

# Report

陈浩贤 18307110276

---

## 地址信息:

---

github repo: <https://github.com/chx7514/assignment-1-for-computer-vision>

模型网盘地址: <https://drive.google.com/drive/folders/1EdK-ofTN0teqOhajmMRh38bz39f6dIK6?usp=sharing>

## 实验设置:

---

激活函数使用 ReLU 和 Sigmoid

两层隐藏层均选择全连接层, 最后一层连接 Softmax 层, 再连接到输出

损失函数选择负对数似然损失

参数更新使用动量法, 其中系数为 0.9

$$\begin{aligned}v^{t+1} &= \beta_t v^t + \alpha_t \nabla f(w^t) \\w^{t+1} &= w^t - v^{t+1}\end{aligned}$$

学习率衰减使用指数衰减, 选择如下的公式:

$$lr = lr^0 \cdot 0.9^{iter/10000}$$

每次训练都训练 4 个 epoch

## 参数查找实验结果

---

### 实验一

---

首先设置查找的四个变量: 激活函数、学习率、正则化强度、隐藏层大小。为了减少训练的时间, 首先看一层隐藏层的情况。实验的具体设置如下, 一共需要测试 96 个模型。

```
1 lr = [1e-4, 5e-4, 1e-3]
2 penalty = [1e-6, 1e-5, 1e-4, 1e-3]
3 activation = ['ReLU', 'Sigmoid']
4 hidden_layers = [[150], [180], [200], [225]]
```

## 实验结果：

详细的结果见 output1.txt

分析结果，可以看出

- Sigmoid 激活函数比 ReLU 效果普遍要更好
- 其他参数相同时，隐藏层越大，通常测试集的准确率更高
- 较小的初始学习率会导致收敛速度太慢，效果更差

得到的最优模型信息如下：

Best accuracy in the test set: 97.14%

The model used: activation function: Sigmoid, learning rate: 0.001, penalty: 1e-06, hidden layers: [200]

## 实验二

---

根据实验一的结果和分析，重新对实验进行设置，使得能找到更好的模型

- 测试更大的学习率
- 只选择 Sigmoid 作为激活函数
- 使用两层的隐藏层，同时隐藏层的大小也增大

实验的具体设置如下，一共需要测试 80 个模型。

```
1 lr = [1e-3, 1e-2, 5e-2, 1e-1]
2 penalty = [1e-7, 1e-6, 1e-5, 1e-4]
3 activation = ['Sigmoid']
4 hidden_layers = [[128, 64], [150, 75], [200, 100], [256, 128], [300, 150]]
```

## 实验结果：

详细的结果见 output2.txt

分析结果，可以看出

- 学习率过高会使得训练效果变差，选择更快的学习率下降策略可能会有改善
- 使用两层的隐藏层时，隐藏层越大，测试集的准确率越高
- 正则化强度不用选择较大的值

得到的最优模型信息如下：

Best accuracy in the test set: 97.93%

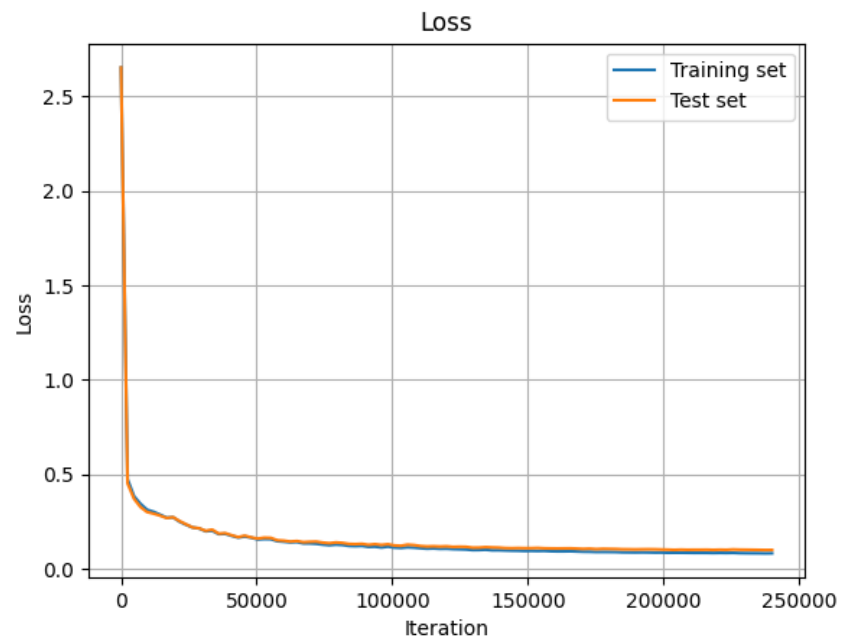
The model used: activation function: Sigmoid, learning rate: 0.01, penalty: 1e-07, hidden layers: [256, 128]

