Worksheet For Exercises From Chapter 22

Record Your Data From The First Study In Table 22-1

Run Number	Learning Rate	Total Sweeps	Converged?	SSE At End	Misses At End
1	0.01	10,000	No	0.69	3
2	0.01	10,000	No	0.73	3
3	0.01	10,000	No	0.85	5
4	0.01	3506	Yes	0.03	0
5	0.01	10,000	No	0.89	6
6	0.001	6085	Yes	0.05	0
7	0.001	3920	Yes	0.06	0
8	0.001	10,000	No	0.76	7
9	0.001	30,114	Yes	0.02	0
10	0.001	14,060	Yes	0.02	0

Table 22-1. Record of results from the first simulation.

Exercise 22.1

1. **From your observations of the network during training at the first learning rate, in a few sentences summarize the characteristics of the dynamics of learning this problem at these settings. For example, did learning proceed in a regular fashion, or did the network "plateau" at certain levels of hits and misses?**

In general, the networks started at reasonably high error (SSE = 8), and this error quickly decreased to below 1. At this point, though, the network stalled – decreases in error were very slow, and the number of hits stopped rising. This was true for all but one of the networks. For this network, the initial decrease in error was more rapid than for the others, and instead of stalling near a solution, the network quickly converged.

2. **From your observations of the network during training at the second learning rate, in a few sentences summarize the characteristics of the dynamics of learning this problem at these settings.**

These networks were the reverse of the ones described above. Two of them converged rapidly. Two others did so too, but then stalled very close to a solution. Both of these converged eventually, but it took a long time to achieve this. The last network behaved like the majority in question 1 – moderately fast SSE decrease, followed by a stall well away from a solution.

3. **In a sentence or two, and on the basis of the two sets of simulations that you have done, what would you say the effect of reducing the learning rate was on this problem for this architecture?**

In terms of likelihood of converging, it is counterintuitive but it would appear that decreasing the learning rate actually improved performance of the value unit networks on this problem.

RECORD YOUR DATA FOR THE SECOND STUDY IN TABLE 22-2

Run Number	Learning Rate	Total Sweeps	Converged?	SSE At End	Misses At End
1	0.01	10000	No	0.41	12
2	0.01	10000	No	9.17	34
3	0.01	3190	Yes	0.05	0
4	0.01	7261	Yes	0.03	0
5	0.01	1773	Yes	0.06	0
6	0.001	10000	No	1.86	30
7	0.001	10000	No	0.14	2
8	0.001	10000	No	12.84	59
9	0.001	10000	No	3.82	26
10	0.001	10000	No	6.45	30

Table 22-2. Record of results from the second simulation.

EXERCISE 22.2

1. **From your observations of the network during training at the first learning rate, in a few sentences summarize the characteristics of the dynamics of learning this problem at these settings. For example, did learning proceed in a regular fashion, or did the network "plateau" at certain levels of hits and misses?**

This network was better behaved than the previous one at the larger learning rate. Several converged quickly to a solution. One reached a plateau far from a solution, the other was closer to the solution, neither converged.

2. **From your observations of the network during training at the second learning rate, in a few sentences summarize the characteristics of the dynamics of learning this problem at these settings.**

The dynamics of these networks were better than before, in the sense that error seemed to drop in a very encouraging way. However, overall performance was much worse – none of the networks converged!

3. **In a sentence or two, and on the basis of the two sets of simulations that you have done, what would you say the effect of reducing the learning rate was on this problem for this architecture? How does this compare to what you found with the 5-parity problem?**

The effect of decreasing the learning rate here definitely hurt performance. The lower learning rate did not lead to any converged networks.

Record Your Data For The Third Study In Table 22-2

Run Number	Learning Rate	Total Sweeps	Converged?	SSE At End	Misses At End
1	0.1	10000	No	0.85	6
2	0.1	10000	No	0.86	6
3	0.1	10000	No	3.1	13
4	0.1	4554		0.05	0
5	0.1	10000	No	8	32
6	0.01	10000	No	4.22	32
7	0.3	10000	No	1.44	6
8	0.3	10000	No	6.73	26
9	0.01	10000	No	8	32
10	0.01	10000	No	8	32

Table 22-3. Record of results from the third simulation.

EXERCISE 22.3

1. **From your observations of the network during training at the first learning rate, in a few sentences summarize the characteristics of the dynamics of learning this problem at these settings. For example, did learning proceed in a regular fashion, or did the network "plateau" at certain levels of hits and misses?**

Four of the five networks rapidly decreased to an error plateau, and then stayed there. Once the plateau was reached, error did not move at all, and the network did not converge. This was in spite of the fact that 5 hidden units were in use. One of the networks behaved in a fashion to the successful networks described earlier, and quickly converged to a solution.

2. **From your observations of the network during training at the second learning rate, in a few sentences summarize the characteristics of the dynamics of learning this problem at these settings.**

Reducing the error rate to 0.01, or increasing it to 0.03, did not lead to any successes.

3. **In a sentence or two, and on the basis of the two sets of simulations that you have done, what would you say the effect of reducing the learning rate was on this problem for this architecture?**

For this architecture, the effect of reducing the learning rate for 5-parity did not seem to be as positive as was the case for the value unit architecture.

4. **In general, from your experience with these simulations, what can you say in a couple of sentences about how the two different architectures (value units and integration devices) deal with this particular problem?**

The data above would suggest that the value unit architecture, with fewer hidden units, has more success with the 5-bit parity problem than does the integration device architecture.

5. ***Bonus question: Choose a learning rate of 0.1, and repeat the above exercise using the 7-bit parity problem, integration devices, and 7 hidden units. How well does a network of integration devices fare with this larger version of the same kind of problem? How does this compare to the value unit architecture?***

Run Number	Learning Rate	Total Sweeps	Converged?	SSE At End	Misses At End
1	0.1	10000	No	32.28	128
2	0.1	10000	No	32.16	128
3	0.1	10000	No	32.44	128
4	0.1	10000	No	32.01	128
5	0.1	10000	No	32.65	128

Table 22-4. Record of results from the bonus simulation.

The larger version of the problem poses larger problems for this network. It never comes close to converging. It is far poorer on this problem than were the networks of value units.