Alessio Mazzone

CS1571 Artificial Intelligence

Homework 4

1. Use the ​Gaussian​​ data (the bottom left data set where the orange and blue blobs are nicely separated), the first two features (i.e., just X​1​ and X​2)​ , 1 hidden layer with 1 neuron. Roughly how many epochs does it take to reach a ​test​ ​loss​ of 0?

I ran many trials to get an estimate of how many epochs it takes for the test loss to reach 0.000. On my first run, it took about 350 epochs. On a second run, it took 438 to reach 0.000. On a third run still it took about 280 epochs. Upon running multiple other trials, I found the range of reaching 0.000 was between 280 and 450 epochs. So, taking an average, the number of epochs it takes to reach a test loss of 0.000 is roughly 350 according to the trials I ran.

2. Change to the ​exclusive or​​ data (top right). After 2,000 epochs, what is the ​test loss​? Is it correctly classifying most of the data? Does it look like it is getting much better?

After switching datasets, I let the network run for 2,000 epochs. I ran multiple trials and found that after 2,000 epochs, the test loss was somewhere between 0.370 and 0.425.

During the many trials I ran, I saw that the network is classifying most of the data correctly. By “most”, I mean here the majority, or at least 50% of the data. So, by this definition, the network is correctly identifying most, or at least 50%, of the data. I found that throughout multiple trials, the network would correctly classify almost the entirety of one label (almost all positive samples or all negative samples, depending on the trial) but then misclassify about half of the other samples correctly and half incorrectly. This means that that, for example, almost all the positive samples would be correctly classified, but then only half of the negative samples would be correctly classified, and the other half incorrectly classified.

It only takes about 200 epochs for the prediction of the network and the test loss to stabilize, and after that it only improves only slightly as it approaches 2,000 epochs. Thus, the test loss and predictions don’t get much better after about only 200 epochs.

3. What is the minimum number of hidden neurons required to quickly (e.g., epochs < 1,000) get nearly all the data right (e.g., ​test loss​ < 0.04) at least two-thirds of the time?

I used the same dataset as #2. I found that with 3 hidden neurons, the test loss after roughly 1,000 epochs was always somewhere between 0.05 and 0.17. The network here with 3 hidden neurons categorized nearly all the data correctly. But, when I added a fourth hidden neuron, I found that the network much more quickly categorized all the data correctly *and* the test loss was almost always below 0.04. So, the minimum number of hidden neurons required to quickly get nearly all the data categorized correctly most of the time is 4.

4. Switch to the circle data. Does your current network work well on that data? How many epochs does it take to get a test loss that is less than 0.01?

Using the circle data, I used one hidden layer with four neurons. Over many trials with this configuration, the range to get the test loss below 0.01 was wider than I expected. For some trials, the test loss took as little as about 460 epochs to fall below 0.01. On other trials, it took as many as about 1,000 epochs for the test loss to fall below 0.01. This is a larger range than I expected, but according to question number 3, this is still relatively quick since it takes less than 1,000 epochs usually.