# 分类与回归

崔雨豪

**摘　要** 分类与回归是机器学习中的基本问题。本文共做了三个实验：1.房价的回归问题。利用线性回归算法对房价进行预测。2.Mnist 分类。利用多项式回归进行分类。3.乳腺癌检测。利用感知机对乳腺癌数据进行分类，讨论过拟合问题，解决方案利用 L2 或 PCA 方法。

关键词 感知机；回归；分类；机器学习

## Image Classification Based On KNN

**Abstract** Classification and regression are fundamental issues in machine learning. This paper has done three experiments: 1. The return of housing prices. The house price is predicted using a linear regression algorithm. 2. Mnist classification. Classification is performed using polynomial regression. 3. Breast cancer testing. The breast cancer data was classified using a perceptron and the fitting problem was discussed，use L2 Regularization or PCA to solve the problem.

**Key words** Perceptron; Regression; Classification; Machine Learning

## 1　引言

分类和回归是机器学习的基本问题，都是对现有的数据分布进行拟合。分类一般是对伯努利分布或是多项式分布进行拟合，而回归则是对高斯分布进行拟合。 感知机模型是神经网络的基础模型，利用仿射函数来提高模型容量，利用激活函数来增加非线性能力。

## 2　本文算法

### 2.1　感知机

感知机算法可以描述为一个三层的线性模型：输入层，隐含层，输出层。隐含层作为特征提取层，对高维数据可以进行降维，对低维数据可以进行充分学习。

$$:t = f(\sum_{i=1}^{n} w_i x_i + b) = f(\mathbf{w}^T \mathbf{x}) \tag{2.1}$$

### 2.2 正则化

正则化是防止过拟合或加大稀疏度的有效手段分为 L1 正则化和 L2 正则化。

L1 正则化：

$$C = C_0 + \frac{\lambda}{n} \sum_{w} |w|. \tag{2.2}$$

L1 更新规则：

$$w \to w' = w - \frac{\eta\lambda}{n}\mathrm{sgn}(w) - \eta\frac{\partial C_0}{\partial w}, \tag{2.3}$$

L2 正则化：

$$C = C_0 + \frac{\lambda}{2n}\sum_w w^2, \tag{2.4}$$

L2 更新规则：

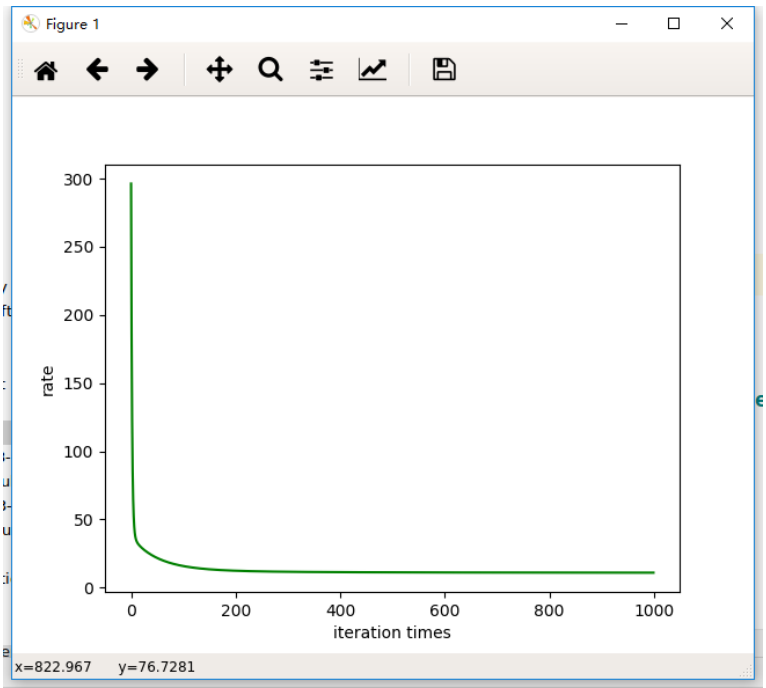
$$\begin{aligned} w \to & w - \eta\frac{\partial C_0}{\partial w} - \frac{\eta\lambda}{n}w \\ = & \left(1 - \frac{\eta\lambda}{n}\right)w - \eta\frac{\partial C_0}{\partial w}. \end{aligned} \tag{2.5}$$

