# Description

In this assignment a two-layered perceptron network was implemented and experiment on in two ways. One of these experiments was the adjustment of the amount of hidden units, either 20, 50, ot 100 units. The other experiment was adjustment of the amount of training data used, which was either one quarter or one half of the training. The training data will be roughly balanced when it comes to the mnist data set.
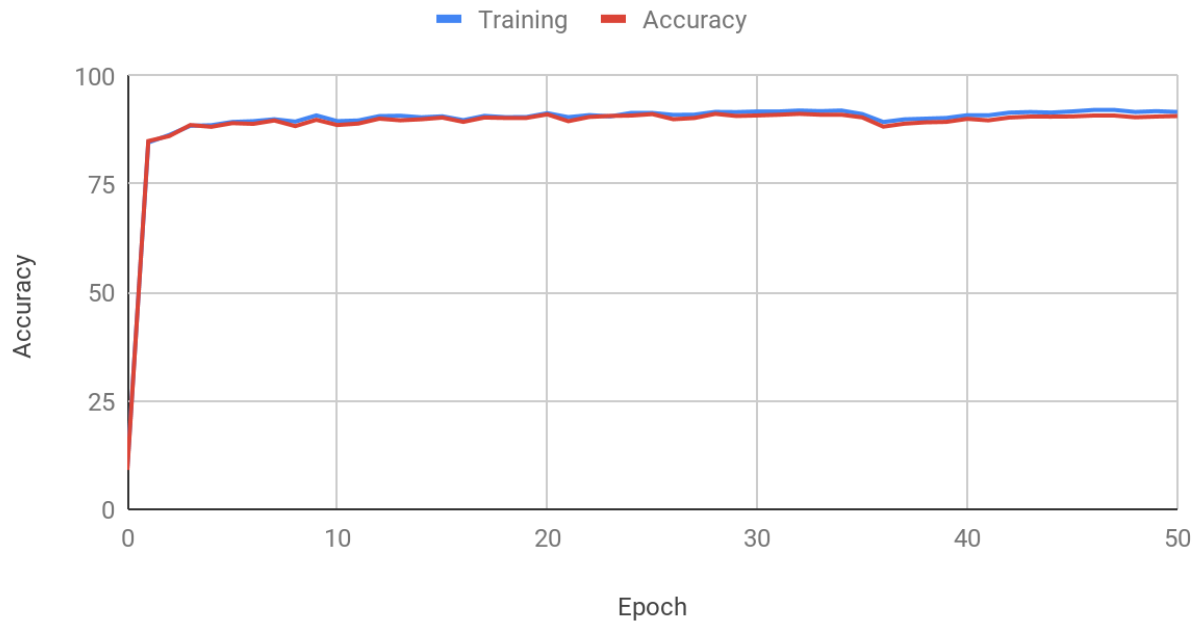
Experiment 1

20 Hidden Units

20 Hidden Units
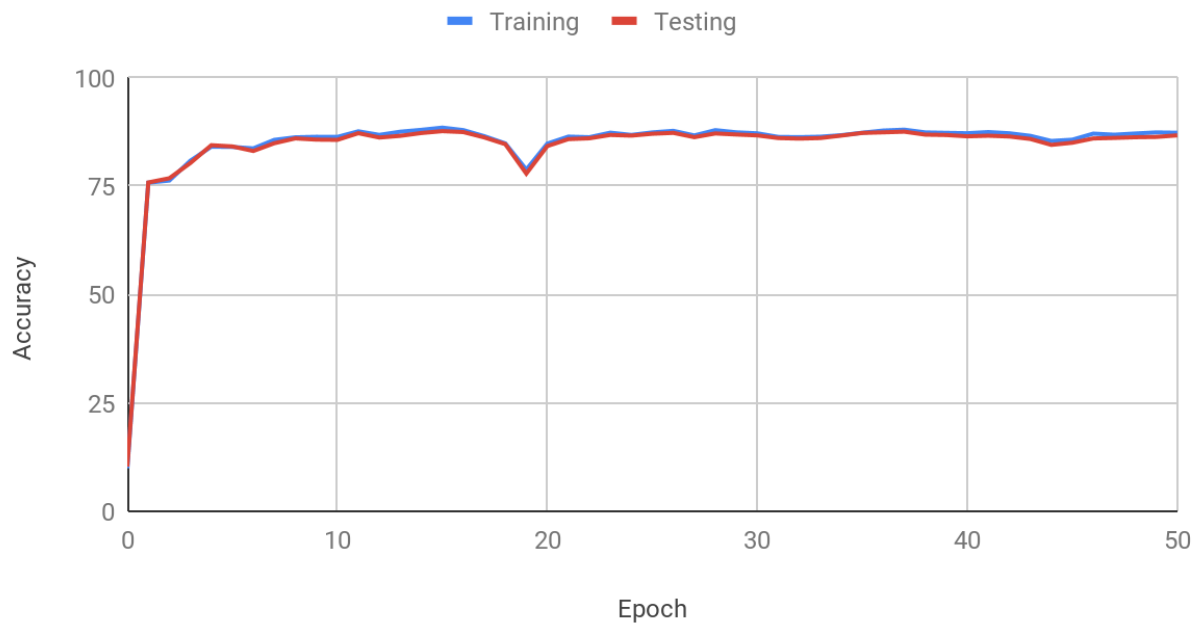
	0	1	2	3	4	5	6	7	8	9
0	938	1	2	8	1	4	6	1	10	9
1	1	1116	0	3	1	6	3	0	5	0
2	20	4	774	36	15	10	33	71	58	11
3	10	2	17	873	2	61	6	9	17	13
4	0	6	5	7	899	17	9	4	16	19
5	14	6	6	67	32	702	11	16	26	12
6	22	5	51	1	8	18	780	0	49	24
7	1	16	22	23	21	18	0	896	11	20
8	30	4	13	15	13	11	16	10	801	61
9	17	5	2	14	28	7	6	4	36	890

