

### 目标函数：

自变量  $X$  属于  $(0, 1)$

$Y = (-\cos(X \cdot 10) + 1.0) / 4.0 + X / 2.0$  （把值域调整到方便观察而已）

离散数据点之间的距离 0.005，一共 199 个数据点

### 模型 1：

层数 2

$$\text{ReLU}(X \cdot w_{11} + b_{11}) = h_1$$

$$\text{ReLU}(X \cdot w_{12} + b_{12}) = h_2$$

$$\text{ReLU}(h_1 \cdot w_{21} + h_2 \cdot w_{22} + b_2) = Y'$$

参数 7 个： $w_{11}, b_{11}, w_{12}, b_{12}, w_{21}, w_{22}, b_2$

### 损失函数：

$$\frac{\sum (Y_i' - Y_i)^2}{n} \quad \text{其中 } n \text{ 是数据点的个数}$$

### 参数初始化：

he\_initializer (<https://zhuanlan.zhihu.com/p/25110150>)

不过最简单的应该是随机初始化

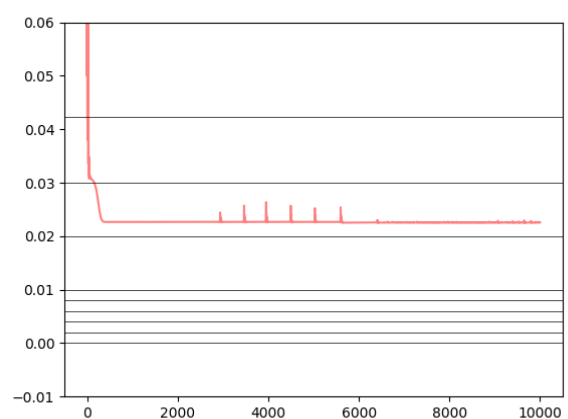
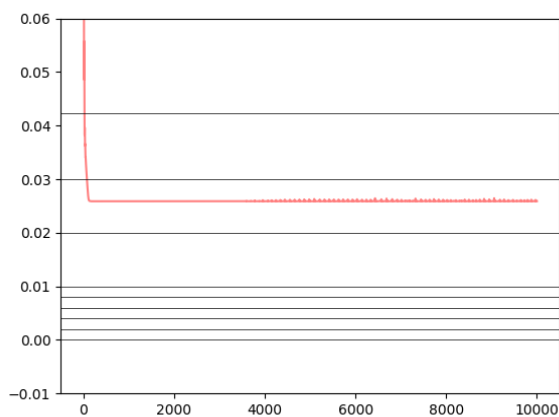
### 优化器：

AdamOptimizer (<https://zhuanlan.zhihu.com/p/22252270>) 初始  $dt=0.1$ ，优化 10000step

不过最简单的应该是随机下降

### 实验结果：

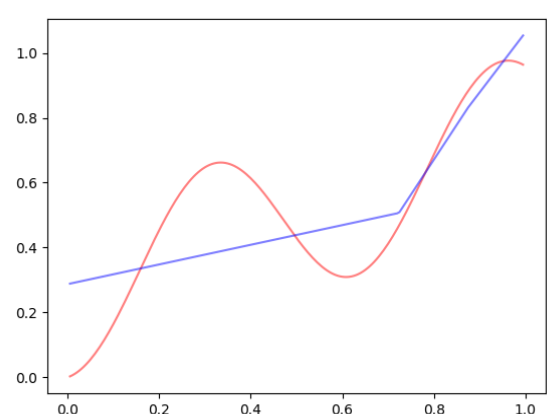
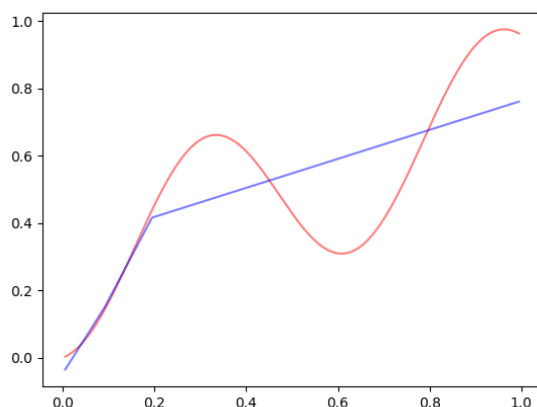
损失函数的下降：（多次实验的曲线结果类似，此处仅显示 2 例）