# 3 实验

实验环境：Ubuntu16.04+Python3.6.5

## 3.1 Price Prediction

3.1.1 loss 下降曲线




## 3.2 Softmax Regression On Mnist

3.2.1 对比试验变量

变量 1：Hidden层Activation Function：Sigmoid，Leaky Relu

变量 2：Learning Rate：0.01，0.001，0.0001

变量 3：Hidden层隐含节点数量：512，128，32
变量 4：Batch Size：10，50，100
3.2.2 Activation Function
不变量：Batch Size=50，Hidden层=512，Weight Initialization=0 均值 1 方差高斯分布/100. Learning Rate=0.001，Iteration=10000

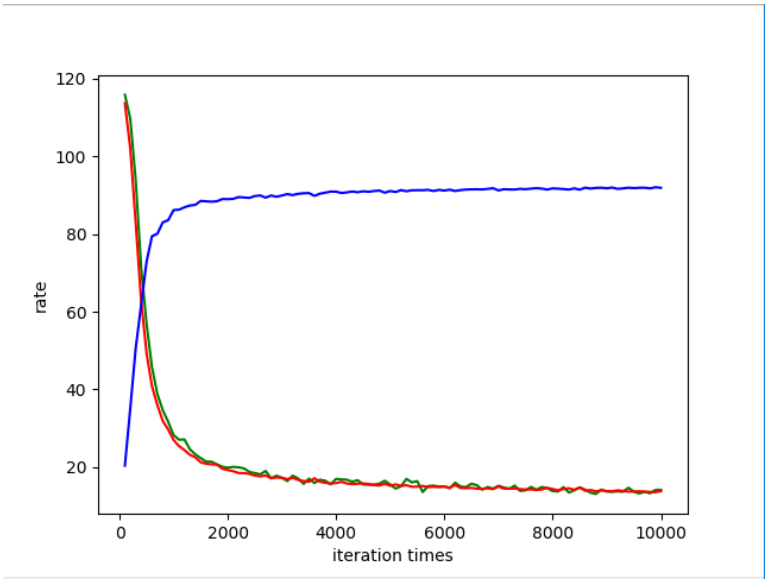



图 3.1 Sigmoid结果

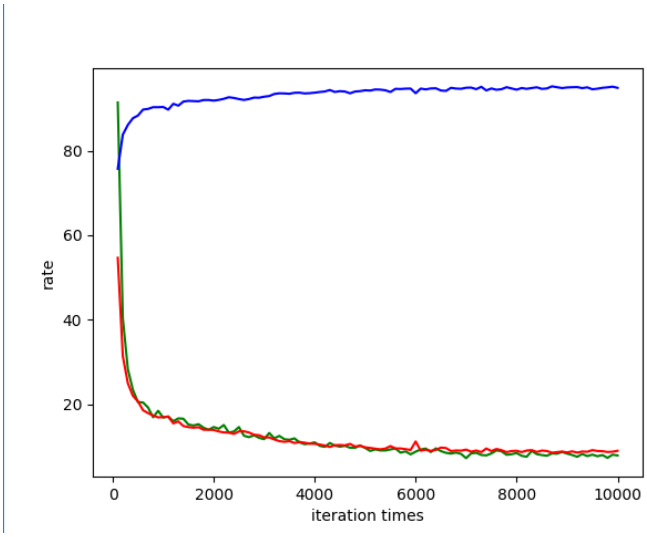



图 3.2 LRelu结果

| Method | ACC | Train Loss | Validation Loss |
|---|---|---|---|
| Sigmoid | 0.9189 | 14.15 | 13.83 |
| LRelu | 0.9486 | 7.88 | 8.98 |

LRelu由于本身梯度较大在 10000 次迭代中收敛速度远高于Sigmoid

3.2.3 Batch Size
不变量：Activation Function =Sigmoid，Hidden层=512，Weight Initialization=0 均值 1 方差高斯分布/100. Learning Rate=0.001，Iteration=10000