50 Hidden Units

50 Hidden Units

	0	1	2	3	4	5	6	7	8	9
0	930	0	3	12	1	2	10	1	20	1
1	0	1059	6	10	1	2	2	0	54	1
2	7	2	945	17	6	2	11	18	24	0
3	6	5	15	883	2	22	8	10	48	11
4	2	1	6	0	838	1	26	8	35	65
5	4	1	8	14	1	843	7	5	7	2
6	11	1	3	6	4	8	903	1	20	1
7	2	3	31	18	5	7	0	939	12	11
8	14	11	8	39	3	17	18	10	832	22
9	14	5	3	17	33	7	15	4	15	896

100 Hidden Units

100 Hidden Units

Training ■ **Testing** ■
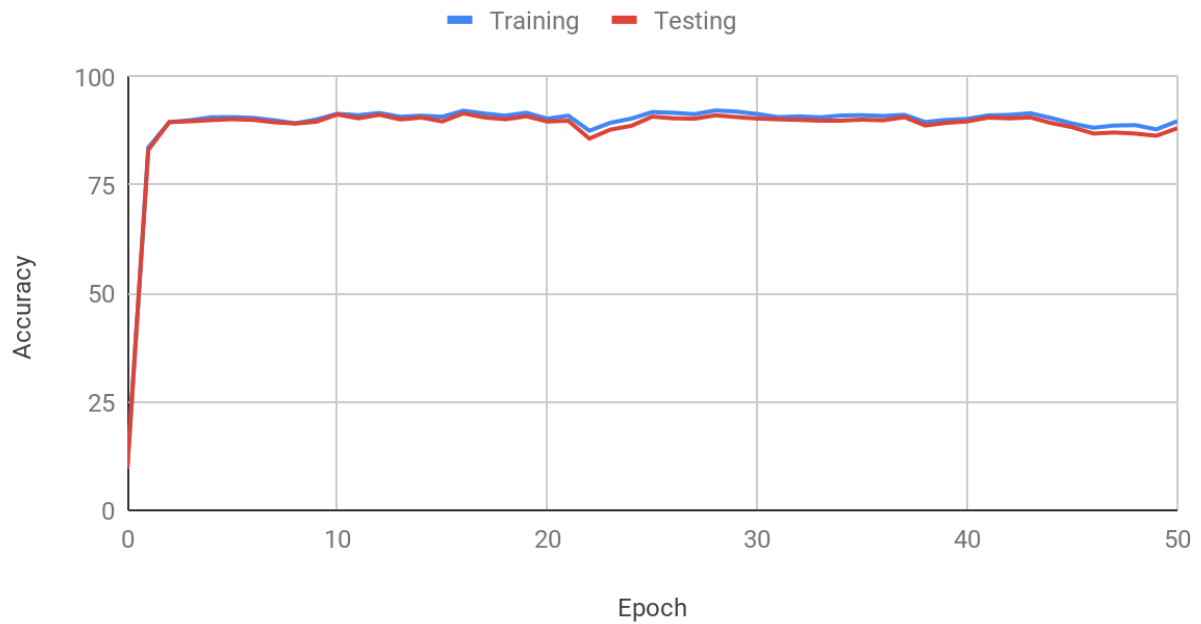
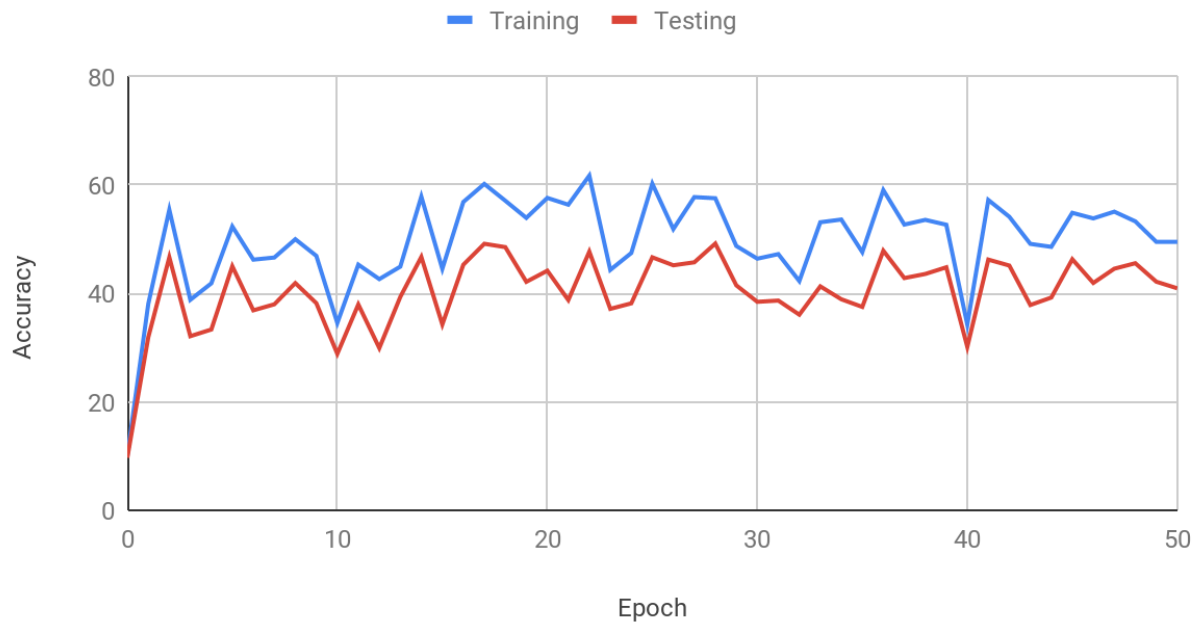|	0	1	2	3	4	5	6	7	8	9
0	930	2	3	3	6	3	16	6	10	1
1	0	1113	2	3	1	5	5	1	5	0
2	2	2	750	9	17	4	23	21	197	7
3	8	6	51	826	8	23	10	13	44	21
4	2	11	4	4	852	4	23	3	13	66
5	7	9	2	8	7	840	6	7	3	3
6	27	6	8	1	36	8	842	0	25	5
7	2	25	3	2	12	3	2	951	10	18
8	5	4	52	25	22	4	42	3	781	36
9	4	5	4	11	16	7	5	5	27	925

