Megan Sin
CS 539
HW 4 – Part 2

30-epoch, 30 hidden-unit, η=3.0 neural network results:

Results 1

Epoch	Correctly Predicted
0	9022 / 10000
1	9208 / 10000
2	9267 / 10000
3	9312 / 10000
4	9361 / 10000
5	9381 / 10000
6	9370 / 10000
7	9393 / 10000
8	9428 / 10000
9	9417 / 10000
10	9468 / 10000
11	9402 / 10000
12	9438 / 10000
13	9426 / 10000
14	9446 / 10000
15	9433 / 10000
16	9469 / 10000
17	9465 / 10000
18	9447 / 10000
19	9453 / 10000
20	9471 / 10000
21	9469 / 10000
22	9483 / 10000
23	9453 / 10000
24	9467 / 10000
25	9461 / 10000
26	9453 / 10000
27	9464 / 10000
28	9483 / 10000
29	**9483 / 10000**

Results 2

Epoch	Correctly Predicted
0	9077 / 10000
1	9235 / 10000
2	9256 / 10000
3	9321 / 10000
4	9353 / 10000
5	9387 / 10000
6	9365 / 10000
7	9440 / 10000
8	9449 / 10000
9	9450 / 10000
10	9440 / 10000
11	9452 / 10000
12	9427 / 10000
13	9479 / 10000
14	9480 / 10000
15	9504 / 10000
16	9493 / 10000
17	9503 / 10000
18	9512 / 10000
19	9499 / 10000
20	9519 / 10000
21	9494 / 10000
22	9525 / 10000
23	9496 / 10000
24	9506 / 10000
25	9489 / 10000
26	9490 / 10000
27	9523 / 10000
28	9495 / 10000
29	**9531 / 10000**

Results 3

Epoch	Correctly Predicted
0	9072 / 10000
1	9228 / 10000
2	9272 / 10000
3	9298 / 10000
4	9363 / 10000
5	9397 / 10000
6	9389 / 10000
7	9395 / 10000
8	9415 / 10000
9	9425 / 10000
10	9447 / 10000
11	9432 / 10000
12	9435 / 10000
13	9432 / 10000
14	9475 / 10000
15	9451 / 10000
16	9467 / 10000
17	9454 / 10000
18	9446 / 10000
19	9452 / 10000
20	9449 / 10000
21	9464 / 10000
22	9458 / 10000
23	9458 / 10000
24	9475 / 10000
25	9489 / 10000
26	9473 / 10000
27	9459 / 10000
28	9477 / 10000
29	**9481 / 10000**

Median result: 9483 / 10000

Observation: with additional epochs, the number of correct predictions also generally increases

Epochs experimentation

20-epoch, 30 hidden-unit, η=3.0 neural network:

Epoch	Correctly Predicted
0	9006 / 10000
1	9207 / 10000
2	9310 / 10000
3	9353 / 10000
4	9370 / 10000
5	9379 / 10000
6	9412 / 10000
7	9429 / 10000
8	9422 / 10000
9	9413 / 10000
10	9460 / 10000
11	9457 / 10000
12	9495 / 10000
13	9475 / 10000
14	9460 / 10000
15	9483 / 10000
16	9474 / 10000
17	9448 / 10000
18	9459 / 10000
19	**9490 / 10000**

60-epoch, 30 hidden-unit, η=3.0 neural network:

Epoch	Correctly Predicted
0	9027 / 10000
1	9178 / 10000
2	9266 / 10000
3	9354 / 10000
4	9357 / 10000
5	9391 / 10000
6	9416 / 10000
7	9406 / 10000
8	9456 / 10000
9	9413 / 10000
10	9435 / 10000
11	9439 / 10000
12	9462 / 10000
13	9472 / 10000
14	9478 / 10000
15	9490 / 10000
16	9470 / 10000
17	9491 / 10000

18	9468 / 10000
19	9462 / 10000
20	9494 / 10000
21	9433 / 10000
22	9453 / 10000
23	9474 / 10000
24	9496 / 10000
25	9494 / 10000
26	9482 / 10000
27	9486 / 10000
28	9485 / 10000
29	9480 / 10000
30	9467 / 10000
31	9473 / 10000
32	9501 / 10000
33	9500 / 10000
34	9487 / 10000
35	9504 / 10000
36	9458 / 10000
37	9460 / 10000
38	9495 / 10000
39	9482 / 10000
40	9475 / 10000
41	9498 / 10000
42	9473 / 10000
43	9493 / 10000
44	9483 / 10000
45	9479 / 10000
46	9478 / 10000
47	9469 / 10000
48	9483 / 10000
49	9462 / 10000
50	9485 / 10000
51	9454 / 10000
52	9486 / 10000
53	9489 / 10000
54	9499 / 10000
55	9487 / 10000
56	9478 / 10000
57	9471 / 10000
58	9466 / 10000
59	**9475 / 10000**

Observation: with less epochs, the (median) results were comparable to 30 epochs. However, with more epochs, the results did not scale and having more epochs did not improve performance.

