WORKSHEET FOR EXERCISES FROM CHAPTER 15

RECORD YOUR DATA IN THE TABLES BELOW

"Subject"	First Training File	Sweeps To Converge
1	9maj.net	2
2	9maj.net	3
3	9maj.net	3
4	9maj.net	4
5	9maj.net	7
6	9maj.net	3
7	9maj.net	4
8	9maj.net	7
9	9maj.net	3
10	9maj.net	5

"Subject"	First File	Sweeps To Converge	Second File	SSE Before Training	Sweeps To Converge	Total Sweeps
1	9maj25.net	6	9maj.net	6	2	8
2	9maj25.net	9	9maj.net	3	1	10
3	9maj25.net	7	9maj.net	1	1	8
4	9maj25.net	10	9maj.net	1	3	11
5	9maj25.net	7	9maj.net	0	0	7
6	9maj25.net	13	9maj.net	10	1	11
7	9maj25.net	9	9maj.net	3	5	14
8	9maj25.net	11	9maj.net	3	3	14
9	9maj25.net	11	9maj.net	0	0	11
10	9maj25.net	10	9maj.net	2	2	12

"Subject"	First File	Sweeps To Converge	Second File	SSE Before Training	Sweeps To Converge	Total Sweeps
1	9maj50.net	8	9maj.net	0	0	8
2	9maj50.net	7	9maj.net	0	0	7
3	9maj50.net	5	9maj.net	0	0	5
4	9maj50.net	4	9maj.net	0	0	4
5	9maj50.net	6	9maj.net	0	0	6
6	9maj50.net	5	9maj.net	0	0	5

7	9maj50.net	5	9maj.net	0	0	5
8	9maj50.net	7	9maj.net	0	0	7
9	9maj50.net	5	9maj.net	0	0	5
10	9maj50.net	4	9maj.net	0	0	4

"Subject"	First File	Sweeps To Converge	Second File	SSE Before Training	Sweeps To Converge	Total Sweeps
1	9maj75.net	3	9maj.net	0	0	3
2	9maj75.net	5	9maj.net	0	0	5
3	9maj75.net	7	9maj.net	0	0	7
4	9maj75.net	8	9maj.net	0	0	8
5	9maj75.net	5	9maj.net	0	0	5
6	9maj75.net	5	9maj.net	0	0	5
7	9maj75.net	4	9maj.net	0	0	4
8	9maj75.net	3	9maj.net	0	0	3
9	9maj75.net	4	9maj.net	0	0	4
10	9maj75.net	4	9maj.net	0	0	4

EXERCISE 15.1

1. **For each of the three experimental conditions, what is the average of the "SSE Before Training" column? In general, what does this tell us about the ability of a perceptron to generalize what it has learned to new instances of the 9-majority problem?**

For the training set consisting of 25% of the patterns, the average SSE is 2.9. For the other two conditions, the average SSE is 0. All in all, this suggests that the ability of the perceptron to generalize on the 9 majority problem is quite high. When trained on 50% of 75% of the patterns, the network is perfect on all of the patterns that it has not seen before. When trained on only 25%, the network is perfect on most of the patterns that it has not seen before. For instance, if you train a network on the entire pattern set, the SSE that you get with a standard starting state is in the order of 250 – a 2.9 SSE is two orders of magnitude smaller.

2. **Does the amount of experience that a perceptron has affect its ability to generalize? (To answer this question, you need to compare the three "SSE Before Training" columns. Appropriate statistical tests are the best option for this question. You can compute independent t-tests between each pair of columns, or you can compute a one-way ANOVA using all three columns as the values to be tested.)**

Yes. The more patterns that the network sees (i.e., the higher the percentage of patterns that it has been presented in training), the better is its generalization. But this only works to a point. Our data shows that after the network has seen 50% of the patterns, it doesn't need to see any more for perfect generalization.

3. **If you found that some of the networks in the three experimental conditions had errors before their second round of training, then answer this question.**

(Otherwise, answer it by simply saying N/A.) Examine the three "Total Sweeps" columns for the experimental conditions, and compare them to the "Sweeps To Converge" column obtained in the control condition. (This is best done with appropriate statistical tests, like a one-way ANOVA or a set of independent t-tests on all possible pairs of columns.) Is there any advantage to training a perceptron on some of the problems before training it on the complete set?

A t-test between the "Sweeps to converge" column from training the entire training set, compared to the "Sweeps to converge" column from training on 25% of the training set produces a t-value associated with a probability of $p < 0.003$. So, the pretraining on part of the training set sped up the learning of the rest of training set in a statistically significant fashion.