

# 同濟大學

TONGJI UNIVERSITY

## 《神经网络与深度学习》

### 实验报告

实验名称

Simple neural network

实验成员

范诗棋 (2252320)

日期

2052 年 3 月 18 日

## 1、实验背景

理论和实验证明，一个两层的 ReLU 网络可以模型任何函数，现在需要复现此实验证明理论的正确性。自定义一个函数，通过在此函数上独立采样生成训练集和测试集，使用基于 ReLU 的神经网络来拟合此函数，并在测试集上检验模型的拟合效果。

实验代码见文件 chap4\_simple neural network\experiment\_curve\_fitting.ipynb。

## 2、函数定义

采用在定义域 $[-\pi, +\pi]$ 上的非线性周期性函数——正弦函数  $y = \sin(x)$  作为拟合目标。

## 3、数据采集

在定义域 $[-\pi, +\pi]$ 上，通过 numpy 的 linspace 函数采样一百个数据点，并通过 sin 函数生成对应的拟合目标。

## 4、模型描述

### 4.1 模型结构

模型有两层全连接神经网络（MLP），用于拟合前述定义的函数  $y=\sin(x)$ 。模型的主要模块包括：

- (1) 全连接层(Linear Layer)：用于将输入特征映射到隐藏层或输出层。
- (2) ReLU 激活函数：引入非线性，增强模型的表达能力。

模型的具体结构如下：

1. 第一层全连接层：将输入  $x$  (形状为  $(N, \text{input\_dim})$ ) 通过权重矩阵  $W1$  (形状为  $(\text{input\_dim}, \text{hidden\_dim})$ )，映射到隐藏层  $h1$  (形状为  $(N, \text{hidden\_dim})$ )。
2. ReLU 激活函数：对隐藏层  $h1$  应用 ReLU 激活函数，得到  $h1\_relu$ 。
3. 第二层全连接层：将  $h1\_relu$  通过权重矩阵  $W2$  (形状为  $(\text{hidden\_dim}, \text{output\_dim})$ ) 映射到输出层  $h2$  (形状为  $(N, \text{output\_dim})$ )。

### 4.2 前向传播

模型的前向传播过程如下：

1. 输入数据  $x$  通过第一层全连接层，得到  $h1 = x \cdot W1$ 。
2. 对  $h1$  应用 ReLU 激活函数，得到  $h1\_relu = \text{ReLU}(h1)$ 。
3.  $h1\_relu$  通过第二层全连接层，得到  $h2 = h1\_relu \cdot W2$ 。

### 4.2 反向传播

模型的反向传播过程如下：

1. 计算 MSE 损失对  $h2$  的梯度：

$$\frac{\partial L}{\partial h2} = h2 - target$$

2. 计算损失函数对  $h1\_relu$  和  $W2$  的梯度:

$$\frac{\partial L}{\partial h1\_relu} = \frac{\partial L}{\partial h2} \cdot W2^T$$

$$\frac{\partial L}{\partial W2} = h1\_relu^T \cdot \frac{\partial L}{\partial h2}$$

3. 计算损失函数对  $h1$  的梯度:

$$\frac{\partial L}{\partial h1} = \frac{\partial L}{\partial h1\_relu} \cdot \frac{\partial h1\_relu}{\partial h1}$$

其中  $\frac{\partial h1\_relu}{\partial h1}$  是 ReLU 的梯度公式。

4. 计算损失函数对  $x$  和  $W1$  的梯度:

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial h1} \cdot W1^T$$

$$\frac{\partial L}{\partial W1} = x^T \cdot \frac{\partial L}{\partial h1}$$

## 4.2 参数更新

模型完成反向传播后, 使用梯度下降法更新模型的参数:

$$W_1 = W_1 - \alpha \frac{\partial L}{\partial W_1}$$

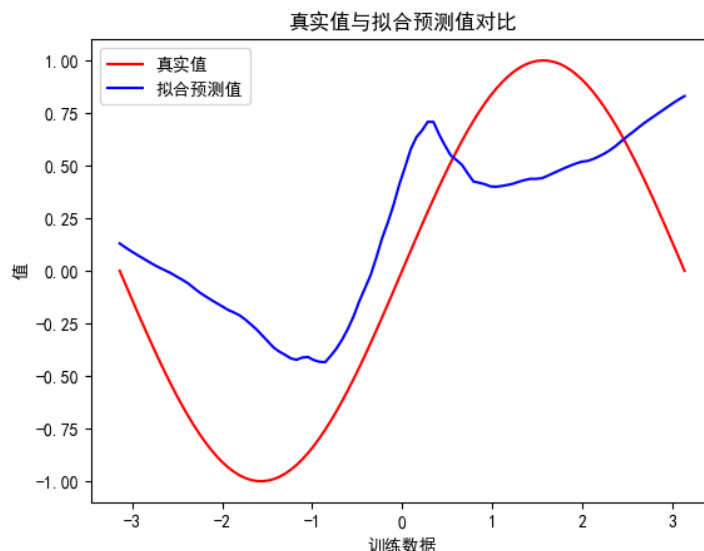
$$W_2 = W_2 - \alpha \frac{\partial L}{\partial W_2}$$

