

# Task 1

## ▼ Task 1

- 代码基本架构

- ▼ 实验

- ▼ 拟合 Sin

- 学习率对网络的影响
    - 正则项对网络的影响
    - 网络层数对网络的影响
    - 单层网络的神经元个数对网络的影响

- ▼ 图像分类

- 学习率对网络的影响
    - 正则项对网络的影响
    - 网络层数对网络的影响
    - 单层网络的神经元个数对网络的影响

- ▼ 对反向传播算法的理解

- 初始化参数的影响
  - 学习率的影响
  - 正则项的影响

## 代码基本架构

- layer. py : 代表的是神经网络的层, 包含了该层**线性变化**和**非线性变化 (可以设置为没有)**
- network. py : 根据输入的参数构建的神经网络, 包含backward forward方法
- sin
  - 文件夹都是数据
  - train\_network\_sin.py : 训练神经网络的程序
  - sin\_data\_model.py : 数据加载器
  - experiencescript. py : 实验脚本代码
  - trained\_network.py : 加载训练好的模型
- img
  - 文件夹都是数据
  - dataloader. py
  - imgdataset. py 以上为模仿pytorch 的dataset 和dataloader写的数据集处理机制
  - train\_network\_img.py : 训练神经网络的程序
  - experiencescript. py : 实验脚本代码
  - trained\_network=.py : 加载训练好的模型

