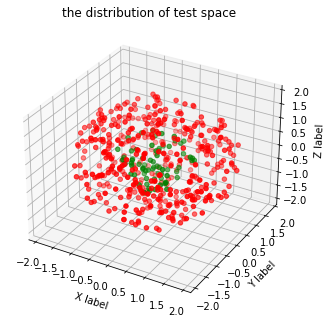
**自编码器性质探究实验**

**实验一：**

数据集：圆内部



如图，绿色点代表单位球内部，作为正样本，红色点代表半径小于2大于1的圆环，作为负样本。

实验：

epoch = 100

batch\_size = 16

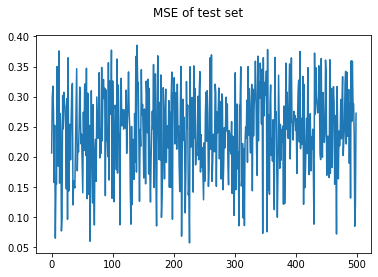
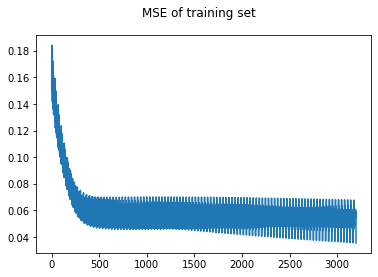
train\_set: 500

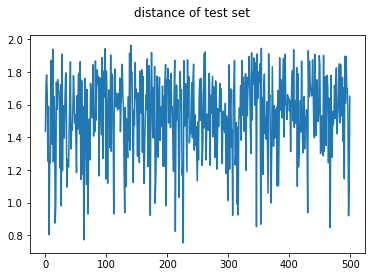
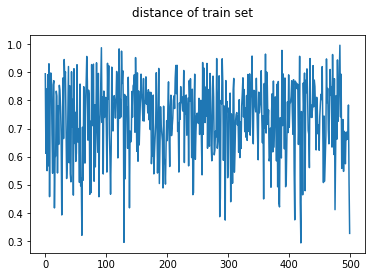
test\_set: 500

模型为三层：输入层，隐层，输出层。

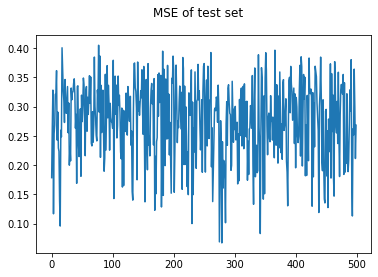
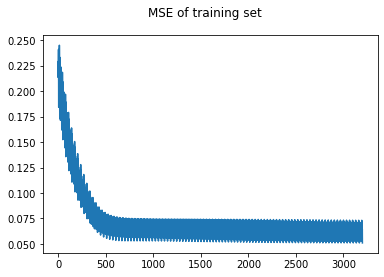
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 输入维度 | 隐层维度 | 阈值 | 训练集准确率 | 测试集准确率 |
| 实验1 | 10 | 3 | 1 | 1.0 | 0.956 |
| 实验2 | 10 | 2 | 1 | 0.996 | 0.992 |
| 实验3 | 10 | 1 | 1 | 0.992 | 0.998 |
| 实验4 | 3 | 2 | 1 | 1.0 | 0.502 |
|  | 3 | 2 | 0.7 | 0.862 | 0.726 |
|  | 3 | 2 | 0.6 | 0.782 | 0.768 |
| 实验5 | 3 | 1 | 1 | 0.996 | 0.776 |
|  | 3 | 1 | 0.7 | 0.688 | 0.880 |
|  | 3 | 1 | 0.6 | 0.550 | 0.890 |
|  |  |  |  |  |  |

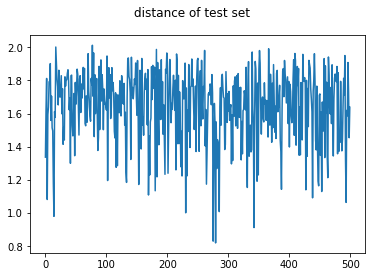
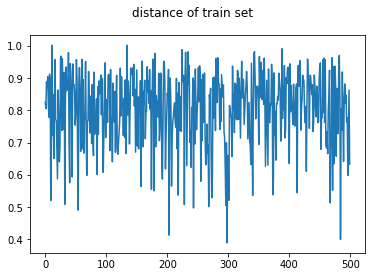
实验1：



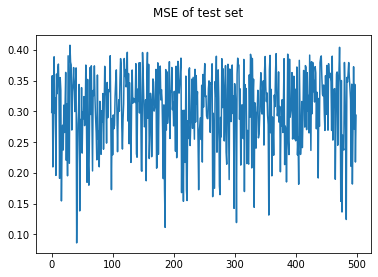
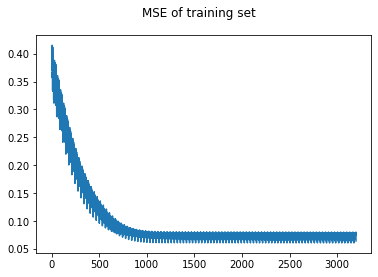


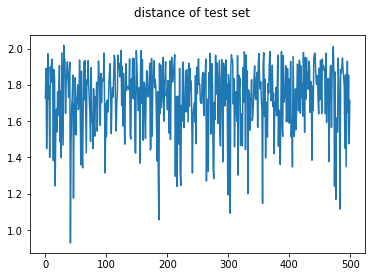
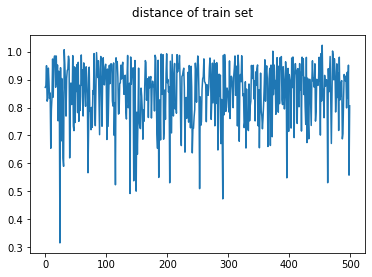
实验2：



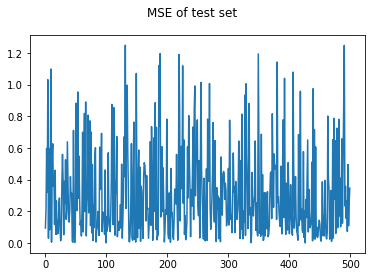
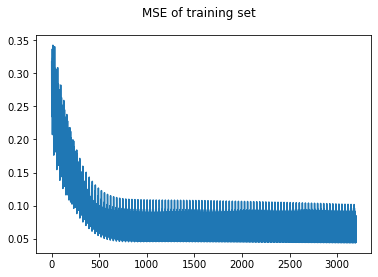


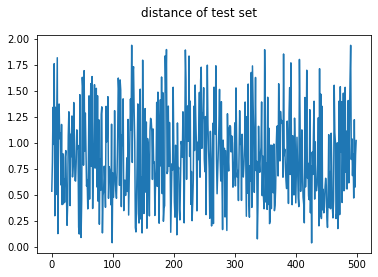
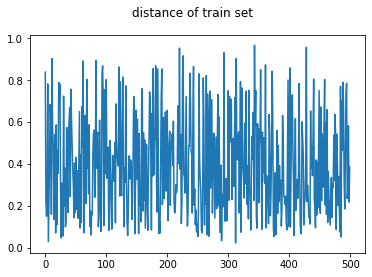
实验3：

