training were required before the network converged?**

The network converged after 1256 iterations.

2. **What was the sum of squared error (SSE) for the network when it converged?**

The total sum of squared error when it converged was approximately 0.04.

3. **Remember that the hidden units and the output units use the logistic activation function, and that the bias of this function is analogous to the threshold of the step function. In the spreadsheet that you saved, examine all of the biases and connection weights in the trained network. Use this information to explain how the network uses its hidden units to solve this linearly separable problem. (Hint: an example of this approach to the network can be found in Dawson (2004)).**

The first hidden unit has a bias of –3.66, and the two weights feeding into it from the inputs are equal to 8.33. As a result, this unit is basically an OR gate: it turns off when both inputs are off, but turns on when one or both of the inputs are on (i.e., off to the first pattern, on to the remaining three). The second hidden unit has a bias of –8.09, and the two weights feeding into it from the inputs are equal to 5.30. As a result, this unit is basically an AND gate: it only turns on when both inputs are on, and turns off otherwise. The output unit has a bias of –2.32; the weight to it from hidden unit 1 is 4.96, and the weight to it from hidden unit 2 is –5.23. As a result, it basically generates the same response as does hidden unit 1 to the first three patterns, because it is excited by hidden unit 1, and there is no signal from hidden unit 2. However, for the last pattern it (correctly) turns off because the excitatory signal from hidden unit 1 is counteracted by a strong inhibitory signal from hidden unit 2.

EXERCISE 19.2

1. **How many sweeps of training were required before the network converged?**

The network took 1565 sweeps to converge.

2. **What was the sum of squared error (SSE) for the network when it converged?**

The sum of squared error at convergence was 0.03.

3. **You probably found in Exercise 19.1 that when the two-hidden unit was trained, it was temperamental, and as a result you might have had to restart training on more than one occasion. Is this also the case for the one-hidden unit network, or was it better behaved? (To answer this question, you should conduct several different runs with the one-hidden unit network.)**

A quick examination of the behaviour of this network indicated that it learned quickly and reliably. It seemed to behave better than the previous network. However, I did not conduct a systematic study of this issue.

4. **Remember that the hidden units and the output units use the logistic activation function, and that the bias of this function is analogous to the threshold of the step function. In the spreadsheet that you saved, examine all of the biases and connection weights in the trained network. Use this information to explain how the network uses its hidden units to solve this linearly separable problem. How does this solution compare to the one that you discovered in the two-hidden unit network? (Hint: an example of this approach to the network can be found in Dawson (2004)).**

In this network, the hidden unit learned to be an OR gate: it turned on to any pattern that had either 1 or 2 inputs on, but did not respond to the pattern with 0 inputs on. The output unit had a bias of –2.57, and the two inputs directly connected to it had weights of –4.59. The hidden unit was connected to the output unit with a weight of 9.56. This circuitry worked as followed: with no inputs, the net input was not large enough to cause a response. With one input unit activated, there was a weak inhibitory signal from an input unit that could not cancel out a much stronger excitatory signal from the OR unit (the hidden unit). With both input units activated, the excitation from the hidden unit was cancelled out, and the network correctly failed to respond to the last pattern.