Assignment2

September 2025

1

Define

$$Z^{[l]} = W^{[l]}a^{[l-1]} + b^{[l]} \tag{1}$$

,and

$$a^{[l]} = \sigma(W^{[l]}a^{[l-1]} + b^{[l]}) \tag{2}$$

for l=2,3,...,L.
For l=1, define $Z^{[1]} = x$ and obviously $\frac{\partial Z^{[1]}}{\partial x} = I$.
Notice that

$$\nabla a^{[L]}(x) = \frac{\partial a^{[L]}}{\partial x} = \frac{\partial a^{[L]}}{\partial Z^{[L]}} \frac{\partial Z^{[L]}}{\partial Z^{[L-1]}} ... \frac{\partial Z^{[1]}}{\partial x} \tag{3}$$

So we compute $\frac{\partial a^{[L]}}{\partial Z^{[L]}}$ and $\frac{\partial Z^{[l]}}{\partial Z^{[l-1]}}$ for l=2,3,...,L then we have,
$\frac{\partial a^{[L]}}{\partial Z^{[L]}} = \frac{\partial}{\partial Z^{[L]}}\sigma(Z^{[L]}) = \sigma^{'}(Z^{[L]})$(which is a scalar since $n_L = 1$),
$\frac{\partial Z^{[l]}}{\partial Z^{[l-1]}} = \frac{\partial Z^{[l]}}{\partial a^{[l-1]}} \frac{\partial a^{[l-1]}}{\partial Z^{[l-1]}} = W^{[l]}diag(\sigma^{'}(Z^{[l-1]}))$.
Hence,

$$\nabla a^{[L]}(x) = \frac{\partial a^{[L]}}{\partial Z^{[L]}} \frac{\partial Z^{[L]}}{\partial Z^{[L-1]}} ... \frac{\partial Z^{[1]}}{\partial x} = \sigma^{'}(Z^{[L]}) \prod_{l=2}^{L} W^{[l]}diag(\sigma^{'}(Z^{[l-1]})) \tag{4}$$

2

What is the good way to determine the learning rate of the optimizer?

3 Report

First, define the Runge function f, then generate the training set by randomly choosing points in [-1,1] which is denoted by x_{train}.

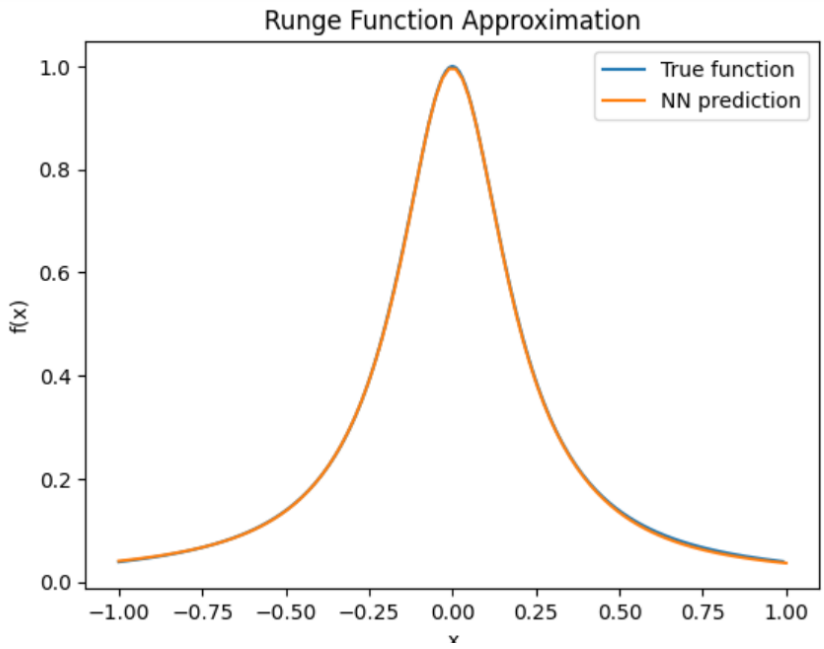
Then I choose 100 points in [-1,1] which are equally spaced and denoted by x_{test}.

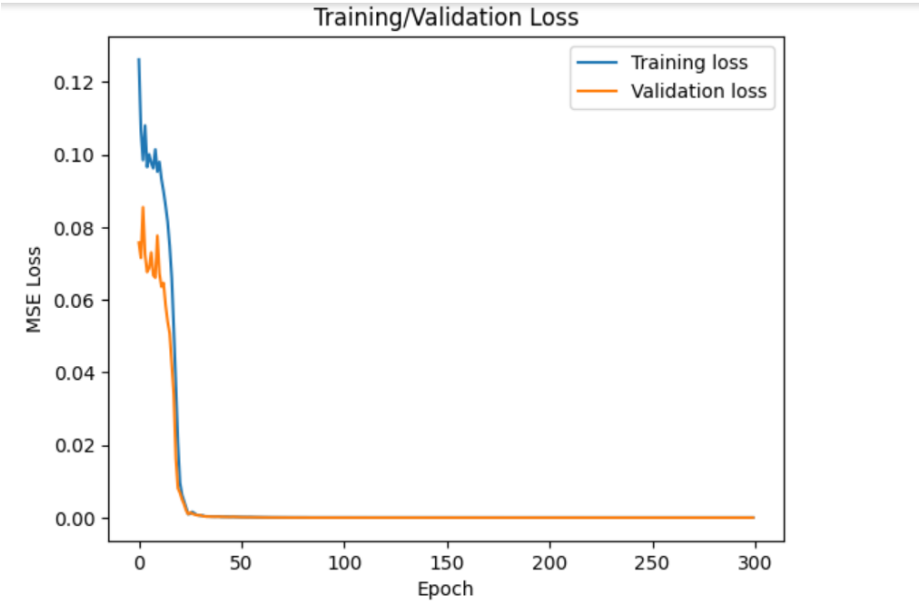
Then define $y_{train} = f(x_{train})$ and $y_{test} = f(x_{test})$.

Then I design a neural network with 1-neuron in the input layer, 20-neuron with the tanh activation function in the first hidden layer, 10-neuron with the sigmoid activation fuction and repeat this pattern for next two hidden layers, then the output layer has 1-neuron.

Then to train the neural network, it will do a forward and back propagation for 300 epochs while using the mini-batch gradient descent for 39 data each time and MSE loss and Adaptive moment estimation as optimizer.
And the validation set is constructed by choosing 30 percent data in the training set.

Next, we have the following result: MSE error: 5.985292021596456e-06





Note that the results may be different since the training set is constructed randomly. And according to my testing, the results are similar, they won't be too different from the result I show above.