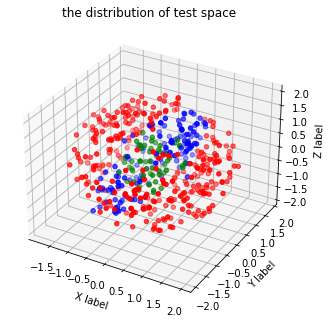




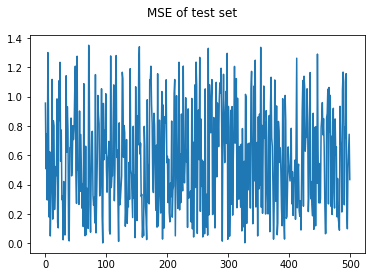
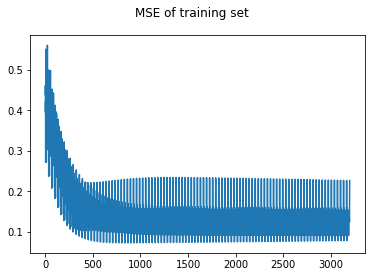
**实验4：**

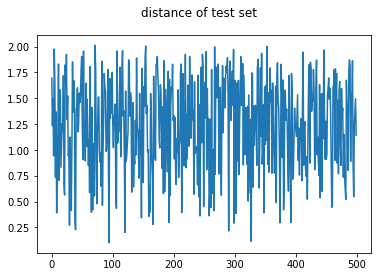
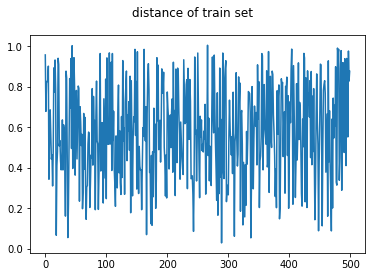


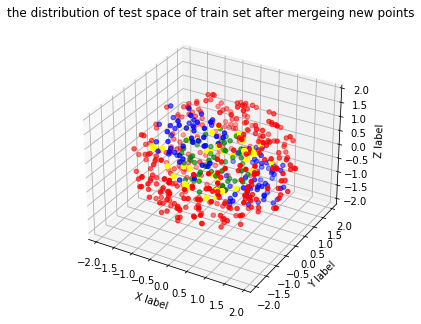
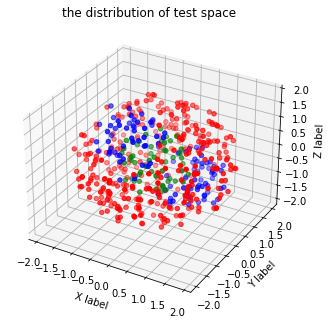




**实验5：**







**通过实验，发现难以看出具有某种性质；而且实验中显示，当数据维度比较大的时候,测试集的准确率更高。**

**构造简单数据集：以0为圆心的球，发现难以发掘到有意义的信息；要从理论的数学推导中，得到是否是一个曲面也是困难的。**

**自编码器性质探究**

## 问题描述与数学表达

自编码器作为一类特殊的神经网络框架，可以通过重构自身的方式，获得一个优异的编码向量。基于这个框架，学者建立了异常检测模型。

基于自编码器的异常检测模型，它的做法十分简单：在训练阶段，用正常数据训练模型，学习正常数据与重构的近似恒同映射关系。然后在测试阶段，输入异常数据和正常数据。在这种映射关系下，异常数据与重构的结果的损失函数会远大于正常数据的损失函数，进而我们通过人为设置阈值，区分正常与异常数据。

但是，由于现有的方法，是自编码器看作是一个黑盒模型，完成一个具体的任务。现在我们想通过异常检测这个任务，解释自编码器的内部机理。因此，我们提出以下几个问题：

问题1. 通过区分异常与正常数据的阈值，我们可以计算出边界点，如下式：

我们想知道，这个阈值边界方程会不会是一个完美的曲面？

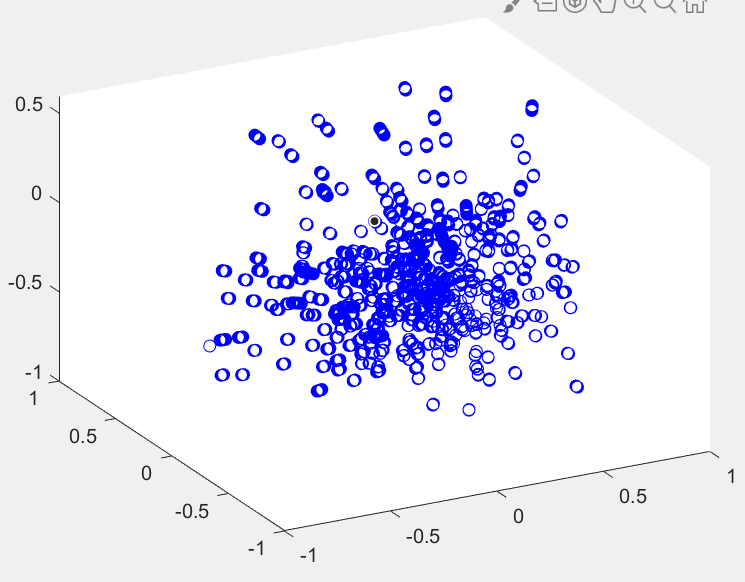
通常，这个方程是非线性函数的复合函数，所以难以直接求得解析解，因此本文采用遍历的方式和方程的数据解法——牛顿法，可以求的方程的解。如果是x是三维及以下的数据，则可以画出曲面图。

其实这个问题是非常好解释的，已知曲面的定义是：若曲面与三元方程，有下述关系：1、曲面上任一点的坐标均满足上述方程；2、不在曲面上的点的坐标都不满足上述方程。那么，上述方程称作曲面的方程，而曲面称作该方程的图形。

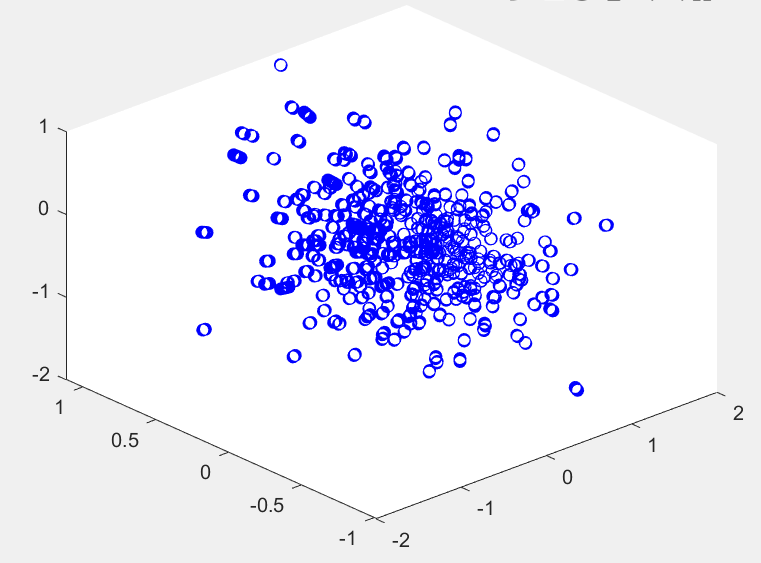
因此，本身就是曲面的方程，根据上述方法解出来的解就是服从这个曲面方程的曲面上的点。

问题2. 自编码器的隐变量具有什么特殊的性质？

没有进行正则约束和特殊设计的自编码器，它的隐变量是不具有特殊性质的，相似数据映射之后，产生的隐变量也是聚作一团。隐变量空间不具有连续性和完备性的。



10维映射到3维的隐变量分布（激活函数：Tanh）



10维映射到3维的隐变量分布（激活函数：无，即MLP）

## 探究实验

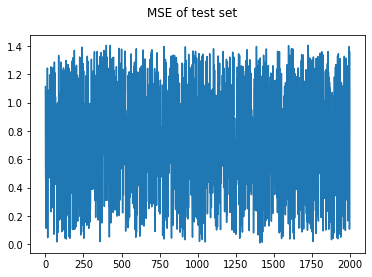
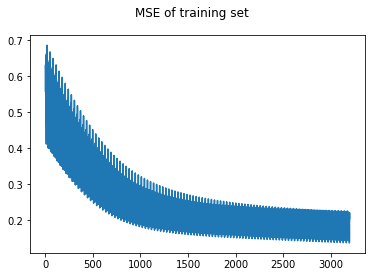
**为了可视化输入数据与阈值边界数据，实验中的输入数据为3维数据，然后分别映射到2维或1维，观察实验结果。**

