

Problem 4

Method

We design a Neural Network to classify digits in the MNIST dataset. The MNIST data was separated into a training set of 60,000 examples, and a test set of 10,000 examples. All examples in the dataset were normalized by 255 in order to get the range of each pixel to be from $[0,1]$.

A neural network was created with a single hidden layer of 10 units. Training was done via backpropagation and gradient descent with a learning rate set to $1e-4$ for 1000 iterations. The activation function employed was the logistic function; all relevant equations for calculating the gradient using the logistic function was used (see Choice1 in HW). MSE for the training data and test data were calculated over all iterations, and the classification error was tracked for the test set.

Out[1]: [Toggle Code](#)

Input: (60000, 785)
Layer0: (785, 10)
Layer1: (10, 10)
Finished Iteration 99, MSE: [0.09803127 0.09030702 0.09834118 0.101062
78 0.0963285 0.08970523
0.0977249 0.10179118 0.096725 0.09803475]
Finished Iteration 199, MSE: [0.09382804 0.01536743 0.09121165 0.09187
902 0.07882009 0.08439562
0.09117999 0.06978791 0.08709561 0.07421357]
Finished Iteration 299, MSE: [0.03200424 0.00999154 0.08321525 0.05783
366 0.04539173 0.07996905
0.06277681 0.02416529 0.0550979 0.05557294]
Finished Iteration 399, MSE: [0.02053916 0.00864589 0.04003289 0.03568
695 0.02370266 0.05293158
0.02768356 0.02432181 0.04266816 0.0343653]
Finished Iteration 499, MSE: [0.01758991 0.00804451 0.02906205 0.03041
215 0.01891216 0.03697163
0.02192865 0.02109149 0.03754186 0.0290419]
Finished Iteration 599, MSE: [0.01466062 0.00763976 0.02510377 0.02700
852 0.01731446 0.03150126
0.01848667 0.01929073 0.03122176 0.02453943]
Finished Iteration 699, MSE: [0.013037 0.00740968 0.02253011 0.02564
751 0.01639801 0.02890445
0.01683564 0.01769058 0.02894271 0.02242815]
Finished Iteration 799, MSE: [0.0120158 0.00724884 0.02094624 0.02477
538 0.01580947 0.02722185
0.01553693 0.01619762 0.02740177 0.02089529]
Finished Iteration 899, MSE: [0.0113379 0.0070707 0.0198793 0.02411
016 0.01536139 0.02602693
0.01448066 0.01494995 0.02626833 0.01983137]
Finished Iteration 999, MSE: [0.01084612 0.006888 0.01913143 0.02355
544 0.01498309 0.02514366
0.01363058 0.01393602 0.02540667 0.01907611]

Final % Classification Error (Test Set):

	0	1	2	3	4	5
6 \						
% Error	0.045918	0.02467	0.102713	0.138614	0.074338	0.197309
						0.0762
	7	8	9			
% Error	0.10214	0.175565	0.113974			

Final Overall % Classification Error (Test Set): 0.10

Discussion

The network reached an overall classification error of 0.1 (90% accuracy). Most of this error seems to be driven by the digits 8 and 5 (which may share similarities visually?). What is interesting is that the network picked out the digit 1 first for classification, and this ended up being the digit with the smallest classification error (0.02 error, 98% accuracy).

At around iteration 500, the network seemed to hit a variation in the error for about 25 iterations. This may be some sort of local minima that it was bouncing around.