Week 2 Program Assignment Report

Design of Network

$$\begin{aligned}
\text{activation function: } & \sigma(x) = \tanh(x) \\
\text{Input layer: } & 1 \text{ input, } x \\
\text{second layer: } & \mathbb{R}^1 \to \mathbb{R}^5, \text{ and } \sigma \\
\text{third layer: } & \mathbb{R}^5 \to \mathbb{R}^5, \text{ and } \sigma \\
\text{output layer: } & \mathbb{R}^5 \to \mathbb{R}^1
\end{aligned}$$

Training Process

$$\begin{aligned}
\text{loss function: } & \text{MSE loss} \\
\text{optimizer: } & \text{Adam} \\
\text{learning rate: } & 0.01 \\
\text{train times: } & 1000
\end{aligned}$$

Training Result

The results show that this model can approximate the original function very well:

1) The curve predicted by the model is obviously very similar to the true function. (see Figure 1)

2) The convergence speed of model training is very fast. In 1000 rounds of training, it basically converges to the minimum value in about 200 rounds. (see Figure 2)

3) The MSE loss of test data and true function value < 0.00001

4) The network structure is simple, takes less time. Even if only use cpu, it only takes about 10 seconds to complete the training.

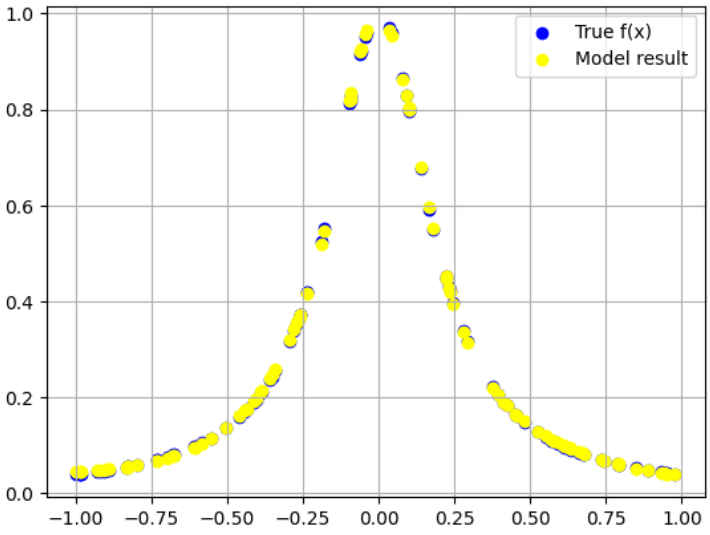

Figure 1: Comparison of prediction and function

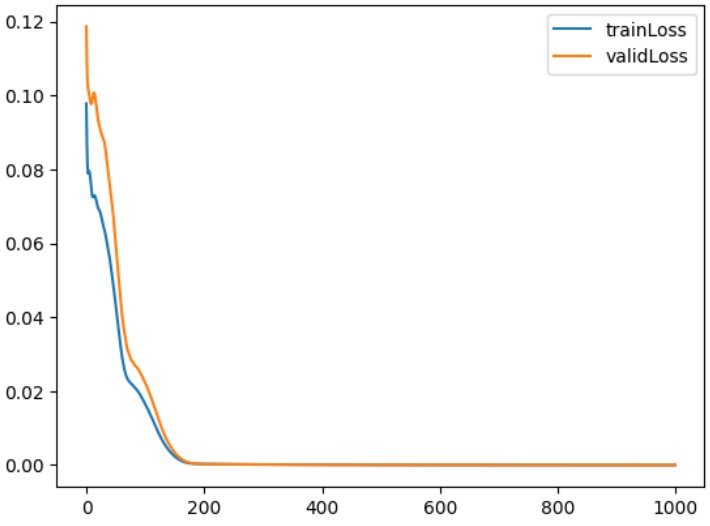

Figure 2: Train and valid error curves