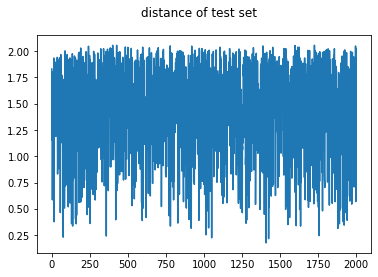
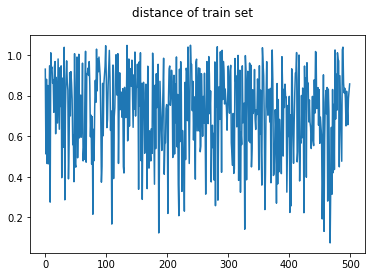
**首先构造数据集：训练集为单位球，测试集为半径为2的球。训练集大小：train\_num = 500，测试集大小：test\_num = 2000。**

**实验如下：**

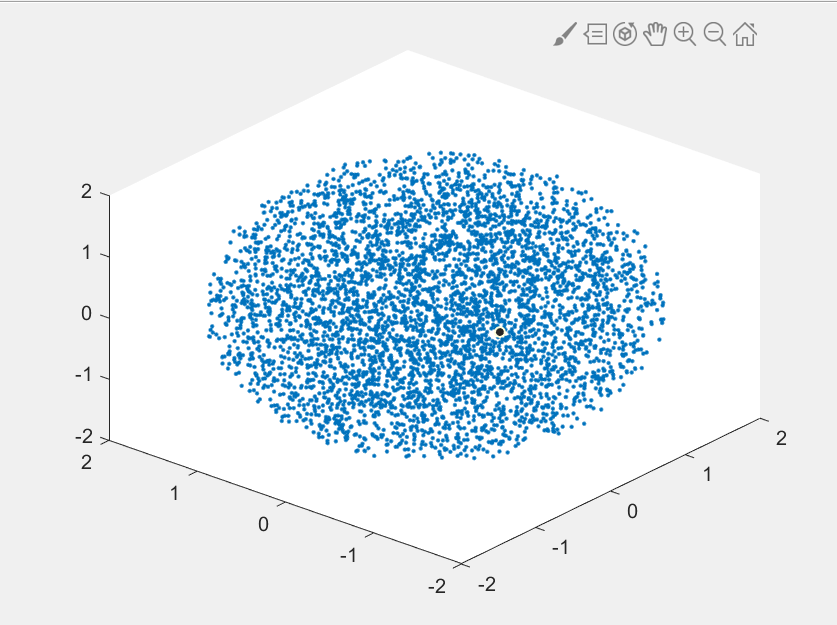
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 激活函数 | 隐层数 | 阈值 | 训练集acc | 测试集acc |
| 实验1 | **Sigmoid** | **2** | **1** | **0.942** | **0.946** |
|  | **Sigmoid** | **1** | **1** | **0.908** | **0.954** |
| 实验2 | **None** | **2** | **1** | **0.998** | **0.432** |
|  | **None** | **1** | **1** | **0.98** | **0.755** |
| 实验3 | **Tanh** | **2** | **1** | **1.0** | **0.531** |
|  | **Tanh** | **1** | **1** | **1.0** | **0.828** |
| 实验4 | **ReLU** | **2** | **1** | **1.0** | **0.894** |
|  | **ReLU** | **1** | **1** | **0.998** | **0.933** |
| 实验5 | **ELU** | **2** | **1** | **1.0** | **0.498** |
|  | **ELU** | **1** | **1** | **0.996** | **0.796** |