### 拟合结果：

红色：原曲线（光滑连接了离散数据点）

蓝色：神经网络拟合曲线



### 修改模型

#### 模型 2:

层数仍然 2，第 1 层多加一个 feature

$$\text{ReLU}(X * w_{11} + b_{11}) = h_1$$

$$\text{ReLU}(X * w_{12} + b_{12}) = h_2$$

$$\text{ReLU}(X * w_{13} + b_{13}) = h_3$$

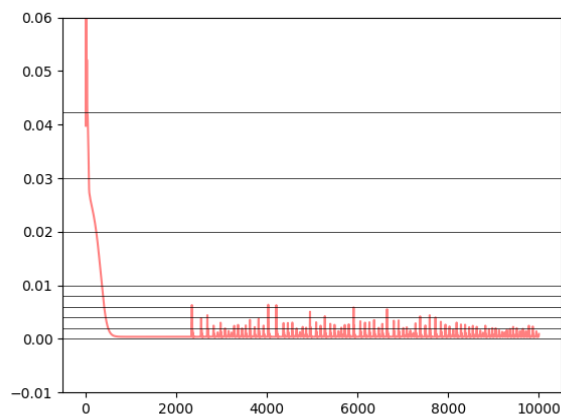
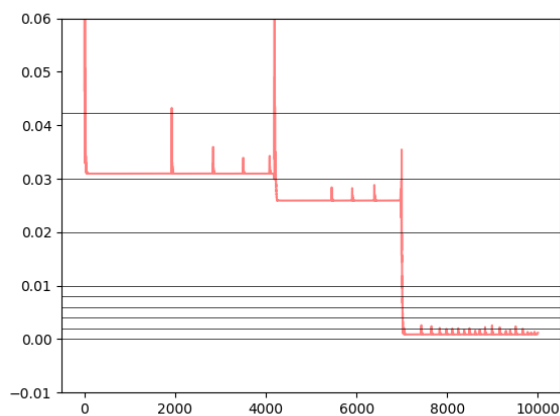
$$\text{ReLU}(h_1 * w_{21} + h_2 * w_{22} + h_3 * w_{23} + b_2) = Y'$$

参数 10 个:  $w_{11}, b_{11}, w_{12}, b_{12}, w_{13}, b_{13}, w_{21}, w_{22}, w_{23}, b_2$

其余设置不变

**实验结果:** (也有失败的时候, 结果图形类似模型 1, 但是不像模型 1 几乎没有成功的时候, 模型 2 一般跑几次就一定会有成功的时候)

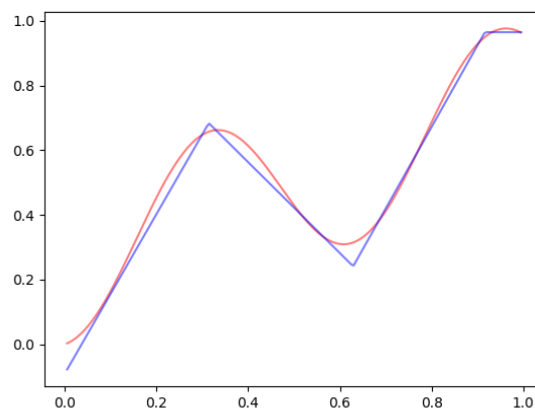
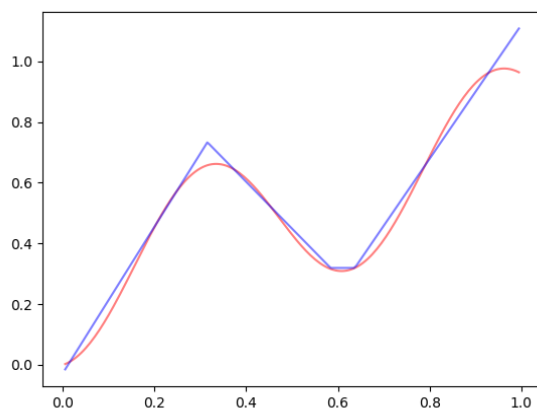
**损失函数的下降:** (多次实验的曲线结果类似, 此处仅显示 2 例)



