CSE 5526 Lab 1 report

In this lab, I implemented a 2-layer neural network from scratch for parity problem.

In the first part, I will present my findings using tanh activation in both layers.

Here is the number of times under different learning rate for a model to reach a stopping criterion without using momentum.

Chart, line chart

Description automatically generated

As you can see from the figure above, as we increase the learning rate, it takes fewer epochs to converge in the general trend.

However, after we include the momentum in weight update, we could see a significant reduction of epochs to reach the convergence.

Chart, line chart

Description automatically generated

In the second part, I replace the tanh activation with sigmoid activation. And unfortunately, the model did not converge to an error less than 0.1. Here is the plot of the loss with one of the learning rates as an example.

Chart

Description automatically generated

Then I updated the model with momentum in weight update. The situation is the same: we have a faster start but still cannot get a small error. here is the plot of the error.

Chart, histogram

Description automatically generated

Some other observations: random start of the weight and bias will have a larger impact to the later training. It is important to figure out a good initialization.