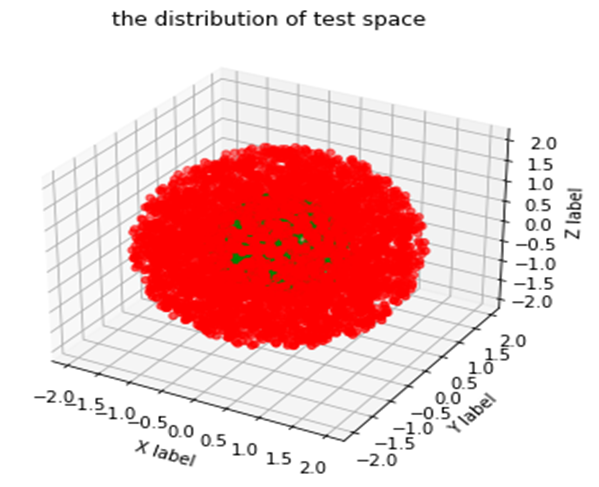
**(1)激活函数为Sigmoid(), threshold =1**



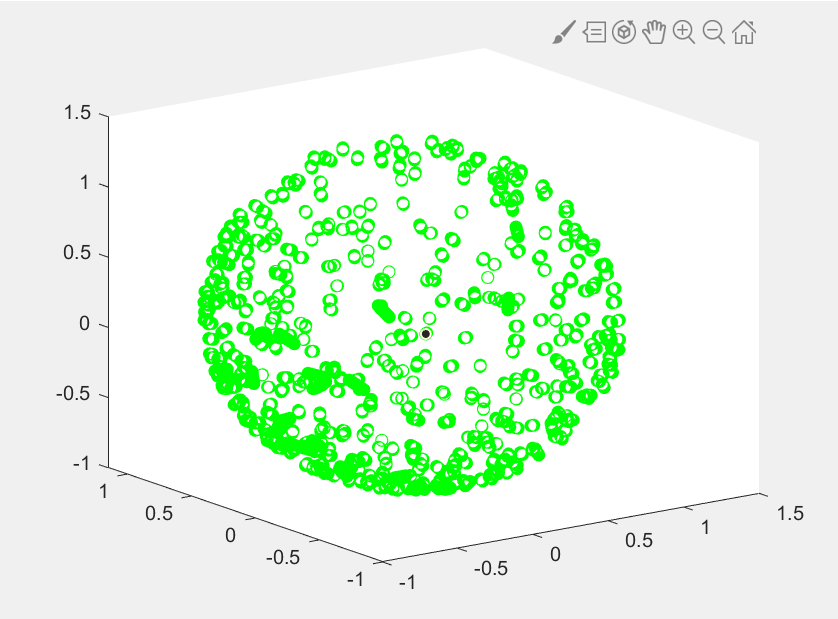


**测试集test\_data：**

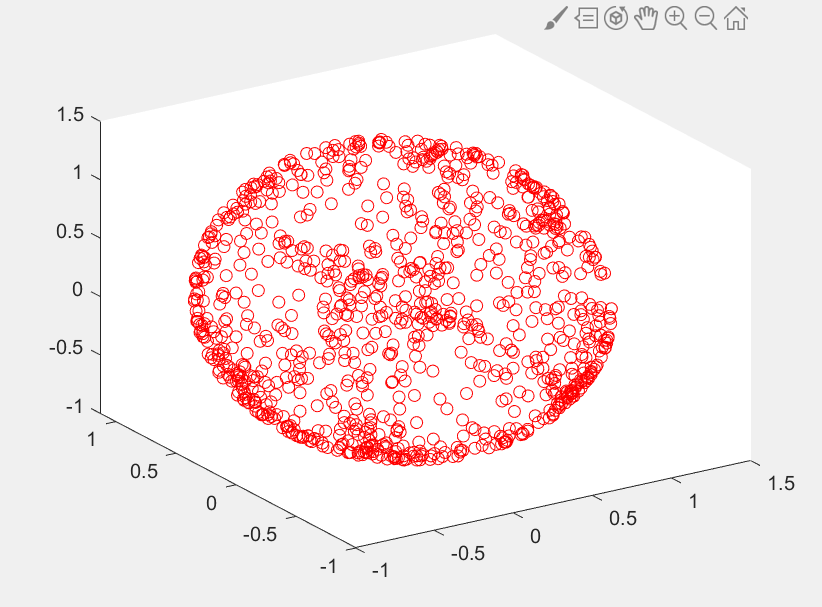


****

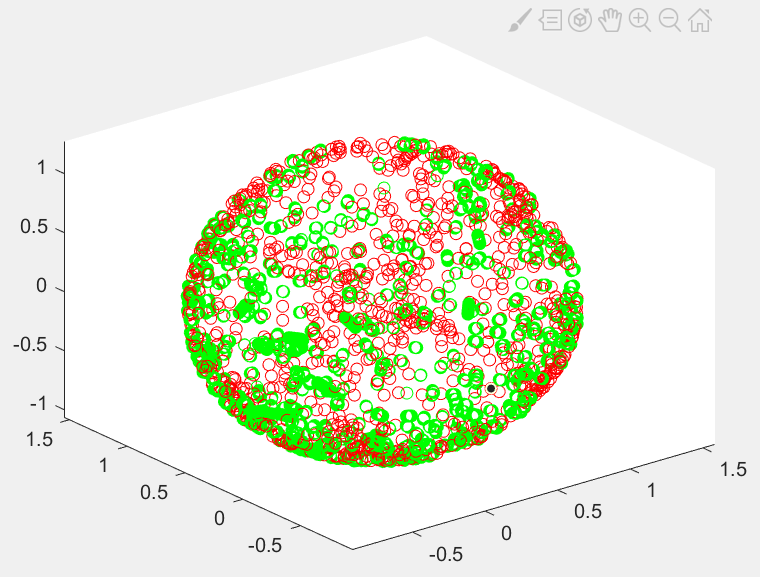
**遍历法求解：**



**Newton法解方程：**

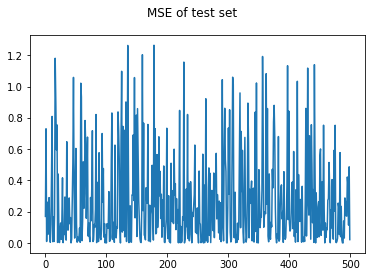
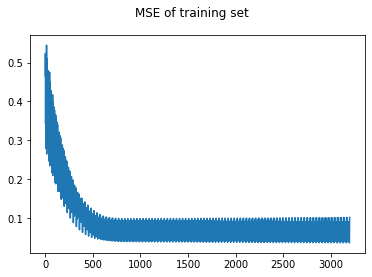


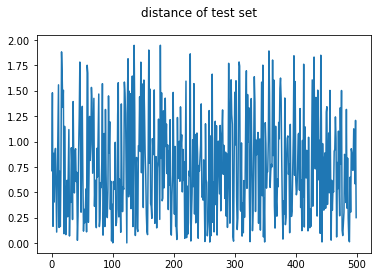
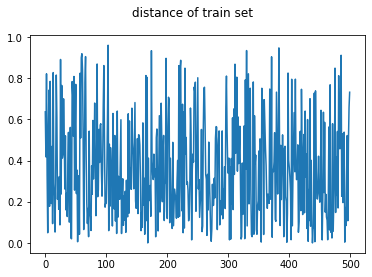
**将这个图画在一起比较：**

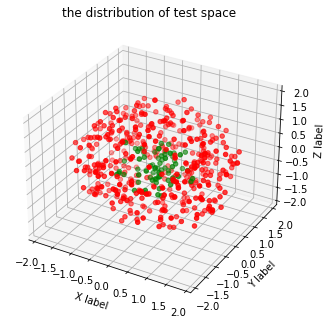


**通过画图发现两者图像近乎重合。**

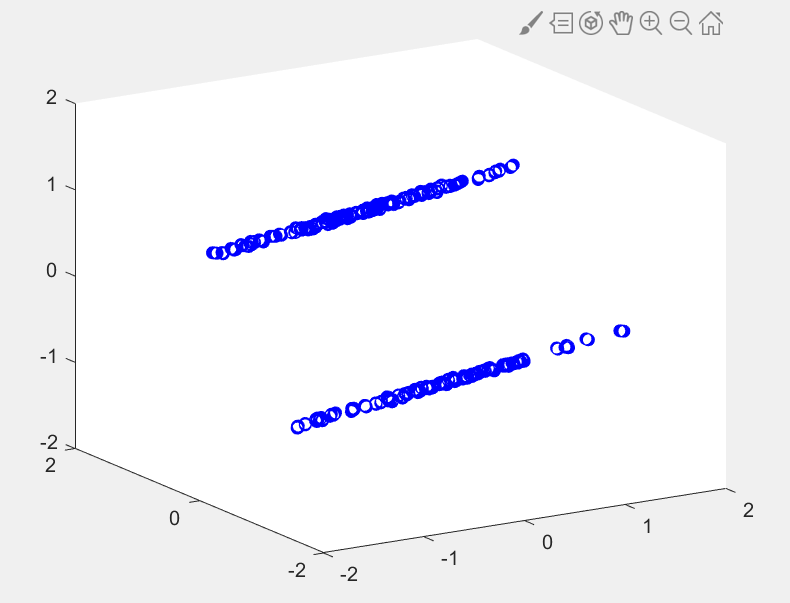
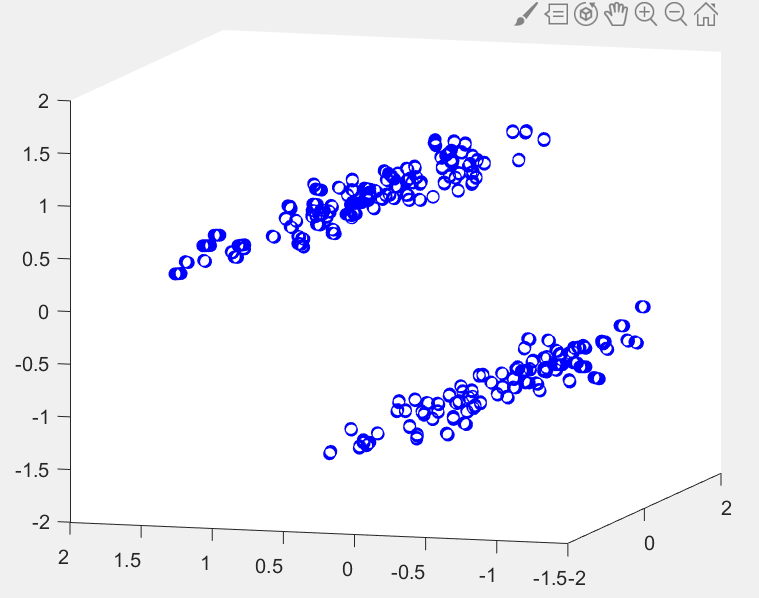
**(2)激活函数去掉，即采用全连接 threshold=1：**