# 实验

## 拟合 Sin

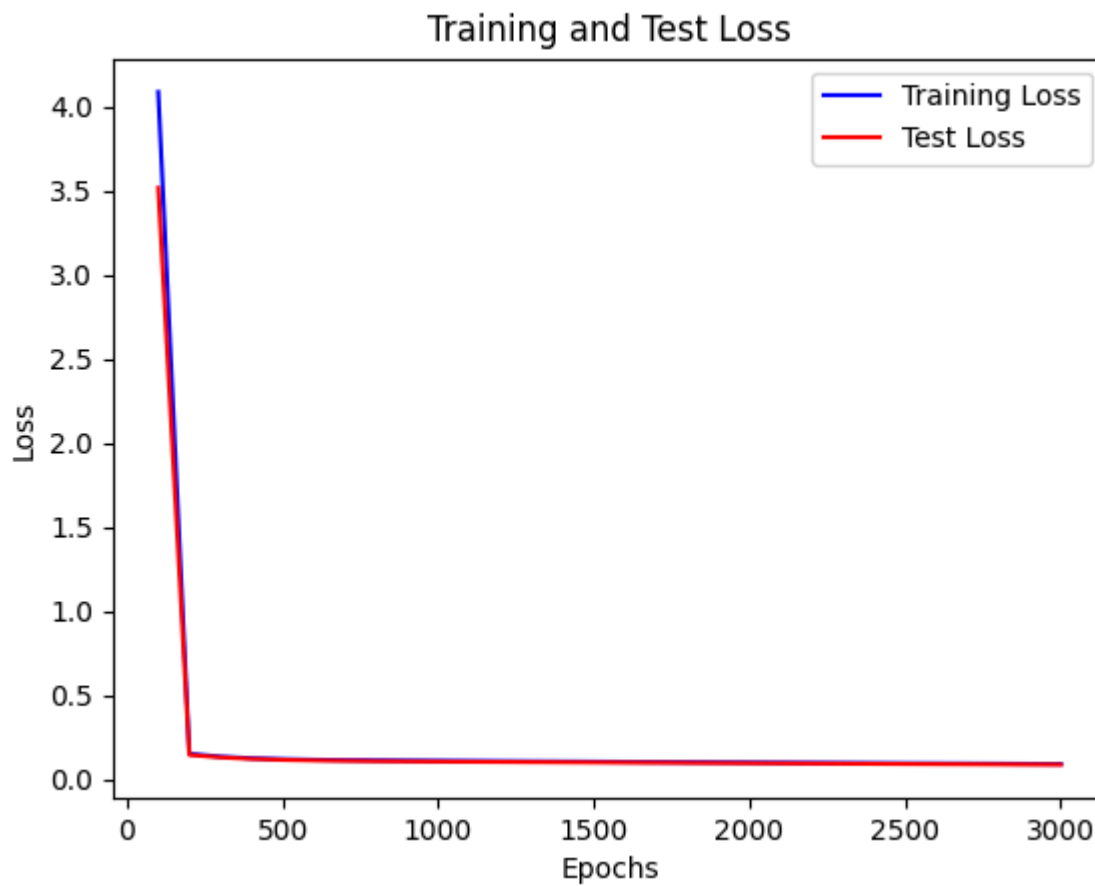
### 学习率对网络的影响

- 基于 单层 32unit sigmoid激活函数 实验
- 上图为lr为0.001 下图为 0.01 的情况

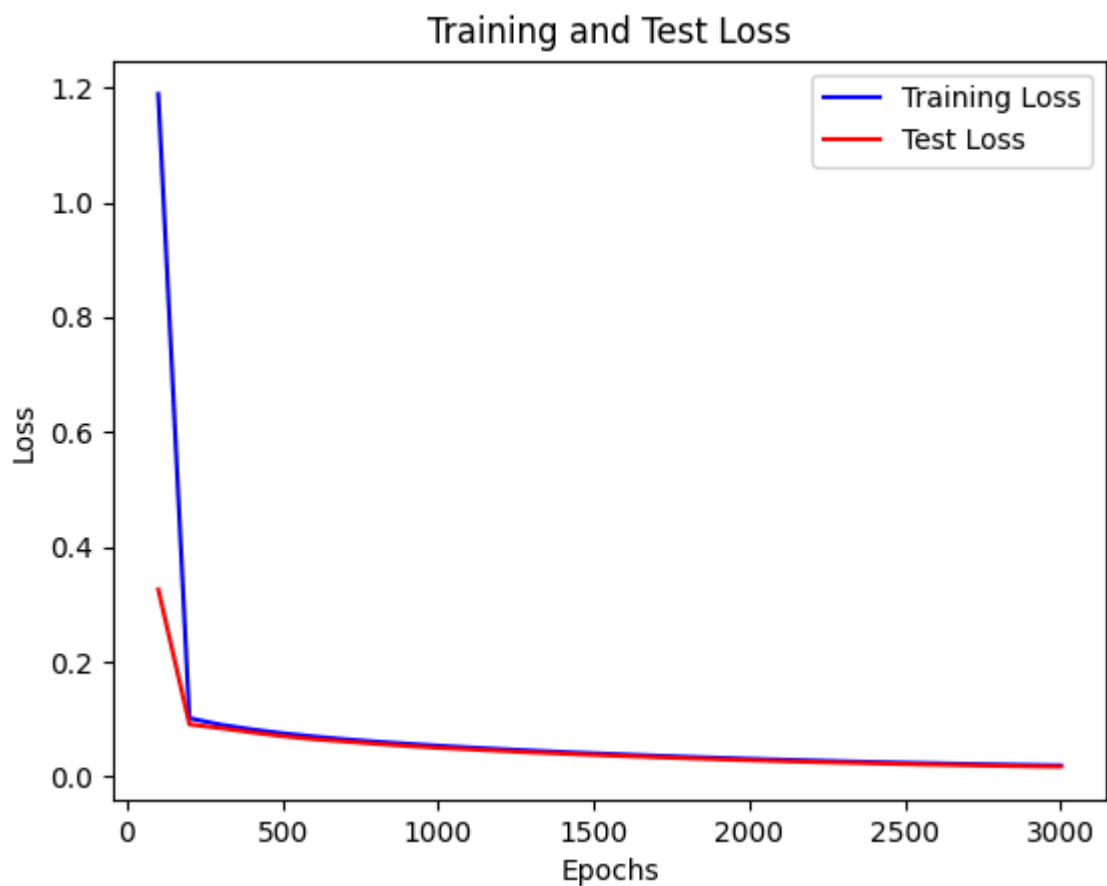
可见 对于sigmoid函数而言

lr太小的话，在部分epoch之后，loss会开始慢速下降，导致训练缓慢，所以需要较大的lr

图像

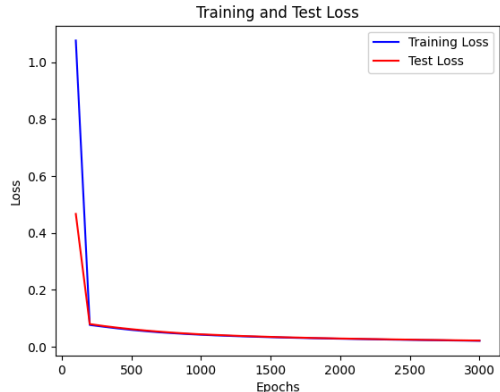


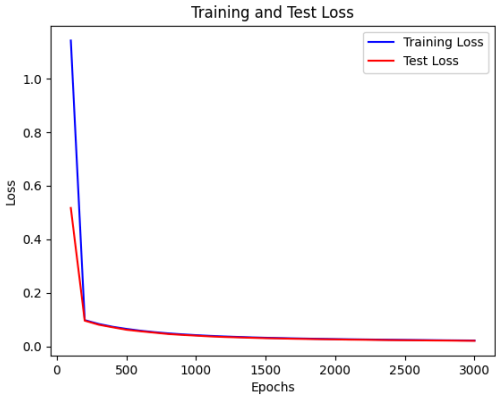
图像



### 正则项对网络的影响

- 下面的两个网络在相同初始化情况下进行迭代
