Worksheet For Exercises From Chapter 23

Record Your Data For The 27-Hidden Unit Condition In Table 23-1

Run	Maximum Weight	Minimum Weight	Weight Sign	Number of Hidden Units	Converged?	Sweeps To Converge
1	1	0	Both	27	Yes	1165
2	1	0	Both	27	Yes	1509
3	1	0	Both	27	Yes	1583
4	1	0	Both	27	Yes	3017
5	1	0	Both	27	Yes	1363
1	2	1	Both	27	Yes	27
2	2	1	Both	27	Yes	51
3	2	1	Both	27	Yes	98
4	2	1	Both	27	Yes	50
5	2	1	Both	27	Yes	56
1	3	2	Both	27	Yes	29
2	3	2	Both	27	Yes	3
3	3	2	Both	27	Yes	5
4	3	2	Both	27	Yes	81
5	3	2	Both	27	Yes	12

Table 23-1. Data record for the first simulation.

Record Your Data For The 18-Hidden Unit Condition In Table 23-2

Run	Maximum Weight	Minimum Weight	Weight Sign	Number of Hidden Units	Converged?	Sweeps To Converge
1	1	0	Both	18	Yes	2544
2	1	0	Both	18	Yes	1680
3	1	0	Both	18	Yes	2143
4	1	0	Both	18	Yes	2126
5	1	0	Both	18	Yes	2263
1	2	1	Both	18	No	5000
2	2	1	Both	18	Yes	51
3	2	1	Both	18	Yes	165
4	2	1	Both	18	Yes	117
5	2	1	Both	18	Yes	89
1	3	2	Both	18	Yes	21
2	3	2	Both	18	No	5000
3	3	2	Both	18	Yes	15
4	3	2	Both	18	Yes	56
5	3	2	Both	18	Yes	6

Table 23-2. Data record for the second simulation.

RECORD YOUR DATA FOR THE 9-HIDDEN UNIT CONDITION IN TABLE 23-3

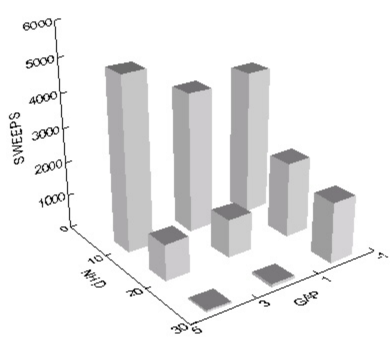
Run	Maximum Weight	Minimum Weight	Weight Sign	Number of Hidden Units	Converged?	Sweeps To Converge
1	1	0	Both	9	No	5000
2	1	0	Both	9	Yes	4948
3	1	0	Both	9	No	5000
4	1	0	Both	9	Yes	1101
5	1	0	Both	9	No	5000
1	2	1	Both	9	Yes	308
2	2	1	Both	9	No	5000
3	2	1	Both	9	No	5000
4	2	1	Both	9	No	5000
5	2	1	Both	9	No	5000
1	3	2	Both	9	No	5000
2	3	2	Both	9	No	5000
3	3	2	Both	9	No	5000
4	3	2	Both	9	No	5000
5	3	2	Both	9	No	5000

Table 23-3. Data record for the third simulation.

EXERCISE 23.1

1. **Examine the results that you have in the tables above. What are your general conclusions about how the manipulations of the starting states of the weights, and the number of hidden units, affect the ability to solve this problem?**

If one examines the data above, two general trends emerge. First, as the number of hidden units is increased, the speed of learning and the likelihood of converging are also increased. Furthermore, as one increases the size of the "gap" in the sampling distribution, the speed of learning and the likelihood of converging are also increased. The graph below plots the means of the different conditions taken from the three tables above:



The appearance of the graph suggests the possibility of an interaction. To test this, one can perform a regression predicting sweeps to learn (using 5000 sweeps as the result for noncoverged nets) from size of gap, number of hidden units, and the product of these two predictors (the interaction term). This is the result:

```
Dep Var: SWEEPS    N: 45    Multiple R: 0.760    Squared multiple R: 0.578

Adjusted squared multiple R: 0.547    Standard error of estimate: 1465.605
```

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	5063.689	913.965	0.000	.	5.540	0.000
GAP	452.667	353.977	0.343	0.143	1.279	0.208
NHID	-143.041	47.009	-0.488	0.400	-3.043	0.004
INTERACTION	-34.606	18.207	-0.562	0.118	-1.901	0.064

From the regression, it can be seen that there is a main effect of the number of hidden units, but there is not a main effect of the size of the gap. Instead, the interaction between gap size and the number of hidden units is significant (it would be if the P was taken as 1 tailed). In other words, the best learning occurs for large sized nets with a large gap size – consistent with the selectionist hypothesis.

2. **In the exercises above, structure was manipulated in a fairly coarse manner, by varying the distribution from which connection weights were randomly initialized. Can you think of other ways in which structure could be inserted into a network prior to the start of training?**

One way in which structure could be manipulated would be by imposing some sort of limited order effect, where hidden units only saw a certain proportion of inputs. The more systematic this manipulation, the better the manipulation of structure.