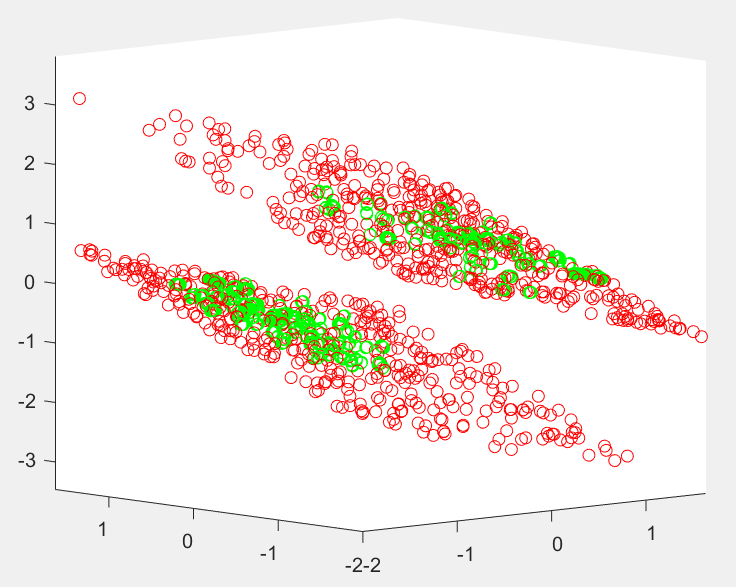




**遍历法：**

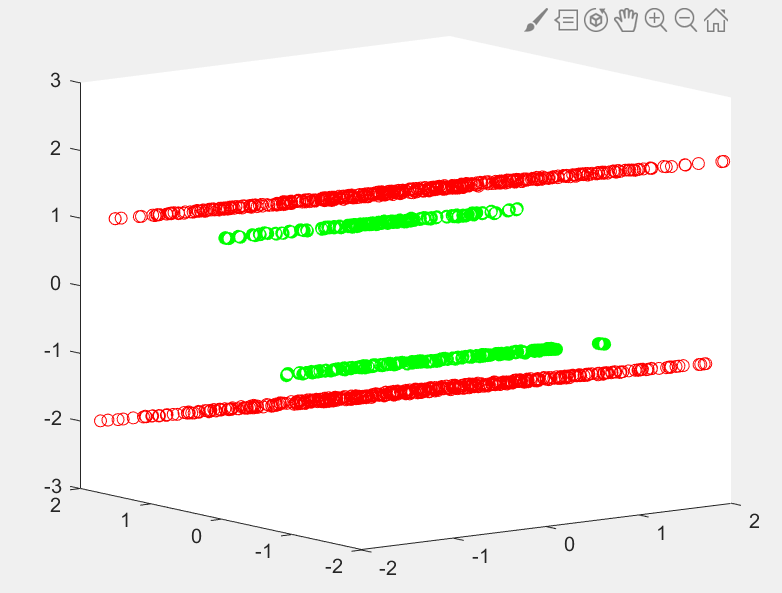


**牛顿法：**



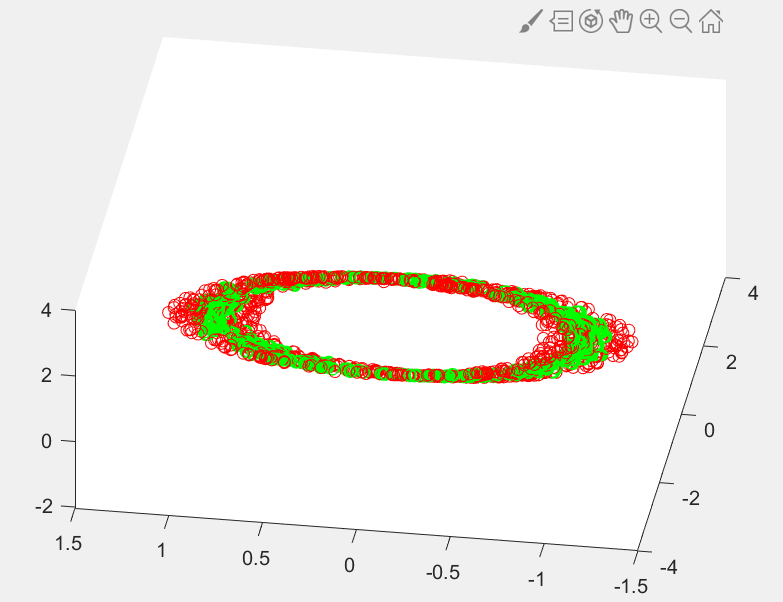
**结果是两个平行的平面**

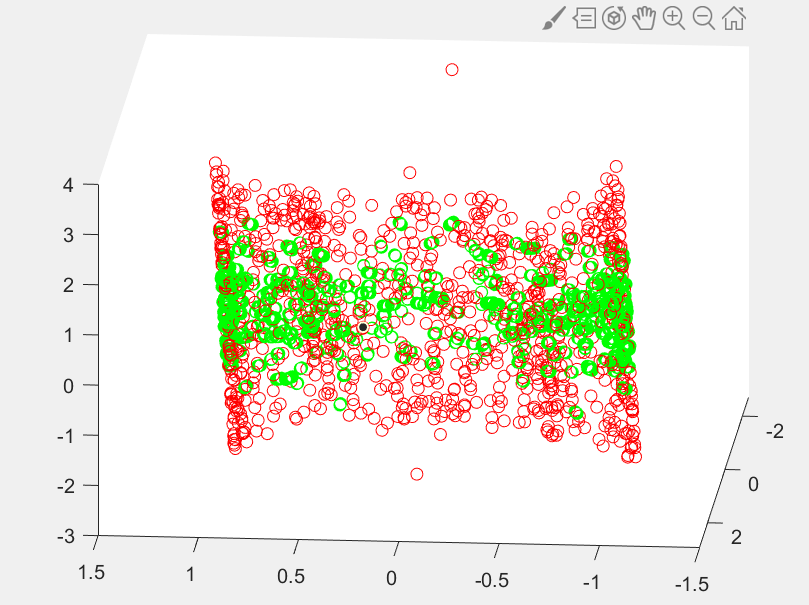
**如果在解牛顿法时使用threshold =2 时，得到如下图：**



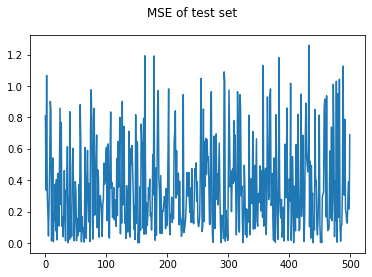
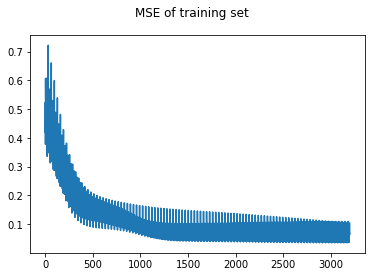
**也就是说平面的性质不变，只是产生了平移。**

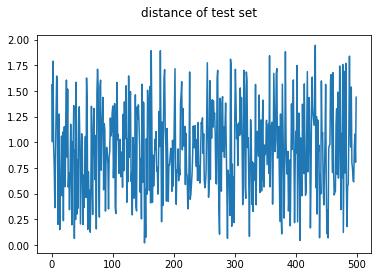
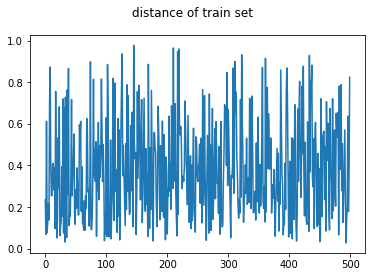
**将隐空间维度降为1维，得到结果图如下：**