## 可视化

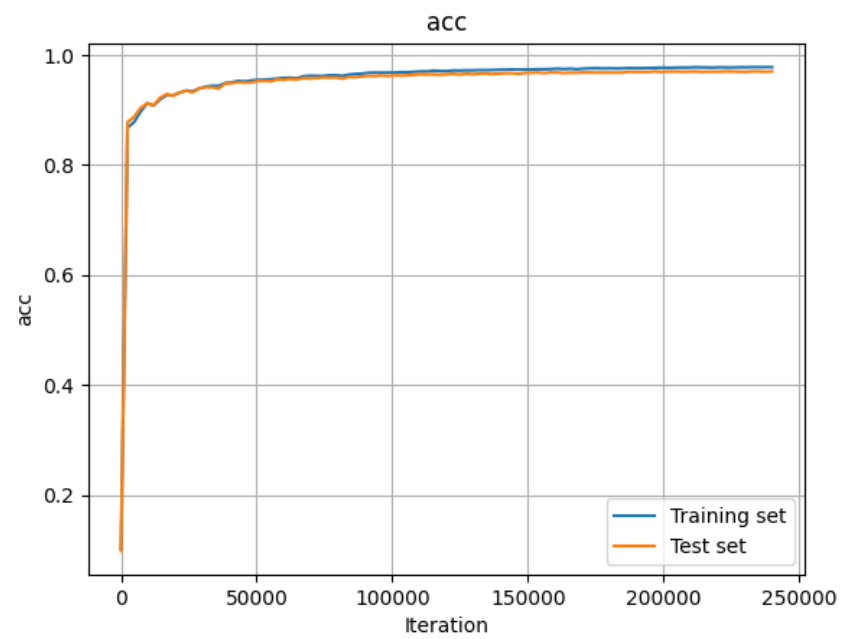
分别可视化了以上两个实验的最佳模型

---

## 模型一 Loss 曲线

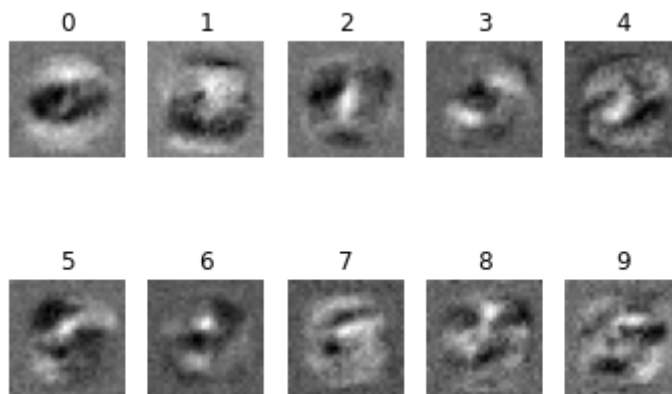


## 模型一 Acc 曲线

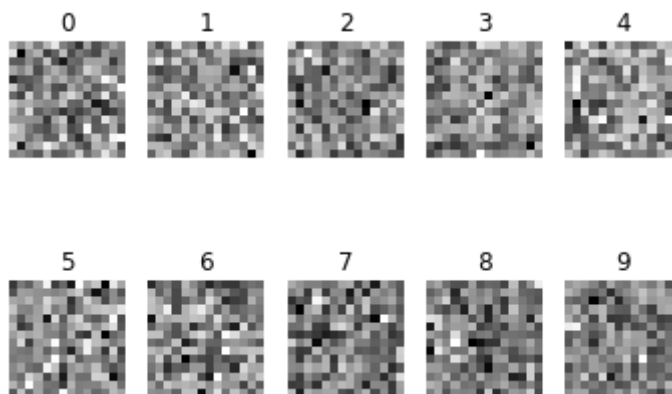


## 模型一网络参数可视化

$W_1$ :



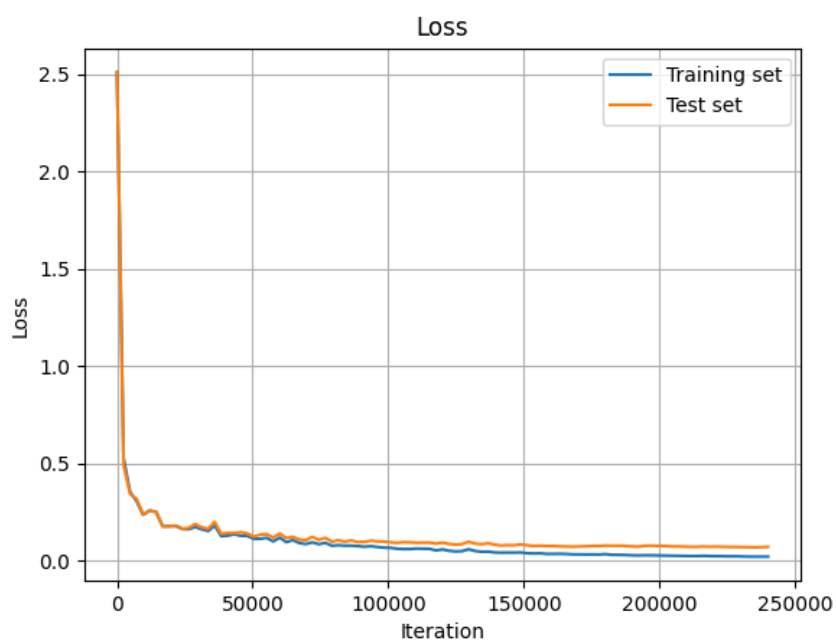
$W_2$ :