其中  $\alpha$  为学习率。

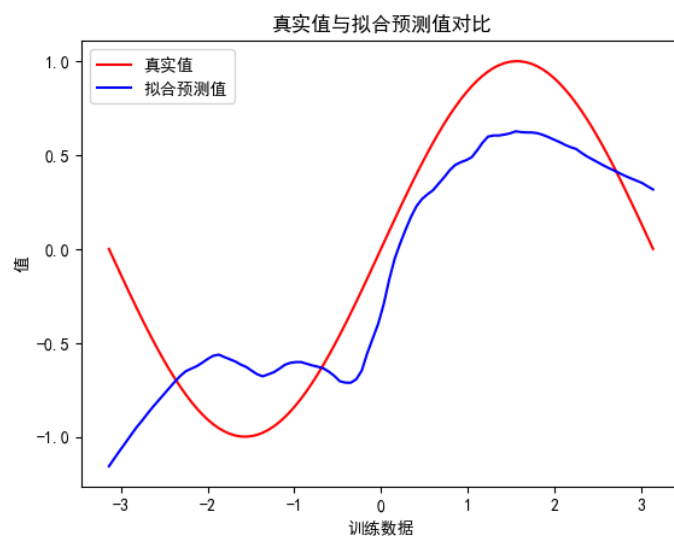
## 5、拟合效果

最初定义网路的隐藏层大小为 500, 学习率为  $1e-5$ 。

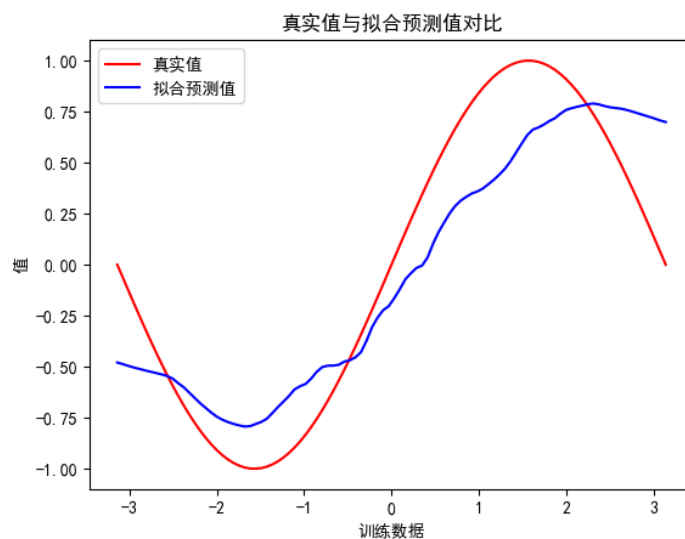
训练轮数为 2000 时, RMSE 损失为 0.4758。



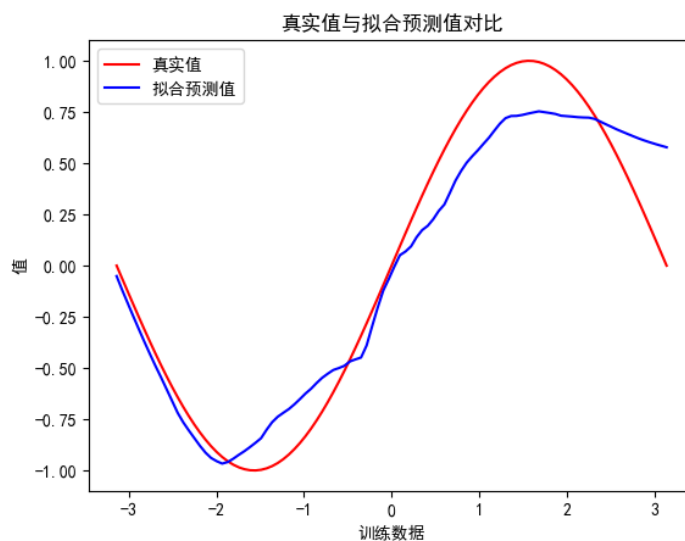
训练轮数为 4000 时, RMSE 损失为 0.3598。



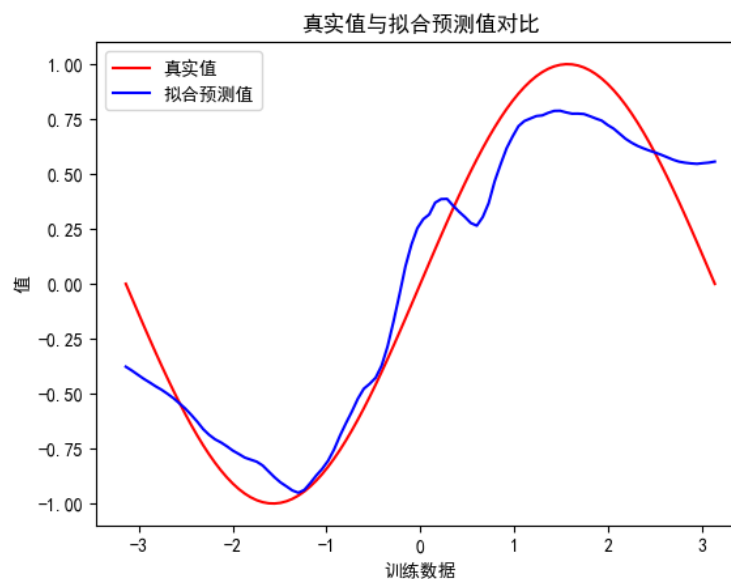
训练轮数为 8000 时，RMSE 损失为 0.2988。



训练轮数为 32000 时，RMSE 损失为 0.1943。



隐藏层大小改为 1000，训练 32000 轮，MSE 损失为 0.1893。



## 5、实验总结

在随着训练次数的增加，两层的 ReLU 全连接神经网络对 $[-\pi, +\pi]$ 区间上的  $y = \sin(x)$  函数的拟合能力逐渐增强，可以预见，当训练轮数和隐藏层神经元个数足够大时，神经网络的拟合损失可以小于任意正数；体现了两层的 ReLU 网络拟合函数的能力。