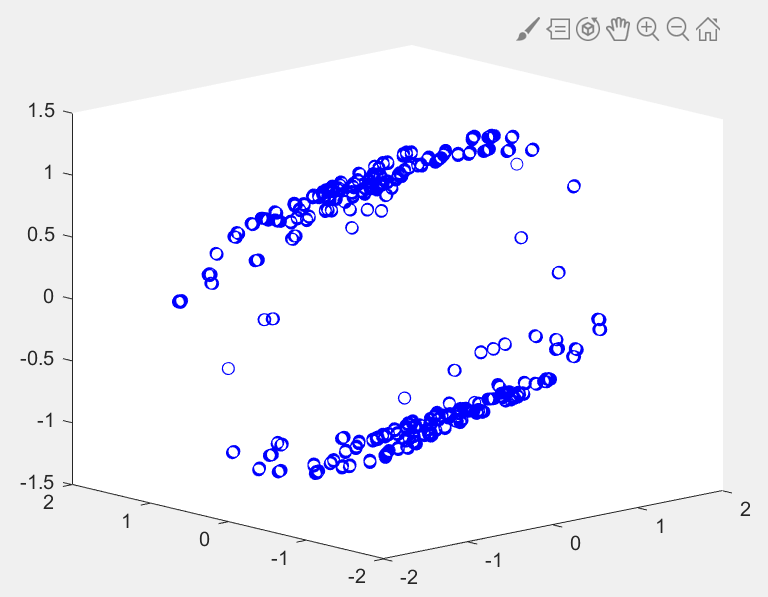


**（3）激活函数是反正切函数，Tanh() threshold= 1**

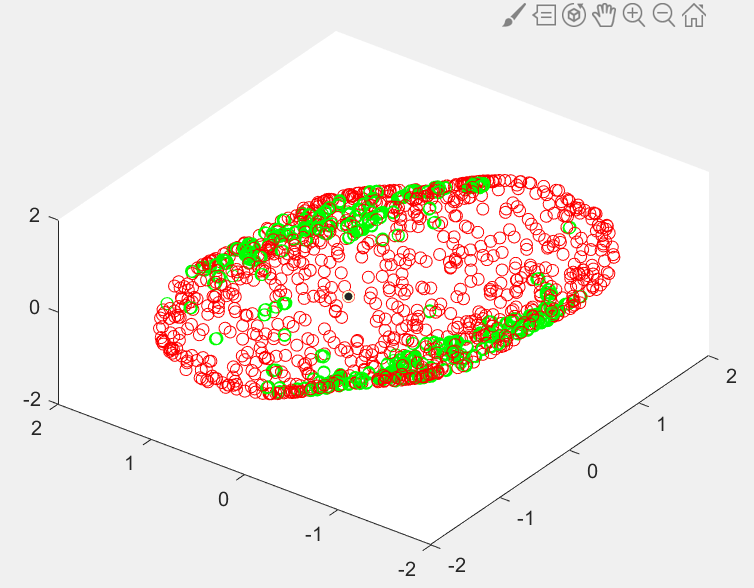
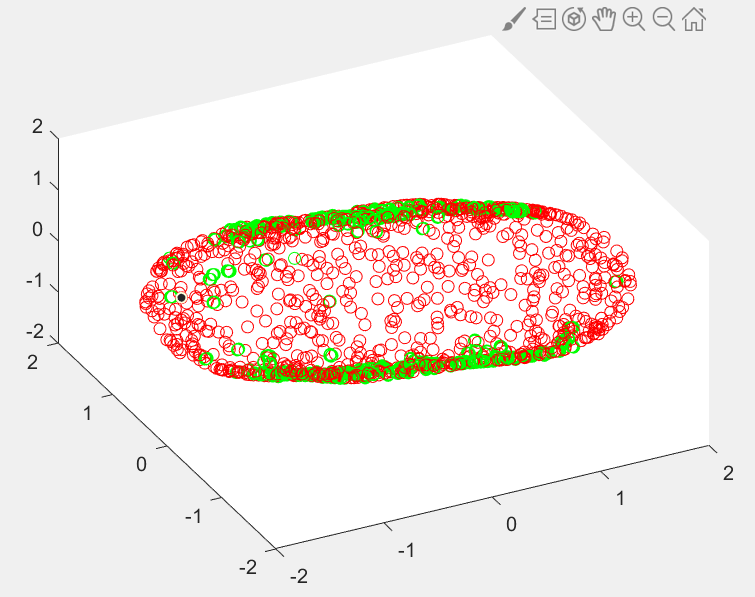


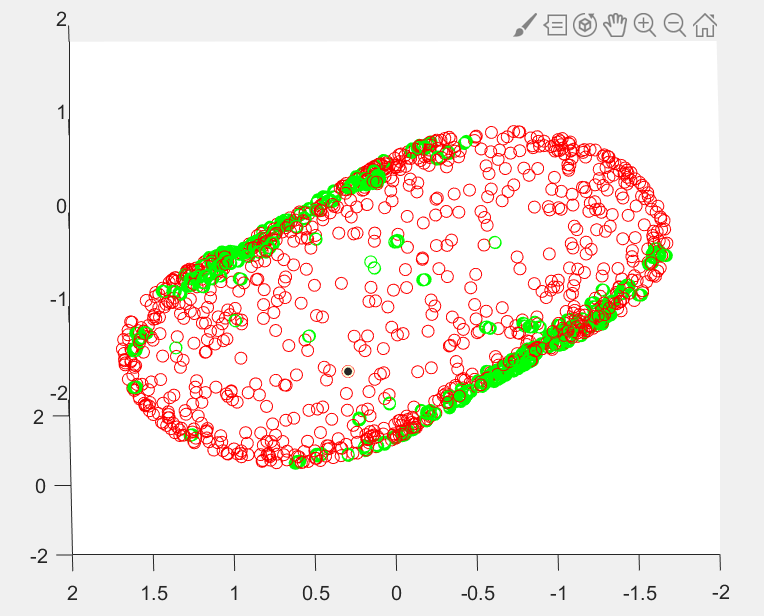


**遍历法：**

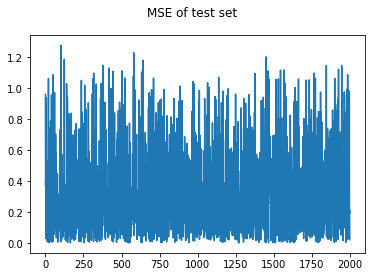
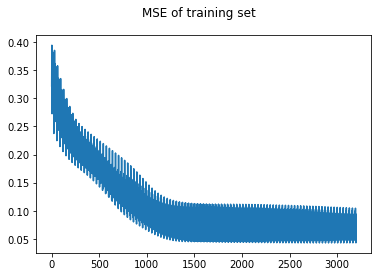


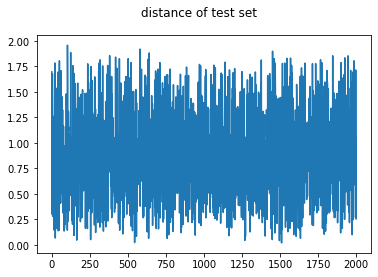
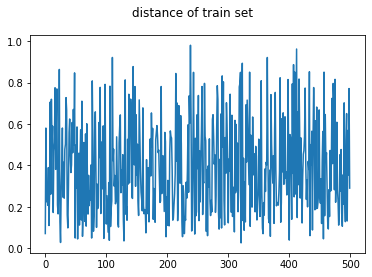
**牛顿法：**