- 可见 正则项会惩罚过大的梯度，导致下降较慢，同时会防止过拟合

网络	L1误差	图像
正则项系数为0	Epoch 2900, Loss: 0.0208701047369986	

网络	L1误差	图像
正则项系数为0.01	Epoch 2900, Loss: 0.022429724348479462	

网络层数对网络的影响

- 实验：对两个不同网络同时进行十次测试 得到结果 取平均
- 可见单层网络收敛速度较快
- 双层网络更容易在不改变学习率的情况下不收敛

网络	3000epoch后的L1误差
sigmoid 单层网络	0.014551067166471124
sigmoid双层网络	0.02297293823671869

单层网络的神经元个数对网络的影响

- 实验：对两个不同单层网络同时进行十次测试 得到结果 取平均
- 可见在拟合任务上 单层情况下 越多的神经元数量会使得拟合的更快 且更准确
- 由于是拟合问题 所以神经元过多导致的过拟合问题就不会显现

网络	3000epoch后的L1误差
32个神经元	0.014551067166471124
128个神经元	0.006307620398259217

图像分类

实验数据存储在 Task1\img\experiencedata\data.json

实验都采取早停策略

此处都使用relu做激活函数，所以当神经元数量增加或减少时，学习率需要调整

学习率对网络的影响

1024 单层网络Relu

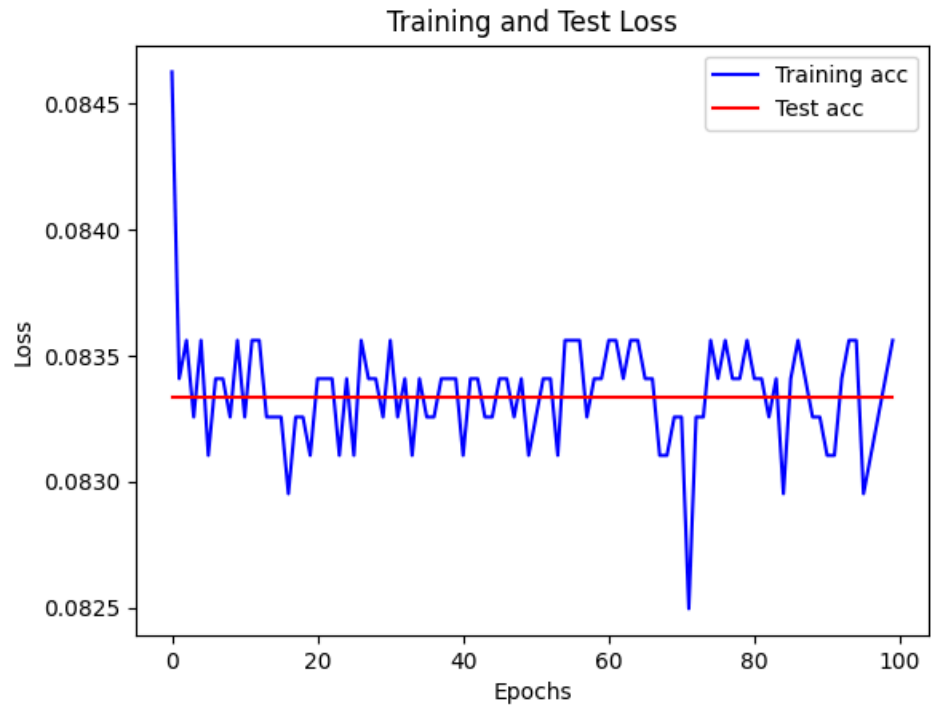
- 高的学习率任意使得网络在不断震荡难以收敛，甚至可能过快的将部分参数变为0，减缓参数的更新速率
- 更新学习率使得网络能够探索更优的解