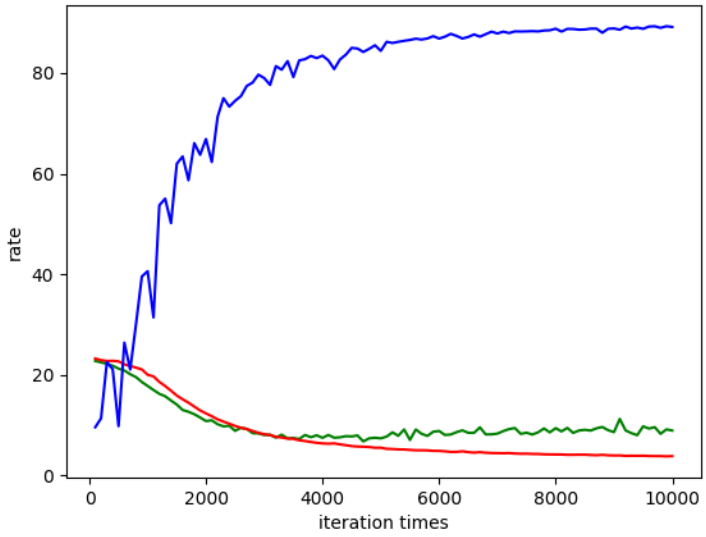



图 3.3 Batch Size=10 结果

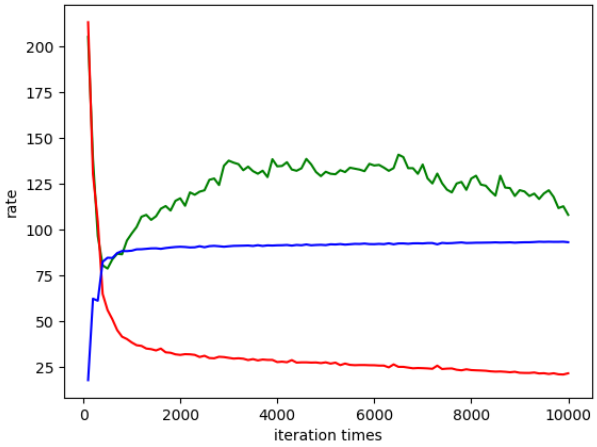



图 3.4 Batch Size=100 结果

Loss由于批次不同，Loss和不同，只看ACC曲线，明显 100 比 10 更加平滑，收敛更快。

| Batch Size | ACC |
|---|---|
| 10 | 0.905 |
| 50 | 0.918 |
| 100 | 0.933 |

3.2.4 Hidden层隐含节点数量

不变量：Batch Size=50，Activation Function =Sigmoid，Weight Initialization=0 均值 1 方差高斯分布/100.Learning Rate=0.001，Iteration=10000

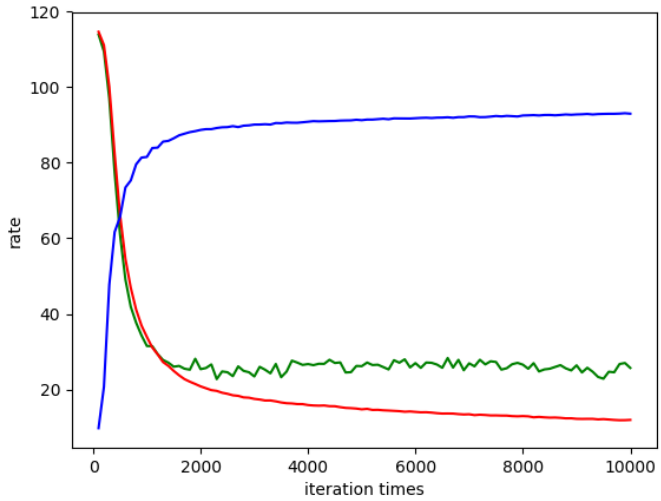



图 3.5 Hidden层=128 结果

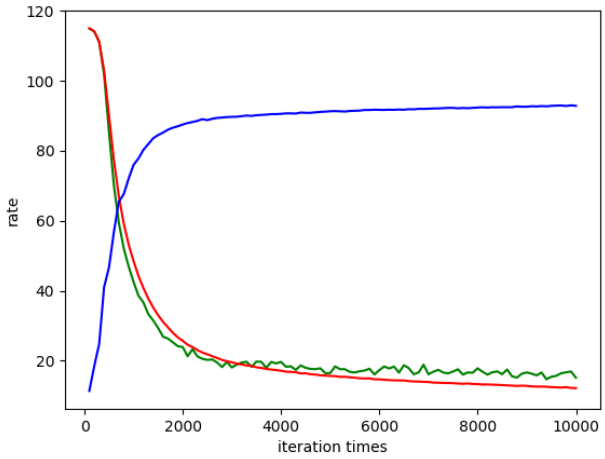



图 3.6 Hidden层=32 结果

| Hidden | ACC | Train Loss | Validation Loss |
|---|---|---|---|
| 32 | 0.9283 | 15.14 | 12.15 |
| 128 | 0.9296 | 25.73 | 11.89 |
| 512 | 0.9189 | 14.15 | 13.83 |