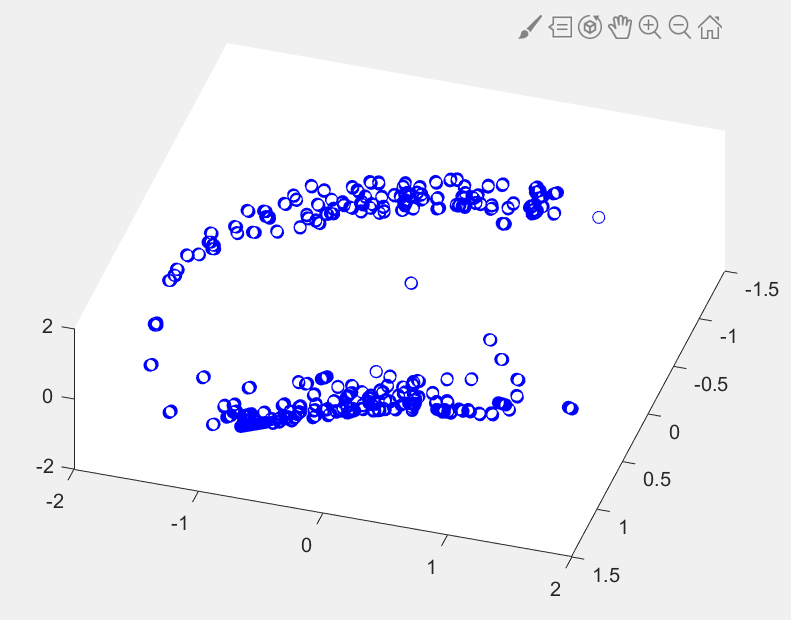
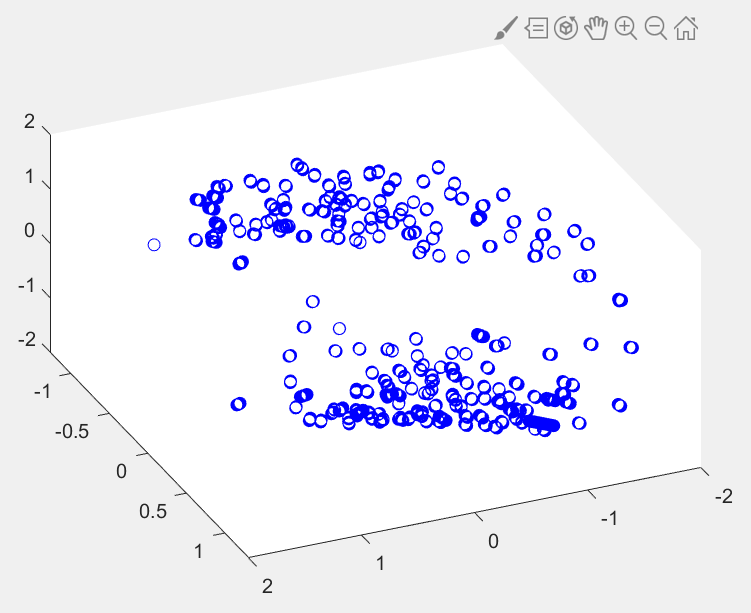


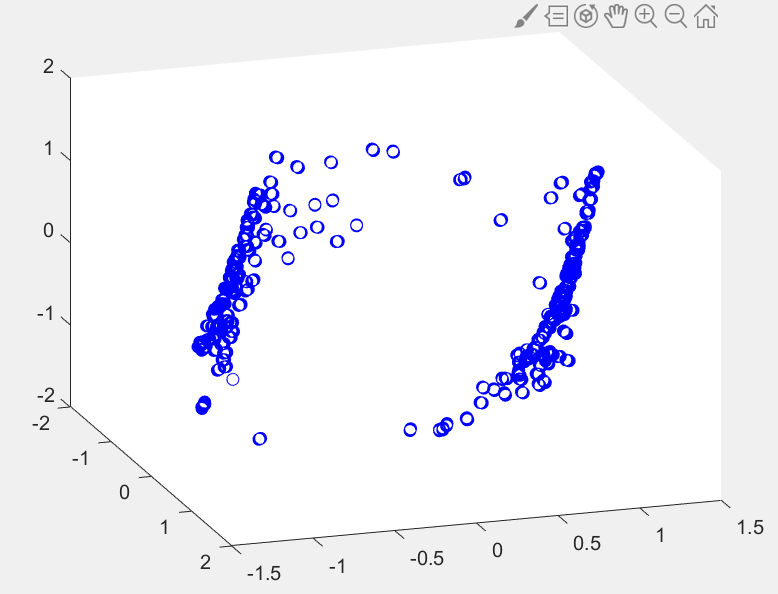
**(4) 激活函数是ELU(), threshold =1**



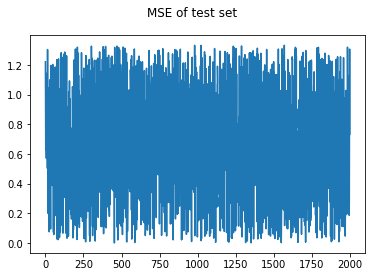
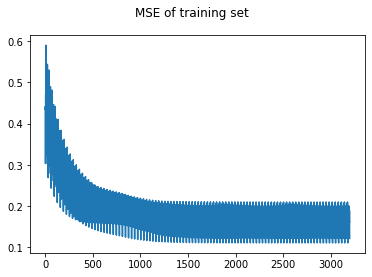


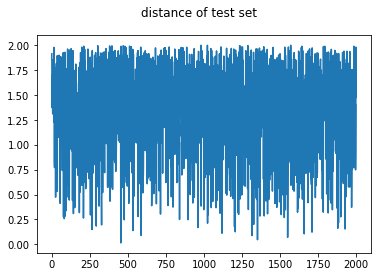
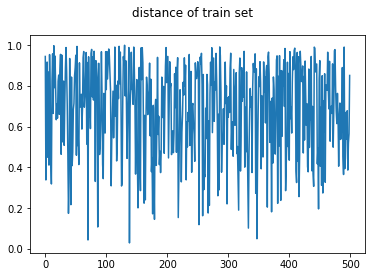
**遍历法：**



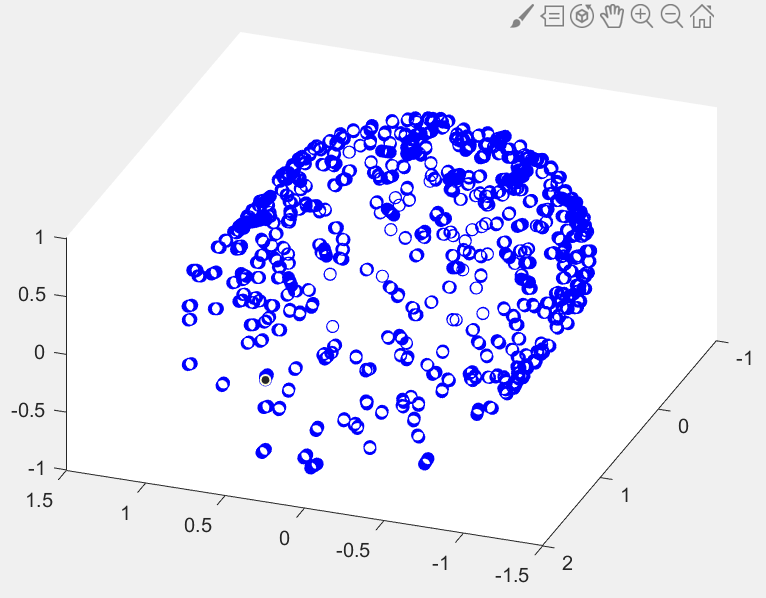
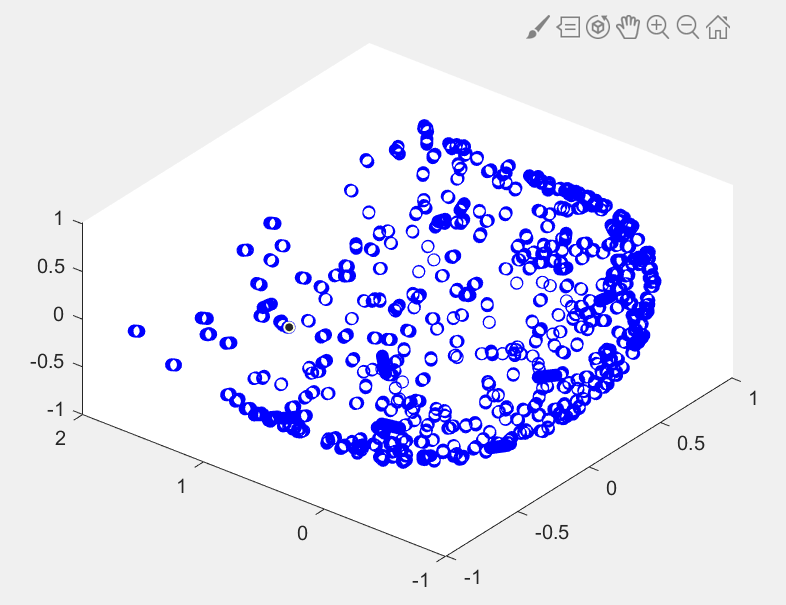