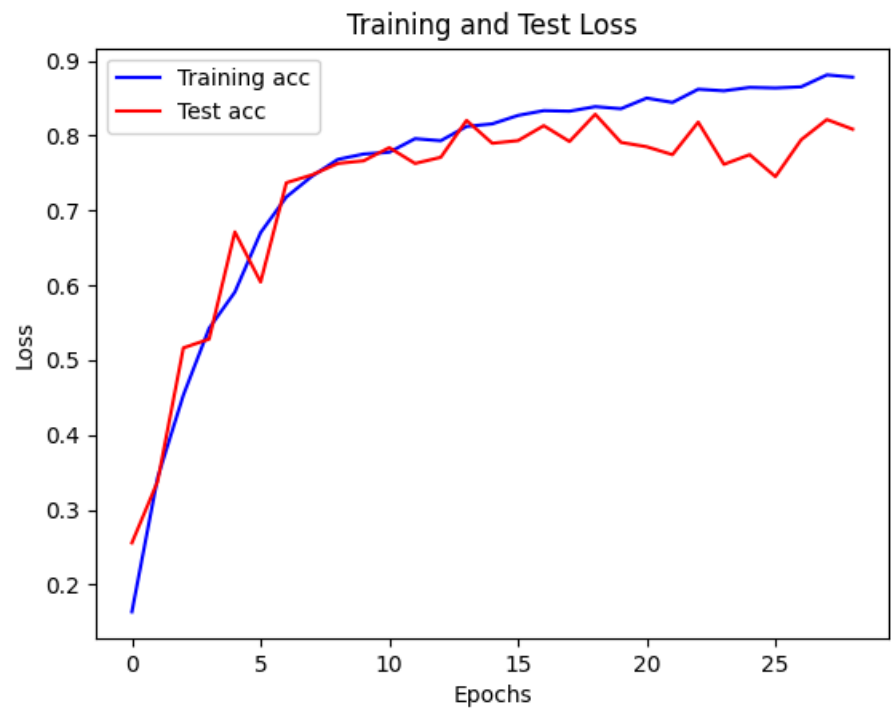
网络

图

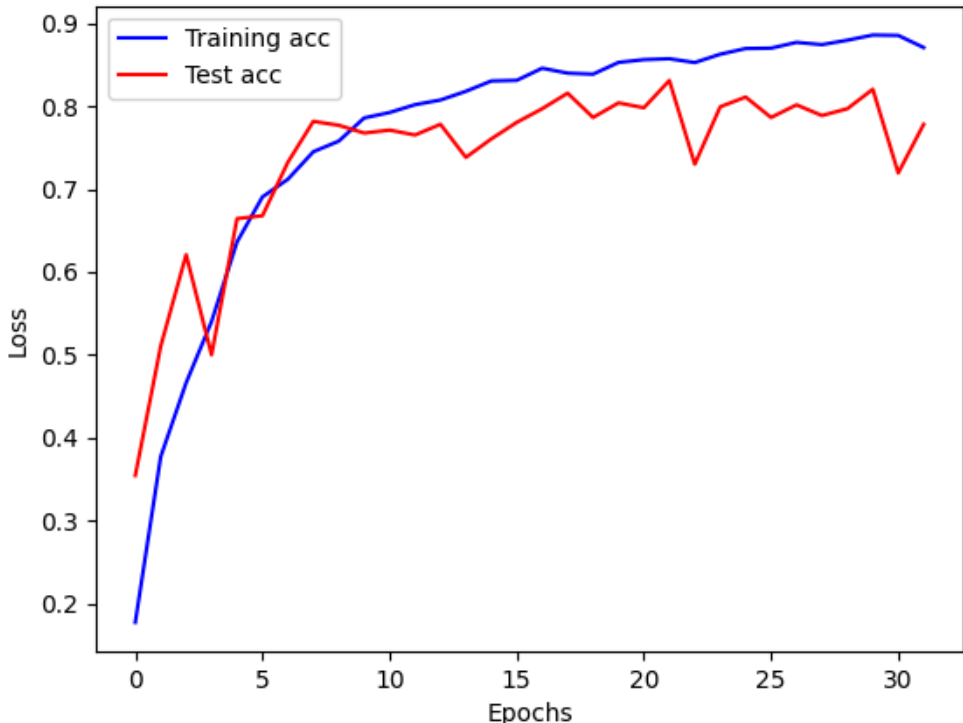
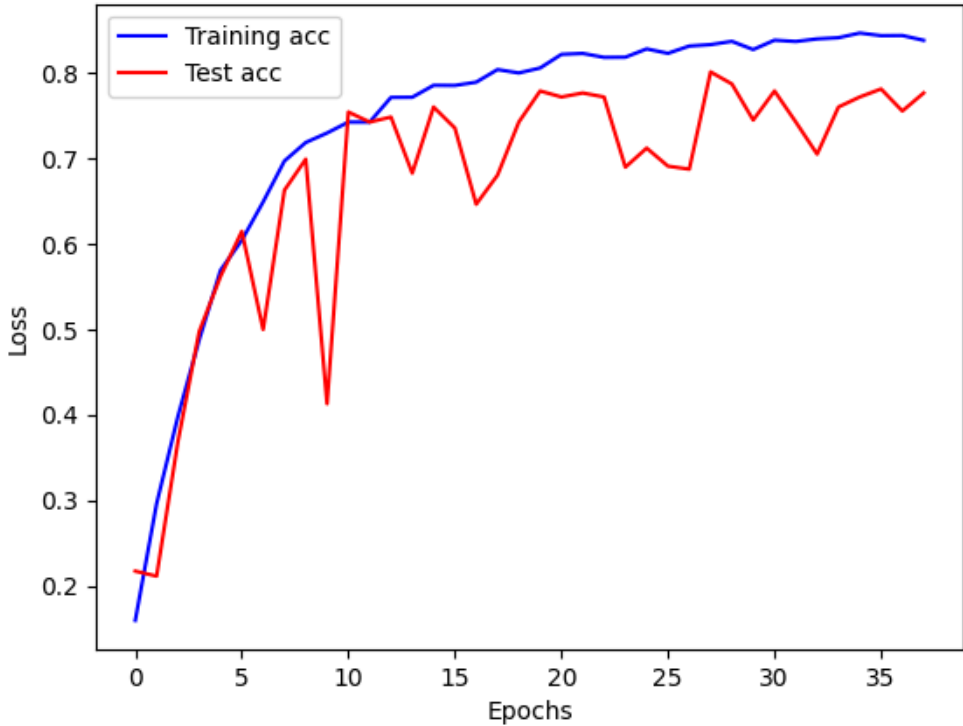
$lr = 0.001$



