1141ML-Week 2 Programming Assignment

Runge Function Approximation

王宣瑋

September 2025

1 Original way

We basically use Neural Network to complete the program. At first, we use ”tanh” nonlinear function to do the question.

1.1 Settings

- dataset: sampling method (**uniform random** over [-1, 1])
- dataset size:
  - training set: 256
  - validation set: 256
  - test set: 1000
- Neural network architecture:
  - hidden layers: 2
  - nonlinear function: tanh
  - tanh units: 64
- optimizer: Adam
- learning rate: $2e-3$
- weight decay: $1e-5$
- epoches: 800
- Loss metric: MSE

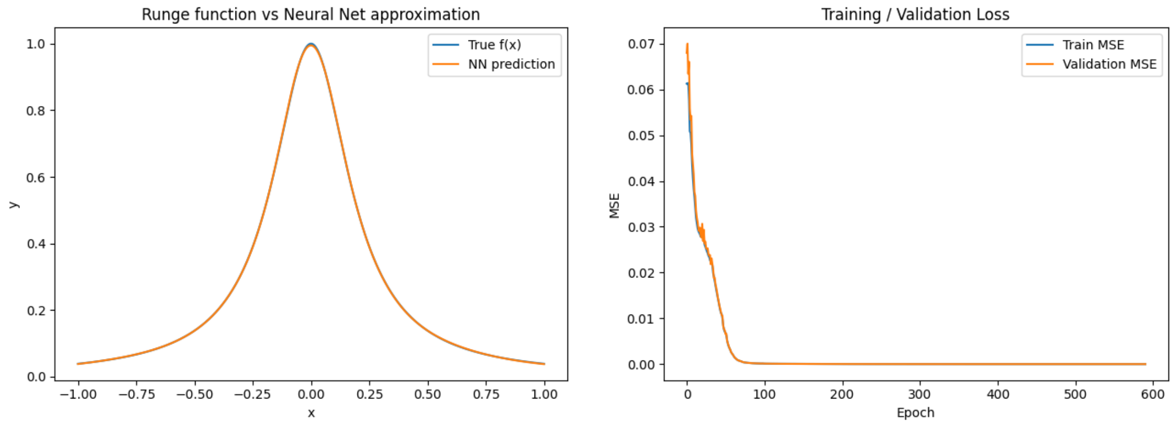
1.2 Result

- Test MSE: 2.070892e-06
- Test Max |err|: 5.308021e-03

1.3 Analysis

Highlight that the NN captured the smooth shape of the Runge function, including the steep slopes near the edges.

And we also found that the MSE is around $2e-6$, which is not low enough.



(a) Runge function vs Neural Net approximation

(b) Training / Validation Loss

Figure 1: Neural network results: (a) Function approximation (b) Loss presentation

2 Enhancement method

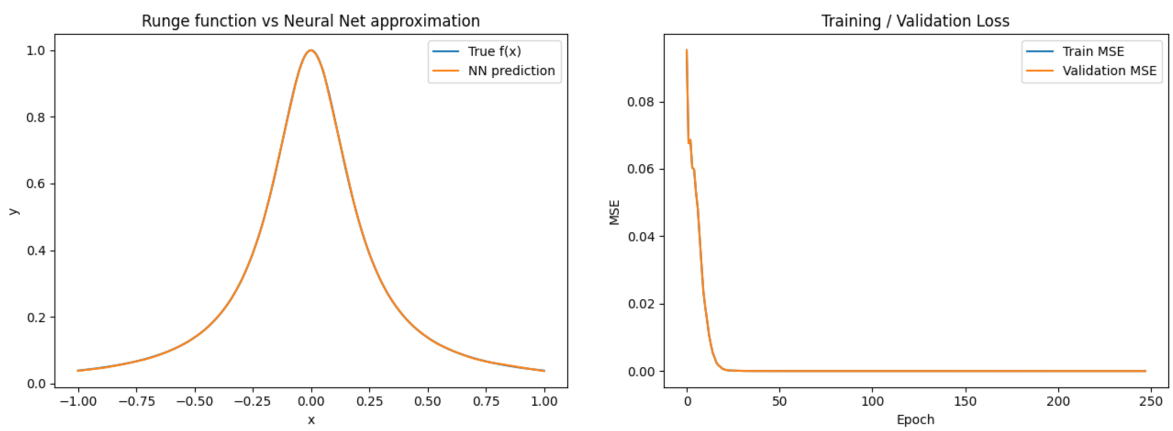
We only show the difference here.

2.1 Settings

- dataset: sampling method (**Chebyshev nodes** over [-1, 1])
- Neural network architecture:
  - nonlinear function: **ReLu**
  - **ReLu** units: 64

2.2 Result

- **Test MSE: 1.404094e-06**
- **Test Max |err|: 5.119538e-03**



(a) Runge function vs Neural Net approximation

(b) Training / Validation Loss

Figure 2: Neural network results: (a) Function approximation (b) Loss presentation

2.3 Analysis

The function approximation plot shows that the ReLU network closely matches the true Runge function across the entire interval $[-1, 1]$, including the steep region near the origin and the flatter boundary regions. The prediction curve is almost indistinguishable from the true function.

The loss curves also illustrate rapid convergence:

- Training loss quickly decreases within the first 50 epochs.
- Validation loss follows a similar trend, indicating no significant overfitting.
- Both losses stabilize at very low levels.

Interpretation:

- The piecewise-linear nature of ReLU allows the network to represent sharp changes in curvature more effectively than smooth saturating activations like Tanh.
- This suggests that ReLU networks can approximate smooth but steep functions with fewer nonlinear distortions, yielding lower error.
- However, one tradeoff is that ReLU networks may introduce "kinks" in the learned function, although this is not visibly significant in this experiment.