Hidden-unit experimentation

(30-epoch, 40 hidden-unit, η=3.0 neural network):

Epoch	Correctly Predicted
0	9143 / 10000
1	9253 / 10000
2	9333 / 10000
3	9370 / 10000
4	9378 / 10000
5	9468 / 10000
6	9439 / 10000
7	9461 / 10000
8	9501 / 10000
9	9483 / 10000
10	9510 / 10000
11	9498 / 10000
12	9538 / 10000
13	9483 / 10000
14	9529 / 10000
15	9543 / 10000
16	9516 / 10000
17	9539 / 10000
18	9539 / 10000
19	9535 / 10000
20	9530 / 10000
21	9539 / 10000
22	9524 / 10000
23	9544 / 10000
24	9558 / 10000
25	9536 / 10000
26	9528 / 10000
27	9550 / 10000
28	9533 / 10000
29	**9549 / 10000**

(30-epoch, 10 hidden-unit, η=3.0 neural network):

Epoch	Correctly Predicted
0	8503 / 10000
1	8753 / 10000
2	8865 / 10000
3	8904 / 10000
4	8906 / 10000
5	8909 / 10000
6	8873 / 10000
7	8956 / 10000

8	8935 / 10000
9	8944 / 10000
10	9008 / 10000
11	9048 / 10000
12	9078 / 10000
13	9080 / 10000
14	9096 / 10000
15	9078 / 10000
16	9106 / 10000
17	9072 / 10000
18	9080 / 10000
19	9067 / 10000
20	9128 / 10000
21	9105 / 10000
22	9104 / 10000
23	9129 / 10000
24	9120 / 10000
25	9033 / 10000
26	9110 / 10000
27	9149 / 10000
28	9107 / 10000
29	**9093 / 10000**

Observation: the initial prediction with more hidden units was sometimes worse than before (30 hidden-units). However, with the same number of epochs as before (30), the prediction accuracy by the end was generally better than before with more hidden layers. The results from the neural network with fewer hidden units did not perform as well.

Learning rate (η) experiment

30-epoch, 30 hidden-unit, η=5.0 neural network:

Epoch	Correctly Predicted
0	9048 / 10000
1	9259 / 10000
2	9346 / 10000
3	9350 / 10000
4	9394 / 10000
5	9413 / 10000
6	9440 / 10000
7	9429 / 10000
8	9446 / 10000
9	9449 / 10000
10	9458 / 10000
11	9465 / 10000
12	9426 / 10000
13	9439 / 10000
14	9469 / 10000
15	9478 / 10000
16	9465 / 10000
17	9445 / 10000
18	9464 / 10000
19	9485 / 10000
20	9482 / 10000
21	9454 / 10000
22	9473 / 10000
23	9501 / 10000
24	9457 / 10000
25	9517 / 10000
26	9511 / 10000
27	9509 / 10000
28	9484 / 10000
29	**9472 / 10000**

30-epoch, 30 hidden-unit, η=1.0 neural network:

Epoch	Correctly Predicted
0	7686 / 10000
1	9050 / 10000
2	9175 / 10000
3	9203 / 10000
4	9261 / 10000
5	9268 / 10000
6	9307 / 10000
7	9315 / 10000

8	9325 / 10000
9	9343 / 10000
10	9353 / 10000
11	9349 / 10000
12	9368 / 10000
13	9380 / 10000
14	9393 / 10000
15	9386 / 10000
16	9381 / 10000
17	9381 / 10000
18	9377 / 10000
19	9397 / 10000
20	9411 / 10000
21	9412 / 10000
22	9409 / 10000
23	9419 / 10000
24	9431 / 10000
25	9409 / 10000
26	9422 / 10000
27	9422 / 10000
28	9431 / 10000
29	**9426 / 10000**

Observation: both increasing and decreasing the learning rate made the performance worse. A good median learning rate η, like the original $\eta=3.0$, does the best job.

Through my experiment of changing the number of epochs, number of hidden units, and the learning rate, I think the best configuration is a relatively medium number of epochs, slightly larger number of hidden units, and a medium learning rate.

An example of this would be the following: 20-epoch, 40 hidden-unit, $\eta=3.0$ neural network.

One possible way to improve the prediction rate is to increase/experiment with the number of hidden layers and not just the hidden units. Another possibility would be to preprocess the data to help reduce any amount of noise.