$lr = 0.0001$   
会更新学习率



正则项对网络的影响

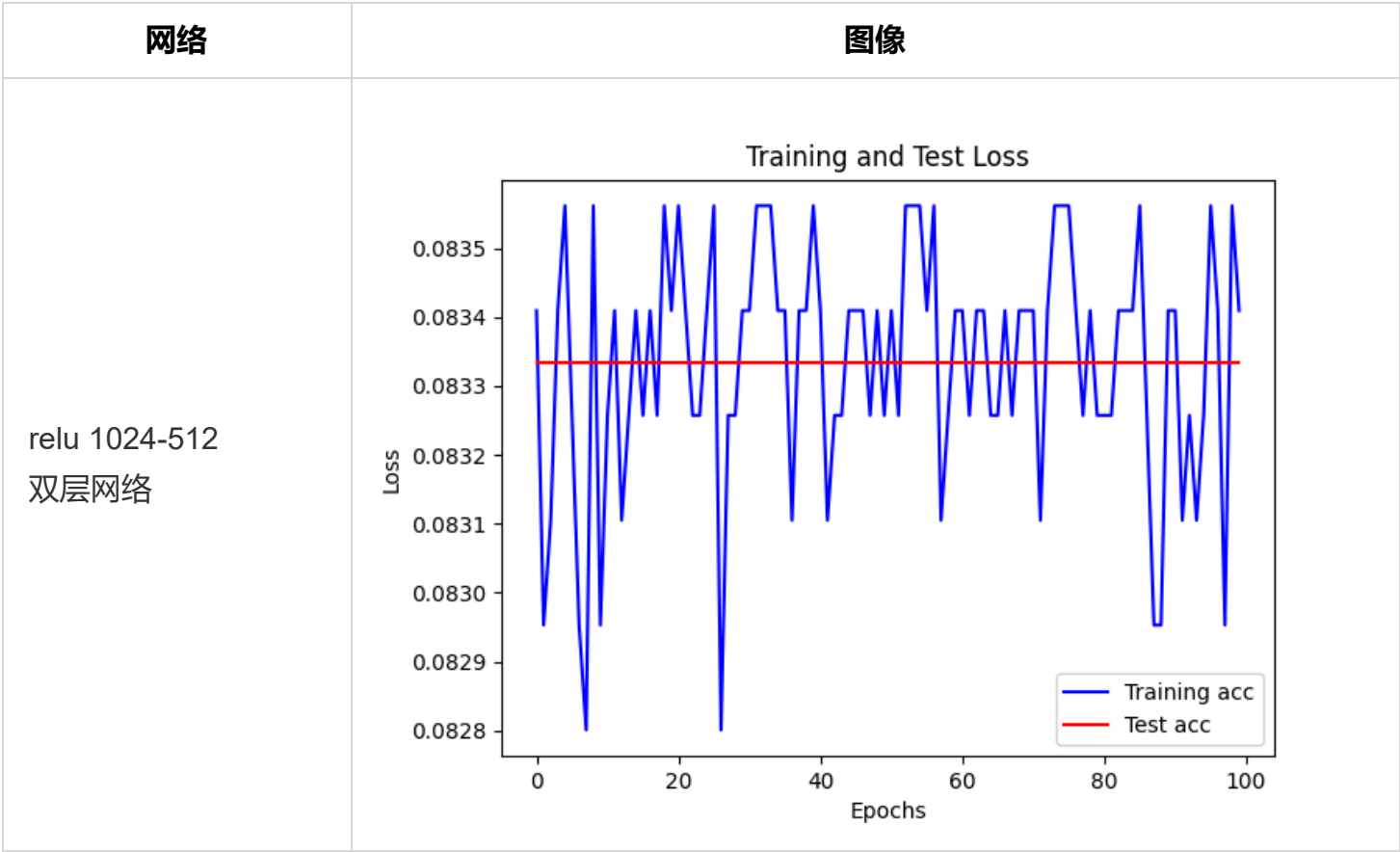
网络	图																											
lamda=0	<div><p>Training and Test Loss</p><table><caption>Approximate data for lamda=0 graph</caption><tr><th>Epochs</th><th>Training acc</th><th>Test acc</th></tr><tr><td>0</td><td>0.18</td><td>0.35</td></tr><tr><td>5</td><td>0.68</td><td>0.68</td></tr><tr><td>10</td><td>0.78</td><td>0.77</td></tr><tr><td>15</td><td>0.83</td><td>0.75</td></tr><tr><td>20</td><td>0.86</td><td>0.80</td></tr><tr><td>25</td><td>0.87</td><td>0.79</td></tr><tr><td>30</td><td>0.88</td><td>0.78</td></tr></table></div>	Epochs	Training acc	Test acc	0	0.18	0.35	5	0.68	0.68	10	0.78	0.77	15	0.83	0.75	20	0.86	0.80	25	0.87	0.79	30	0.88	0.78			
Epochs	Training acc	Test acc																										
0	0.18	0.35																										
5	0.68	0.68																										
10	0.78	0.77																										
15	0.83	0.75																										
20	0.86	0.80																										
25	0.87	0.79																										
30	0.88	0.78																										
lamda = 0.00001	<div><p>Training and Test Loss</p><table><caption>Approximate data for lamda = 0.00001 graph</caption><tr><th>Epochs</th><th>Training acc</th><th>Test acc</th></tr><tr><td>0</td><td>0.15</td><td>0.22</td></tr><tr><td>5</td><td>0.60</td><td>0.60</td></tr><tr><td>10</td><td>0.75</td><td>0.75</td></tr><tr><td>15</td><td>0.79</td><td>0.75</td></tr><tr><td>20</td><td>0.82</td><td>0.78</td></tr><tr><td>25</td><td>0.83</td><td>0.70</td></tr><tr><td>30</td><td>0.83</td><td>0.78</td></tr><tr><td>35</td><td>0.83</td><td>0.78</td></tr></table></div>	Epochs	Training acc	Test acc	0	0.15	0.22	5	0.60	0.60	10	0.75	0.75	15	0.79	0.75	20	0.82	0.78	25	0.83	0.70	30	0.83	0.78	35	0.83	0.78
Epochs	Training acc	Test acc																										
0	0.15	0.22																										
5	0.60	0.60																										
10	0.75	0.75																										
15	0.79	0.75																										
20	0.82	0.78																										
25	0.83	0.70																										
30	0.83	0.78																										
35	0.83	0.78																										

