实验证明

函数定义

$$f(x) = 2sin(3x) + 0.5x^2 + 0.1e^x$$

数据采集

- **训练集**：在区间 [−5,5] 内随机采样 1000 个点。
- **测试集**：在相同区间内均匀采样 200 个点。

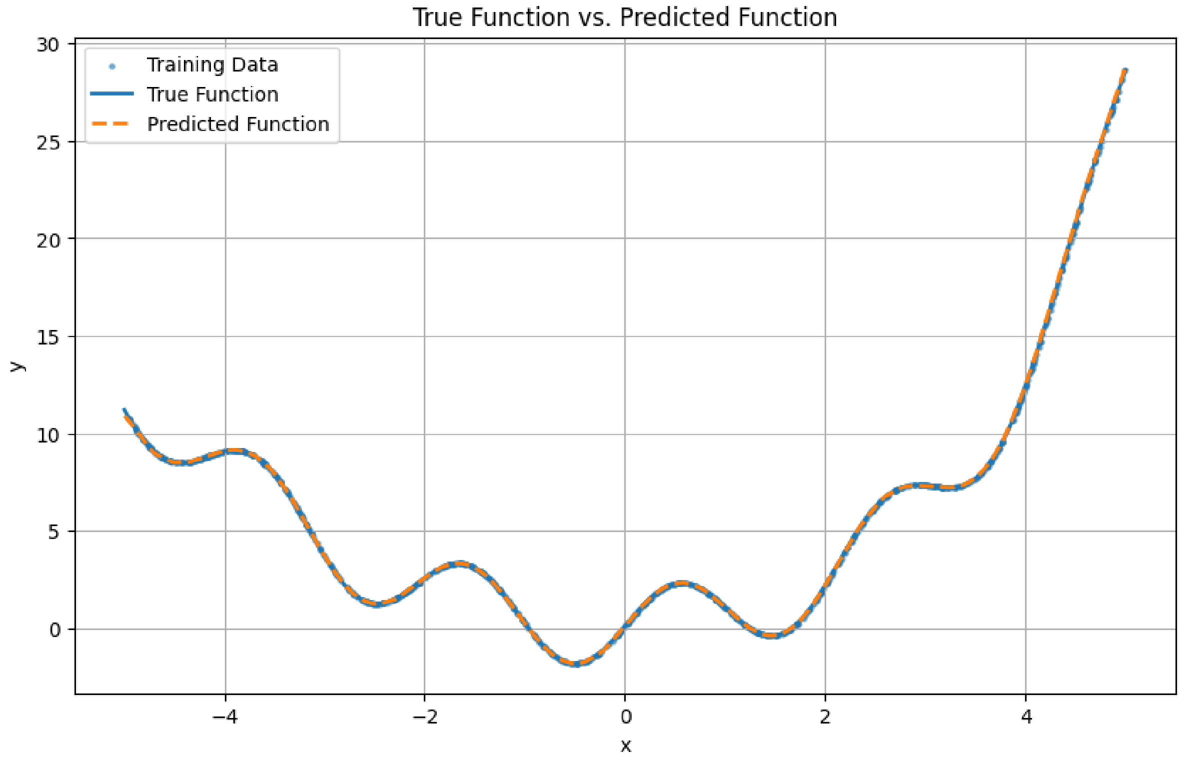
模型描述

- **网络结构**：两层 $ReLU$ 网络（两个隐藏层），每层 200 个神经元。
    - 输入层 → 隐藏层1 ($ReLU$) → 隐藏层2 ($ReLU$) → 输出层。
- **损失函数**：均方误差（MSE）。
- **优化器**：Adam，学习率设为 0.001。
- **训练轮数**：2000 epochs，批量大小 32。

拟合效果

- **损失曲线**：训练和测试损失均随着训练轮数增加而稳步下降，最终收敛到较低值（约0.001）。

- **预测对比**：预测结果（虚线）与真实函数（实线）几乎完全重合，表明模型成功拟合了目标函数。训练数据点（散点）均匀分布在真实曲线周围，进一步验证了拟合效果。



理论证明

1. **通用逼近定理基础**
   Cybenko（1989）[1]和 Hornik 等人（1989）[2]证明了使用 $Sigmoid$ 类激活函数的单隐藏层神经网络能够以任意精度逼近紧集上的连续函数。关键条件是激活函数需为连续、有界、非恒定的单调函数。
2. **扩展至非多项式激活函数**
   Leshno 等人（1993）[3]进一步扩展了这一结论，指出只要激活函数是非多项式函数（即使是无界的或非单调的），单隐藏层神经网络即可成为通用逼近器。 $ReLU$ 函数（$\sigma(x) = max(0, x)$）是分段线性且非多项式的，因此符合条件。
3. $ReLU$ **的通用逼近性**
   - **非多项式性质**：$ReLU$ 无法表示为有限项多项式组合，因此满足 $Leshno$ 定理的核心条件。
   - **构造性证明**：通过组合多个 $ReLU$ 单元，可形成分段线性函数。增加隐藏单元数量可提升分段数量，从而逼近任意连续函数（如图像由折线段逼近曲线）。
   - **高维推广**：每个 $ReLU$ 单元对应一个超平面划分输入空间，多单元组合可构造复杂区域，逼近多维函数。
4. **两层网络结构**
   - **隐藏层**：使用 n 个 $ReLU$ 单元，每个单元计算 $h_i = ReLU(w_i^T x + b_i)$。
   - **输出层**：线性组合隐藏层输出，即 $f(x) = \sum_{i=1}^{n} c_i h_i + c_0$。
   - **逼近过程**：当 n 足够大时，通过调整 wi，b_i，c_i，$f(x)$可逼近目标函数。
5. **实践支持**
   Nair & Hinton（2010）[4]和 Glorot 等人（2011）[5]通过实验验证了 $ReLU$ 在深层网络中的有效性，间接支持其逼近能力。

参考文献：

[1] G. Cybenko. 1989. Approximation by superpositions of a sigmoidal function.

[2] K. Hornik, M. Stinchcombe, and H. White. 1989. Multilayer feedforward networks are universal approximators.

[3] Moshe Leshno, et al. 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function

[4] Vinod Nair and Geoffrey E. Hinton. 2010. Rectified linear units improve restricted boltzmann machines.

[5] Xavier Glorot, Antoine Bordes, Yoshua Bengio. 2011. Deep Sparse Rectifier Neural Networks. PMLR 15:315-323.