Hidden层并不是越多越好，越多TrainLoss可以下降很低但是在Validation上并不能有很好的效果，反而是 128 时虽然TrainLoss很大但是Validation上结果很不错

3.2.5 Learning Rate

不变量：Batch Size=50，Activation Function =Sigmoid，Weight Initialization=0 均值 1 方差高斯分布/100.Hidden Size=512，Iteration=10000

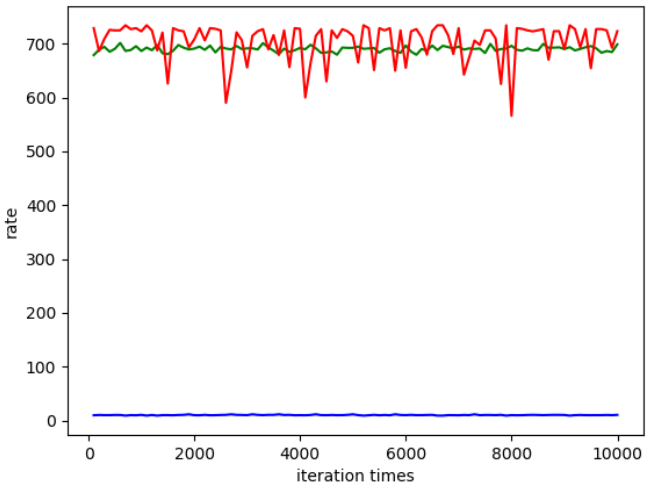



图 3.7 lr=0.01 结果

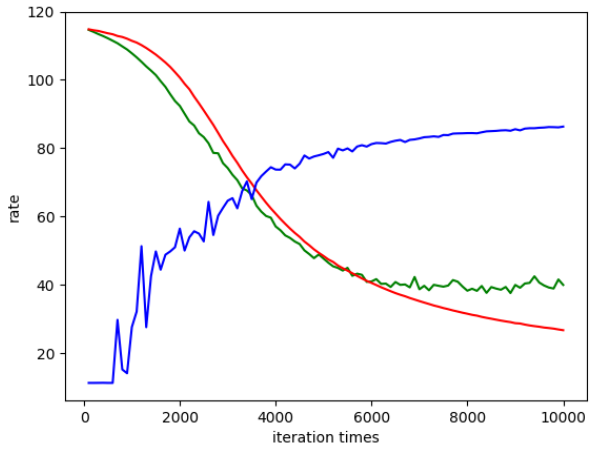



图 3.8 lr=0.0001 结果

从图中可以看出学习率小时训练曲线非常平滑，大时会波动甚至无法训练

| Lr | ACC | Train Loss | Validation Loss |
|---|---|---|---|
| 0.01 | 0.1028 | 678.65 | 726.76 |
| 0.001 | 0.9189 | 14.15 | 13.83 |
| 0.0001 | 0.8633 | 40.01 | 26.77 |