#### 拟合结果:

红色: 原曲线 (光滑连接了离散数据点)

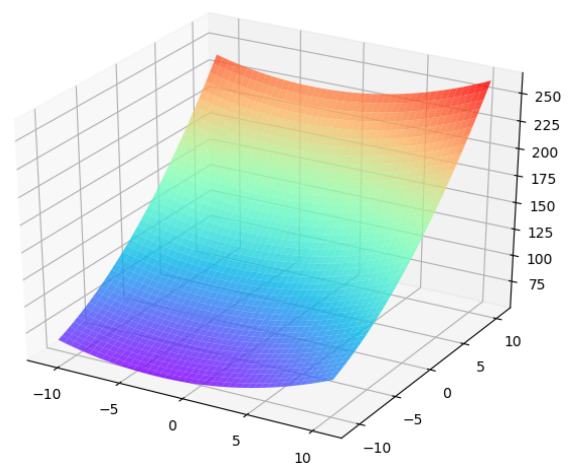
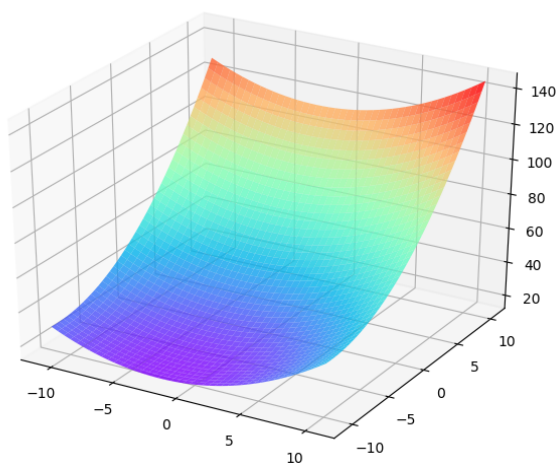
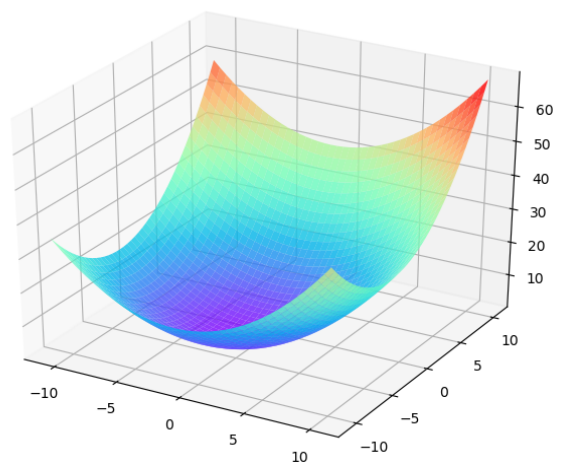
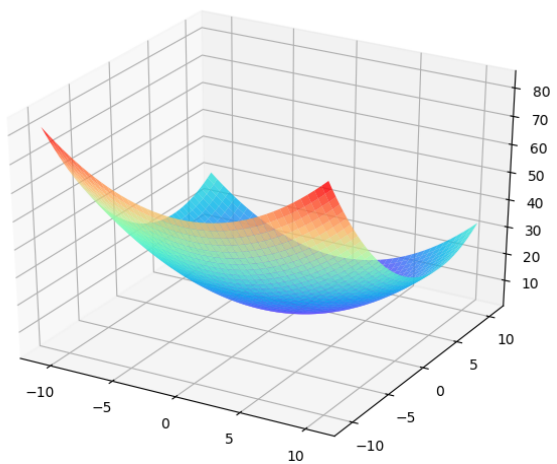
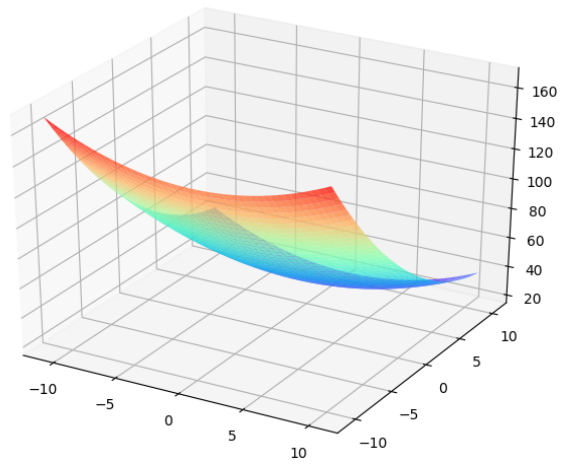
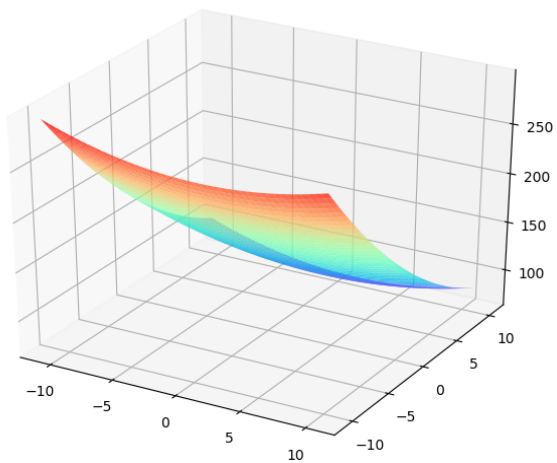
蓝色: 神经网络拟合曲线