Discussion

What can be seen in the 20 hidden unit network, the 50 hidden unit network, and the 100 hidden unit network is at each epoch, the testing data and the training overlap almost as if they were the same plot. Overall, the more hidden units used, the higher the maximum accuracy rate the network reaches. This might be overfitting, but it might not be. I still somewhat do not understand overfitting, but overall, the training data and testing data fit quite well. Overall, it takes less than 5 epochs to reach 80%, but if there were more epochs, itis possible 50 hidden units will reach 100% accuracy quicker than 20% and possibly 100 hidden units, since the 100 hidden unit network does decrease somewhat gradually in the last ten epochs, so it might take longer for for the accuracy to increase for the 100 hidden unit. Overall, this was better than the first assignment, since the step activation has limits and one layer can cause some problems.

Experiment 2

Quarter Training Data

Quarter Training Data

	0	1	2	3	4	5	6	7	8	9
0	828	0	9	0	0	128	0	15	0	0
1	0	1017	61	0	0	57	0	0	0	0
2	33	6	694	0	0	296	0	3	0	0
3	10	1	141	0	0	827	0	31	0	0
4	10	3	5	0	300	649	0	15	0	0
5	14	2	35	0	1	814	0	26	0	0
6	101	1	64	0	1	787	1	3	0	0
7	18	21	187	0	2	359	0	441	0	0
8	0	2	129	0	1	839	0	3	0	0
9	10	4	8	0	42	908	0	37	0	0

Half Training Data

Half Training Data

	0	1	2	3	4	5	6	7	8	9
0	22	915	0	0	7	2	28	6	0	0
1	0	1129	0	0	0	0	6	0	0	0
2	0	770	0	0	61	4	175	22	0	0
3	3	916	0	0	15	0	57	19	0	0
4	0	32	0	0	910	0	35	5	0	0
5	7	622	0	0	49	104	91	19	0	0
6	0	24	0	0	16	1	916	1	0	0
7	0	72	0	0	17	0	1	938	0	0
8	0	827	0	0	71	0	64	12	0	0
9	8	715	0	0	218	2	0	66	0	0

Discussion

For the quarter data and the half data, it seems to oscillate and does not converge and is not as accurate, since the stochastic gradient descent seems to require a lot more data to work properly. Looking at the plots, it requires more training than usual and it seems to not work on some of the values, which brings up some questions, since it also brings massive confusion on those values. It is not recommended to reduce the training data, since it causes confusion within the network itself, due to the data collected.