学习率在适中时才有更快的收敛

## 3.3 Softmax Regression On Breast Cancer

### 3.3.1 实验结果

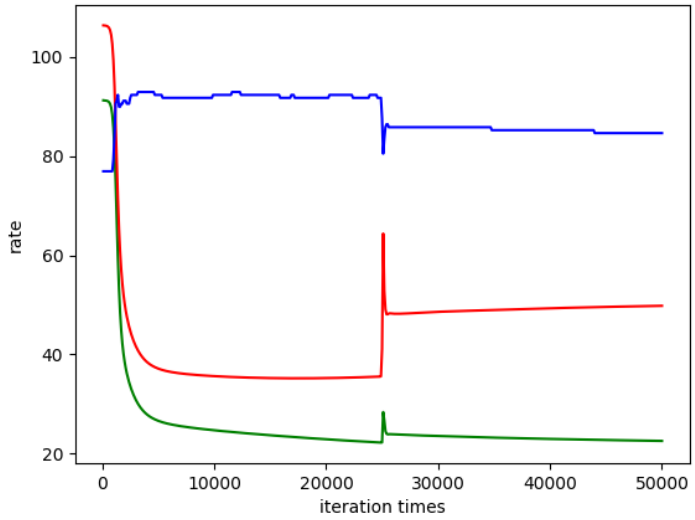



图 3.9 随机试验结果 1

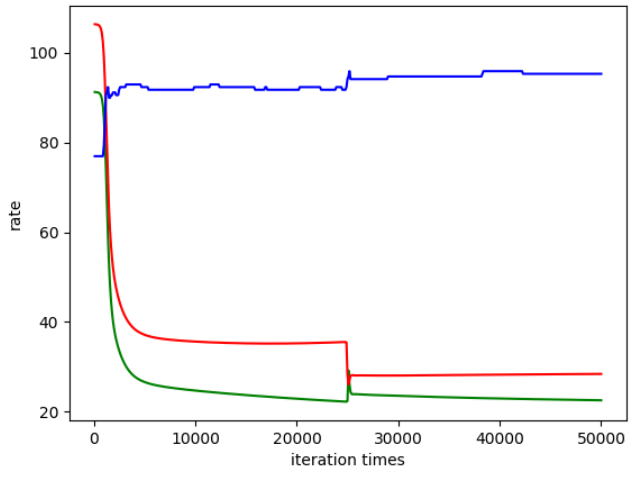



图 3.10 随机试验结果 2

两次具有代表性的随机试验，1 完全过拟合，可以发现虽然TrainLoss在下降但是ValidationLoss已经在上升且ACC突然变低。2 虽然ValidationLoss在后续趋势在上升，但是总体ACC一直在变大即虽然过拟合但并没有陷入局部最小值

| Result | ACC | Train Loss | Validation Loss |
|---|---|---|---|
| 1 | 0.84 | 22.58 | 49.81 |
| 2 | 0.95 | 22.56 | 28.43 |

### 3.3.2 L2 正则化

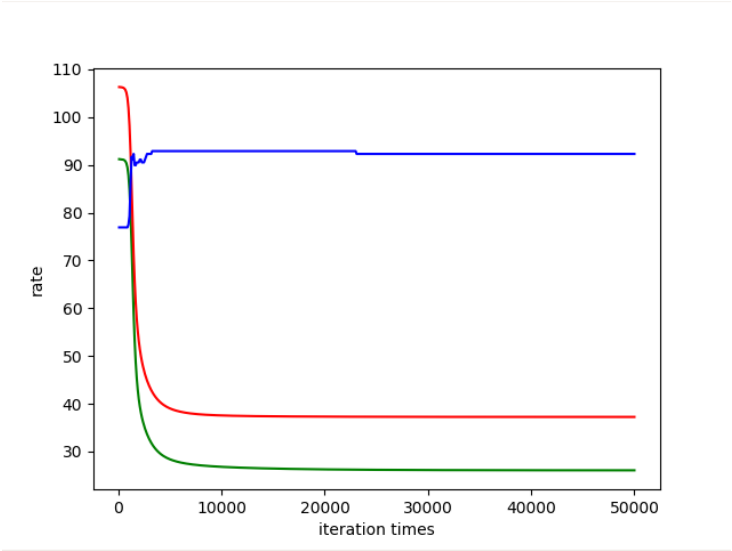



图 3.11 比例为 0.0001

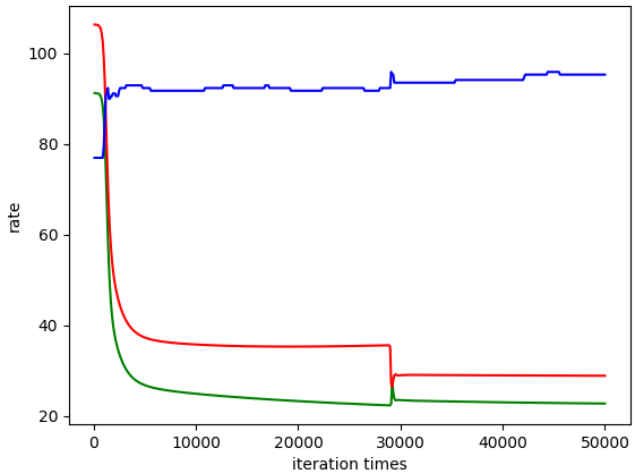



图 3.12 比例为 0.00001

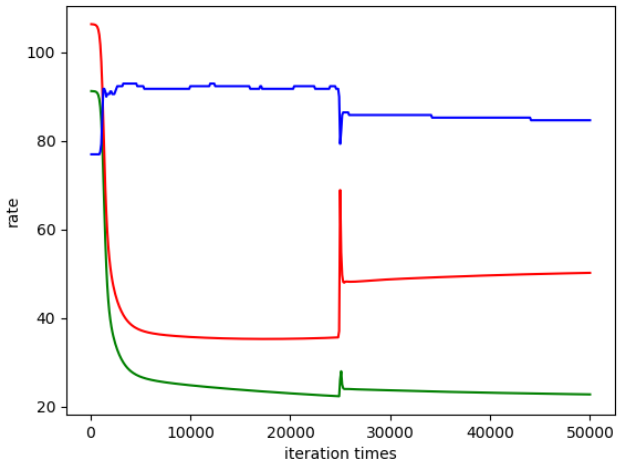



图 3.13 比例为 0.0000001

| Ratio | ACC | Train Loss | Validation Loss |
|---|---|---|---|
| 0.0001 | 0.923 | 26.05 | 37.23 |
| 0.00001 | 0.952 | 22.72 | 28.85 |
| 0.0000001 | 0.846 | 22.69 | 50.14 |