网络层数对网络的影响

- 实验：对两个不同网络同时进行测试 得到结果
- 可见单层网络收敛速度较快
- 双层网络更容易不收敛
- 此处引入早停的技术
  - 单层网络在30epoch内 未调整学习率 任意过早的早停
  - 双层网络则是在震荡
- 说明复杂任务中，更需要较为合适的参数


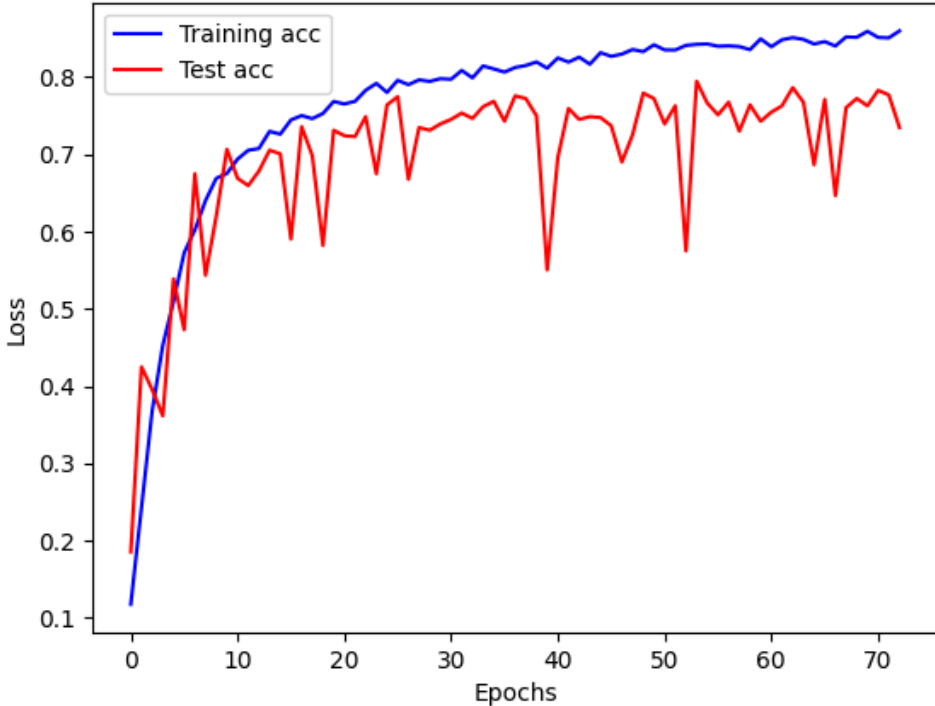
网络	图像																																																			
relu 1024单层网络	<div><p>Training and Test Loss</p><table><caption>Approximate data points from the Training and Test Loss graph</caption><tr><th>Epoch</th><th>Training Loss</th><th>Test Loss</th></tr><tr><td>0</td><td>0.18</td><td>0.35</td></tr><tr><td>2</td><td>0.45</td><td>0.62</td></tr><tr><td>4</td><td>0.65</td><td>0.67</td></tr><tr><td>6</td><td>0.75</td><td>0.78</td></tr><tr><td>8</td><td>0.78</td><td>0.77</td></tr><tr><td>10</td><td>0.80</td><td>0.77</td></tr><tr><td>12</td><td>0.81</td><td>0.75</td></tr><tr><td>14</td><td>0.83</td><td>0.78</td></tr><tr><td>16</td><td>0.84</td><td>0.81</td></tr><tr><td>18</td><td>0.84</td><td>0.79</td></tr><tr><td>20</td><td>0.85</td><td>0.80</td></tr><tr><td>22</td><td>0.85</td><td>0.73</td></tr><tr><td>24</td><td>0.87</td><td>0.81</td></tr><tr><td>26</td><td>0.88</td><td>0.79</td></tr><tr><td>28</td><td>0.88</td><td>0.80</td></tr><tr><td>30</td><td>0.88</td><td>0.72</td></tr></table></div>	Epoch	Training Loss	Test Loss	0	0.18	0.35	2	0.45	0.62	4	0.65	0.67	6	0.75	0.78	8	0.78	0.77	10	0.80	0.77	12	0.81	0.75	14	0.83	0.78	16	0.84	0.81	18	0.84	0.79	20	0.85	0.80	22	0.85	0.73	24	0.87	0.81	26	0.88	0.79	28	0.88	0.80	30	0.88	0.72
Epoch	Training Loss	Test Loss																																																		
0	0.18	0.35																																																		
2	0.45	0.62																																																		
4	0.65	0.67																																																		
6	0.75	0.78																																																		
8	0.78	0.77																																																		
10	0.80	0.77																																																		
12	0.81	0.75																																																		
14	0.83	0.78																																																		
16	0.84	0.81																																																		
18	0.84	0.79																																																		
20	0.85	0.80																																																		
22	0.85	0.73																																																		
24	0.87	0.81																																																		
26	0.88	0.79																																																		
28	0.88	0.80																																																		
30	0.88	0.72																																																		