**(4) 激活函数是Relu(), threshold =1**





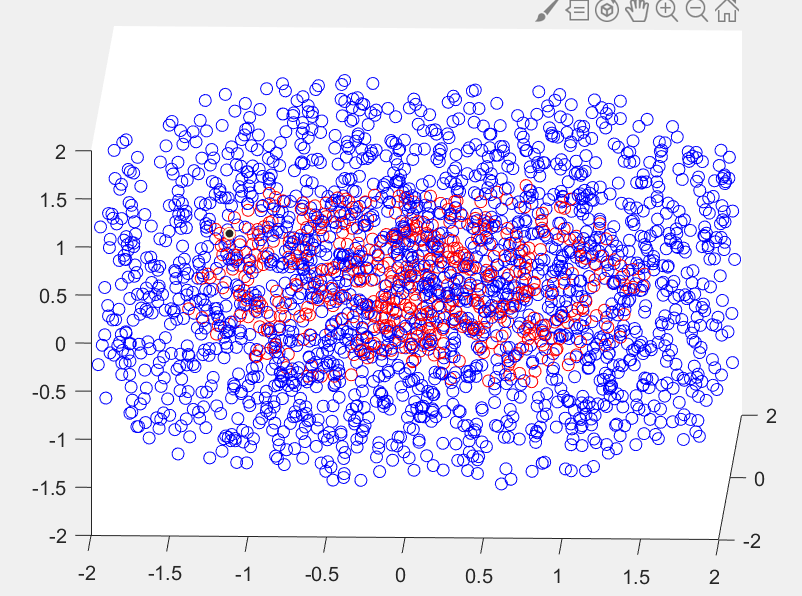
**遍历法：**



**再次尝试数据集：**

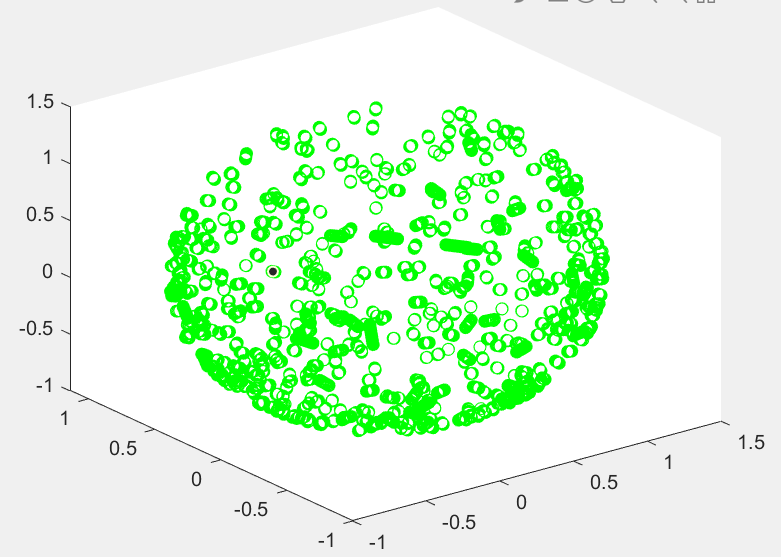
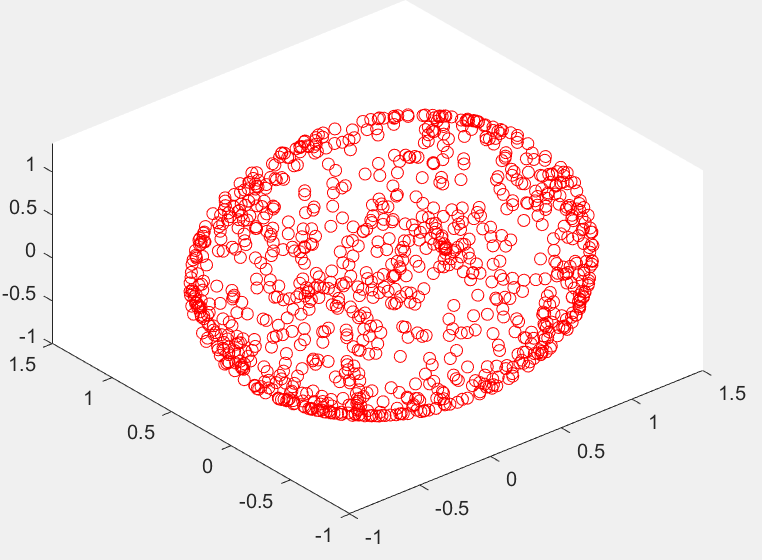
**构造数据集两个球互相嵌入。训练集中，两球的圆心分别为（-0.5,0,0），（0.5,0,0），两球半径均为1。如下图的红色点。测试集为范围在[-2,2]的立方体中的点。模型训练采用的激活函数为sigmoid().**

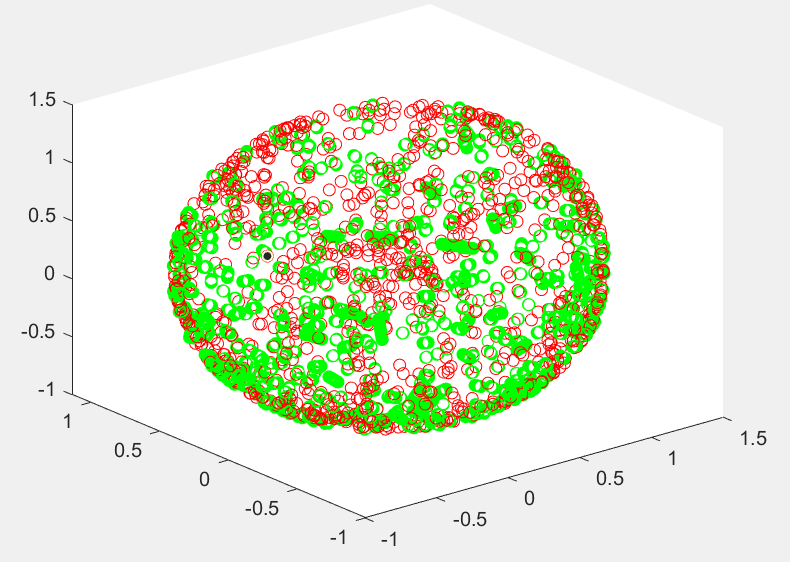
**如下图，红色点为训练集，蓝色点为测试集。**



**计算阈值边界点（threshold = 1）**

**如下图，红色点为牛顿法求解得到的点，蓝色点为遍历法求解得到的点**





**可以看出，如同单位球数据的结果，阈值边界点构成的曲面还是一个类似球面。**

## 结果总结

从目前的实验可以看出，自编码器本身是一个从自身到自身的映射，由于编码向量要么维度小于输入维度，要么是稀疏编码，这样依据信息瓶颈理论，它会损失一些输入的信息，从而重构的结果势必不会完全同输入一样。自编码器可以做异常检测可以从两个角度去解释：1. 函数梯度下降法求解最优值；2. 极大似然的思想。