

## 1. Question

Use a neural network to approximate the Runge function:

$$f(x) = \frac{1}{1+25x^2}, x \in [-1, 1]$$

## 2. Setup

1000 samples, 80% for training , 20% for validation. With a learning rate of 0.01 and Adam optimizer. The total epoch is 300.

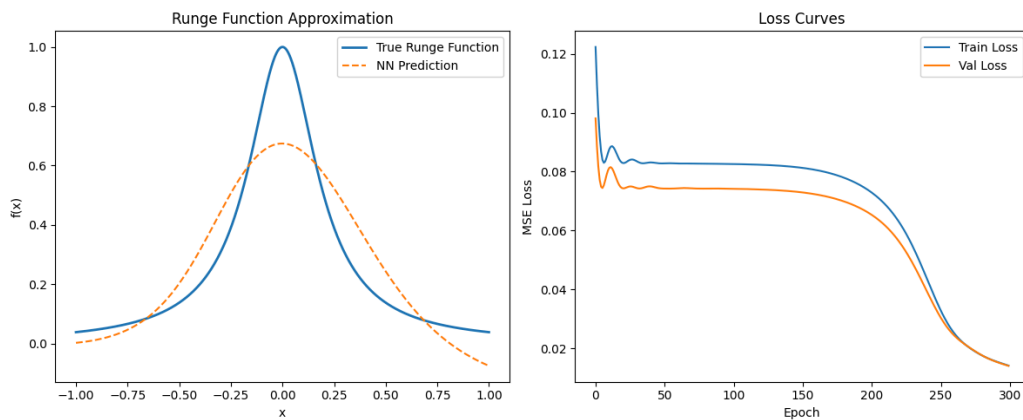
## 3. Method:

First, I use simple network of (1, 6, 6,1) in order to see the in this condition, what will the performance of different activation functions, include:

- (1) Sigmoid
- (2) Tanh
- (3) Relu
- (4) Leaky

The result is down here:

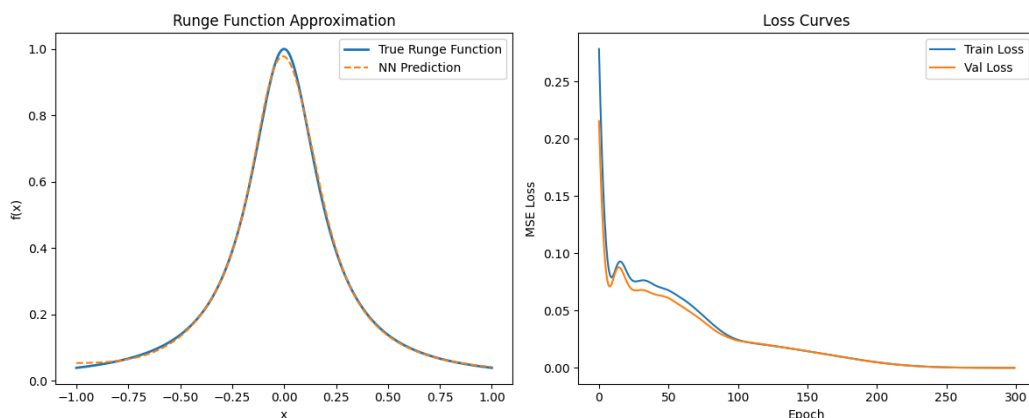
### (1) Sigmoid:



MSE: 0.014015

I guess there is a gradient vanish problem. So I change the learning rate to 0.03 and the MSE came down to 0.000333

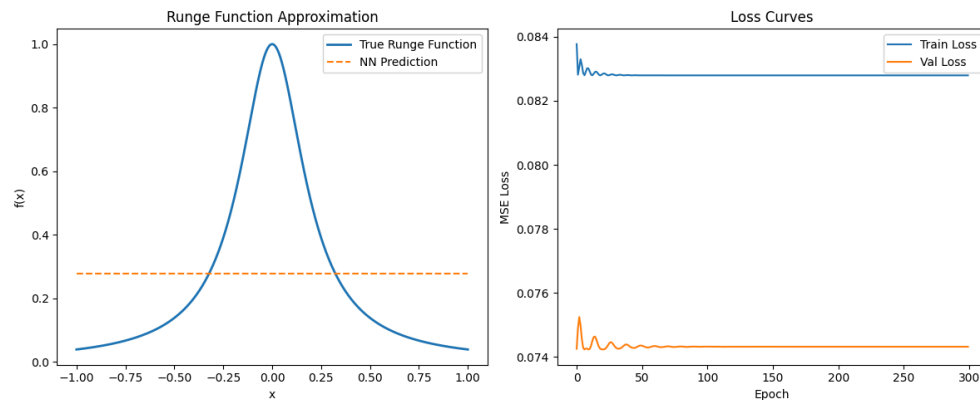
### (2) Tanh



MSE: 0.000050

Since in the input domain, tanh is close to linear, so it won't have the vanishing problem like sigmoid.

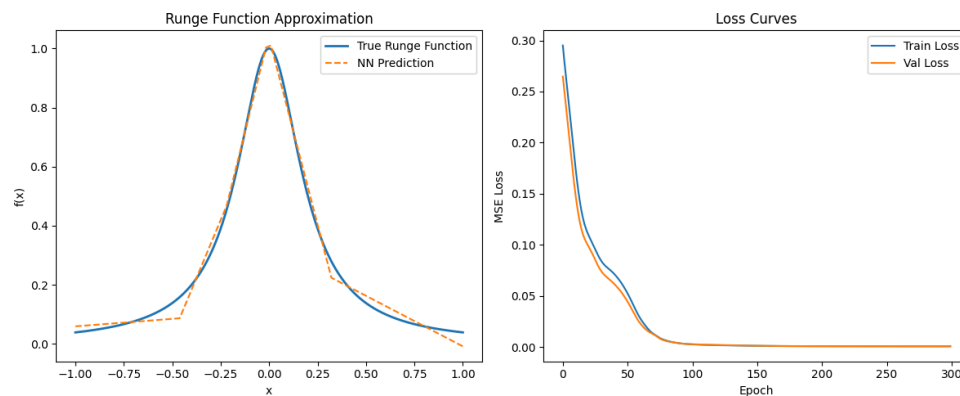
### (3) Relu



MSE: 0.081100

I guess that it is the Dead Relu problem, so I changed the learning rate to 0.006, which is lower than the original one to test it again, the result MSE came down to 0.001648.

### (4) Leaky



MSE: 0.000612

Leaky can solve the problem of Dead Rulu problem

## 4. Conclusion

In this experiment, I try to use different activation functions and try to learn how to increase their performance.