单层网络的神经元个数对网络的影响

- 过多易过拟合，过少难拟合
- 1024个神经元较容易拟合，而512个神经元则容易欠拟合，到一定程度就开始震荡，过早收敛

网络	图像																											
relu 1024单层网络	 <p>Training and Test Loss</p> <p>Y-axis: Loss (0.2 to 0.9)</p> <p>X-axis: Epochs (0 to 30)</p> <p>Legend: Training acc (blue line), Test acc (red line)</p> <table><tr><th>Epochs</th><th>Training acc</th><th>Test acc</th></tr><tr><td>0</td><td>0.18</td><td>0.35</td></tr><tr><td>5</td><td>0.68</td><td>0.68</td></tr><tr><td>10</td><td>0.78</td><td>0.78</td></tr><tr><td>15</td><td>0.83</td><td>0.75</td></tr><tr><td>20</td><td>0.86</td><td>0.80</td></tr><tr><td>25</td><td>0.87</td><td>0.79</td></tr><tr><td>30</td><td>0.88</td><td>0.78</td></tr></table>	Epochs	Training acc	Test acc	0	0.18	0.35	5	0.68	0.68	10	0.78	0.78	15	0.83	0.75	20	0.86	0.80	25	0.87	0.79	30	0.88	0.78			
Epochs	Training acc	Test acc																										
0	0.18	0.35																										
5	0.68	0.68																										
10	0.78	0.78																										
15	0.83	0.75																										
20	0.86	0.80																										
25	0.87	0.79																										
30	0.88	0.78																										
relu 512 单层网络	 <p>Training and Test Loss</p> <p>Y-axis: Loss (0.1 to 0.8)</p> <p>X-axis: Epochs (0 to 70)</p> <p>Legend: Training acc (blue line), Test acc (red line)</p> <table><tr><th>Epochs</th><th>Training acc</th><th>Test acc</th></tr><tr><td>0</td><td>0.12</td><td>0.18</td></tr><tr><td>10</td><td>0.68</td><td>0.68</td></tr><tr><td>20</td><td>0.78</td><td>0.78</td></tr><tr><td>30</td><td>0.80</td><td>0.75</td></tr><tr><td>40</td><td>0.82</td><td>0.75</td></tr><tr><td>50</td><td>0.83</td><td>0.75</td></tr><tr><td>60</td><td>0.84</td><td>0.75</td></tr><tr><td>70</td><td>0.85</td><td>0.78</td></tr></table>	Epochs	Training acc	Test acc	0	0.12	0.18	10	0.68	0.68	20	0.78	0.78	30	0.80	0.75	40	0.82	0.75	50	0.83	0.75	60	0.84	0.75	70	0.85	0.78
Epochs	Training acc	Test acc																										
0	0.12	0.18																										
10	0.68	0.68																										
20	0.78	0.78																										
30	0.80	0.75																										
40	0.82	0.75																										
50	0.83	0.75																										
60	0.84	0.75																										
70	0.85	0.78																										

# 对反向传播算法的理解

## 初始化参数的影响

- 初始化参数，极大的影响了网络的拟合速度和可行性
  - 在使用relu时如果参数值过大，极容易发生溢出的情况，使得网络难以训练
  - 而合适的参数能够使得初始的位置比较好，可以快速拟合

## 学习率的影响

- 学习率与网络结构以及激活函数都相关
  - 对网络结构，越大的网络，初始学习率越小
  - 对激活函数，relu激活函数所需的学习率较小，若太大易溢出；sigmoid需要的较大，太小的话参数更新速度慢

## 正则项的影响

- 正则化是用来惩罚过大的参数变化的，避免参数直接转变到不好的状态，使得网络难以调整