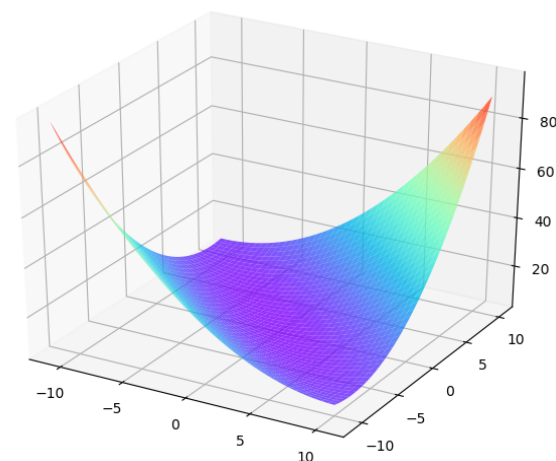
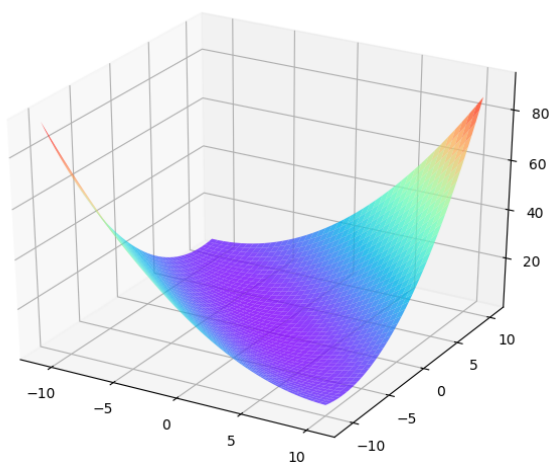
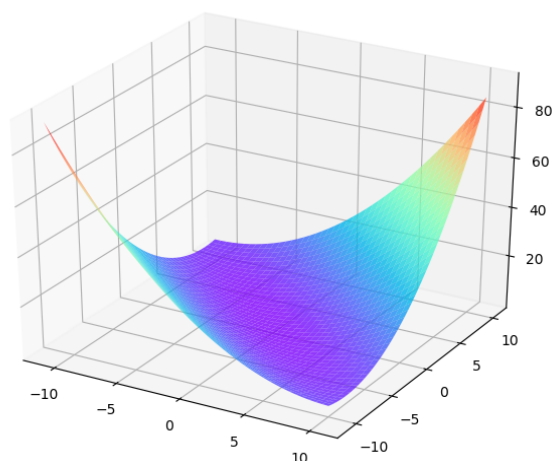
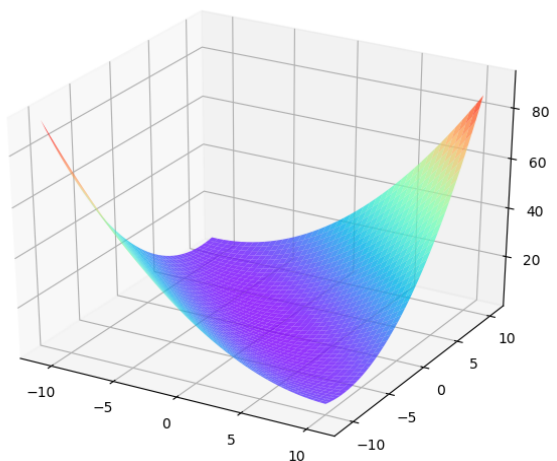
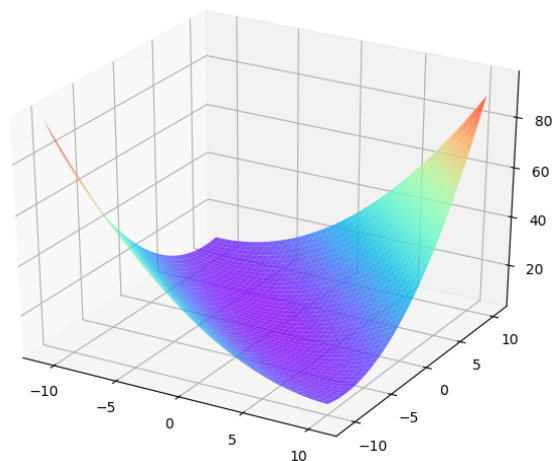
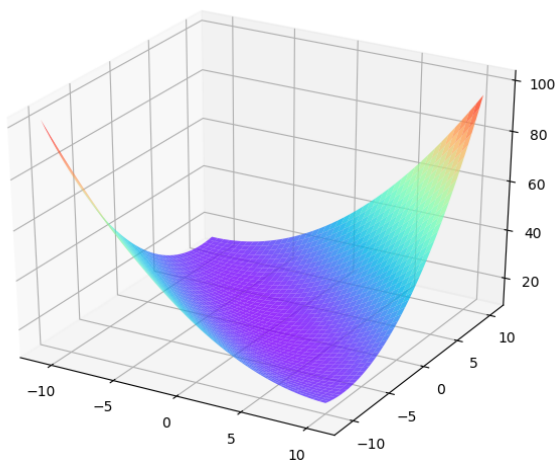
#### 简单考察:

学习成功时  $w_{21}, w_{22}, w_{23}$  形成的 loss surface 地形如下:

当  $w_{21}$  分别为 -10, -6, -2, 2, 6, 10 时,  $w_{22}, w_{23}$  构成的 loss surface 如下:



也有拟合出 Z 字形折线，但是  $w_{21}$  分别为 -10, -6, -2, 2, 6, 10 时， $w_{22}, w_{23}$  构成的 loss surface 变化不大的，如下：



困扰：缺乏高维地形特征的观测手段，不知道应该看什么量，不知道怎么用数学描述问题，更别说解了，等等