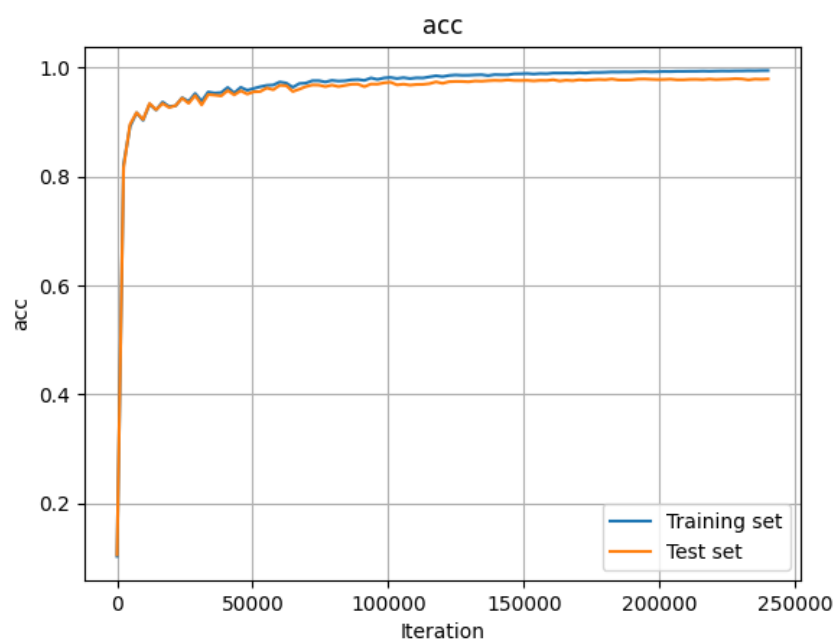
通过对模型一的可视化，有如下发现：

- Loss 曲线出现波动，并非一直下降，原因是使用了动量法，每次参数更新会吸取上一步的经验
- 测试集和训练集之间的 Loss 和 Acc 差不大，说明模型的泛化性能较好
- 对  $W_1$  的可视化可以看出，主要的特征提取在图像的中间，并且从可视化图片中能看出一些数字的大致形状
- 对  $W_2$  的可视化则不能看出什么特征，难以解释

## 模型二 Loss 曲线

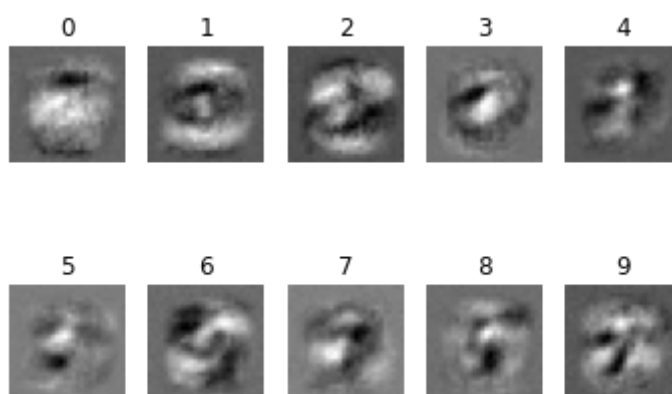


## 模型二 Acc 曲线

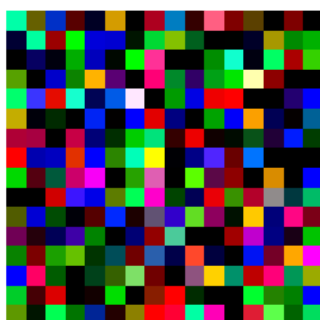


## 模型二网络参数可视化

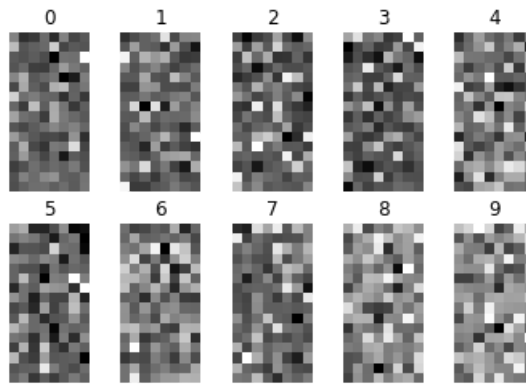
$W_1$ :



$W_2$ :



$W_3$ :



通过对模型一的可视化，有如下发现：

- 相对于单层隐藏层，两层隐藏层的训练曲线中，训练集和测试集之间的 gap 更大，说明模型更复杂的时候，泛化能力会变差，更容易出现过拟合
- 对  $W_1$  的可视化可以看出，主要的特征提取在图像的中间，并且从可视化图片中能看出一些数字的大致形状
- 对  $W_2$ 、 $W_3$  的可视化则不能看出什么特征，难以解释