当Ratio大时难以训练，当Ratio小时正则化不起作用，只有设置合适才能有较好的效果。

### 3.3.3 PCA

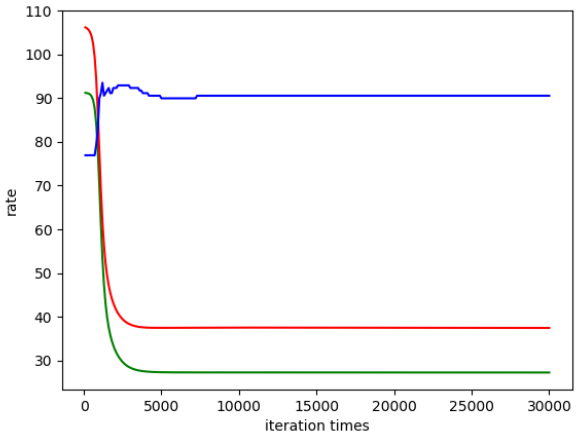



图 3.14 PCA取前 2 维

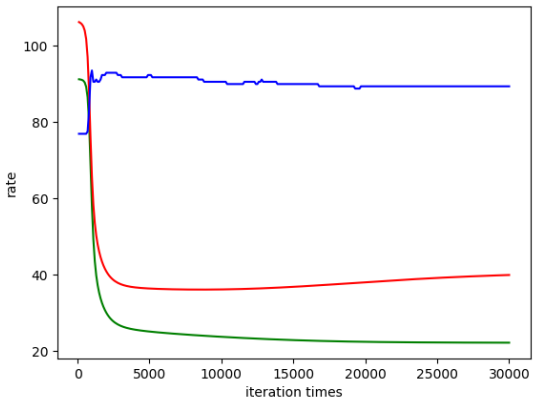



图 3.15 PCA取前 5 维

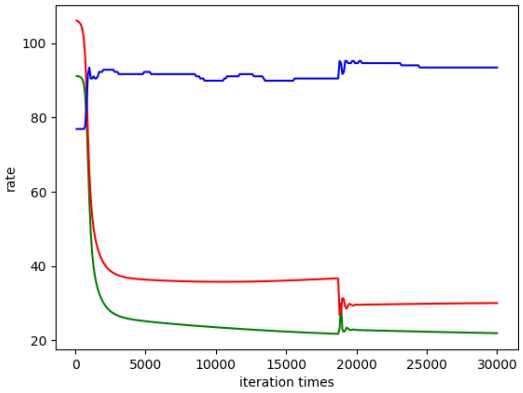



图 3.16 PCA取前 10 维

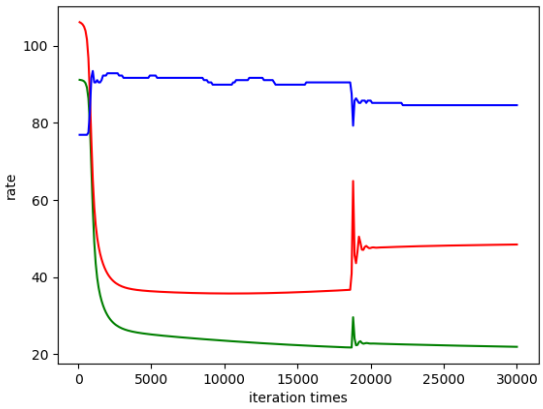



图 3.17 PCA取前 15 维

| PCA | ACC | Train Loss | Validation Loss |
|---|---|---|---|
| 2 | 0.905 | 27.27 | 37.45 |
| 5 | 0.89 | 22.25 | 39.96 |
| 10 | 0.93 | 21.88 | 29.99 |
| 15 | 0.84 | 21.88 | 48.44 |

PCA在取前 15 的时候，已经过拟合了，当取 10 效果比较好，2，5 时特征丢失过多，效果不明显。