

# 使用多层感知机解决多分类问题

1120191286      周彦翔

- 1 实验目的
- 2 多层感知机模型
- 3 数据集划分
- 4 初始化方法
- 5 损失函数
- 6 优化方法
- 7 反向传播计算梯度
- 8 训练过程
  - 8.1 0.001
  - 8.2 0.0001
  - 8.3 0.00001
- 9 总结

## 1 实验目的

参考多层感知机建模和训练方法，利用交叉熵、梯度下降训练模型，实现给定样本分类。

训练数据为150个带噪声的标记数据。

## 2 多层感知机模型

使用的多层感知机的结构为：

- 输入层，4个神经元
- 隐层1，7个神经元
- 隐层2，15个神经元
- 输出层，3个神经元

其中隐层和输出层均为全连接层，激活函数均为ReLU函数，输出层的输出最后经过Softmax函数进行归一化。

## 3 数据集划分

训练数据总计150个，样本量偏小，预测训练结果可能比较差。

在这里设计了两种数据集划分的方式，一种是简单地将所有样本分成训练集和测试集两部分，另一种则是使用K折交叉验证的方式。

第一种方法中，将样本的前80%划作训练集，后20%划作测试集。由于 $150 \times 20\% = 30$ ，测试集样本太小，所以测试得到的结果会有很大的随机性，个人认为难以作为评判的标准。

第二种方法中，将所有样本均匀分成K份，每个epoch中取其中K-1份作为训练集，1份作为测试集，统计K轮epoch的测试结果求平均作为最终测试结果，相对准确性会高一些。这也是当样本量很小的时候的解决办法。

## 4 初始化方法

因为采用ReLU作为激活函数，所以采用He初始化方法，即初始化的参数都是采样于 $\mu = 0, \sigma = \frac{2}{N}$ ,  $N$ 为输入神经元个数的正态分布。

## 5 损失函数

损失函数采用交叉熵。

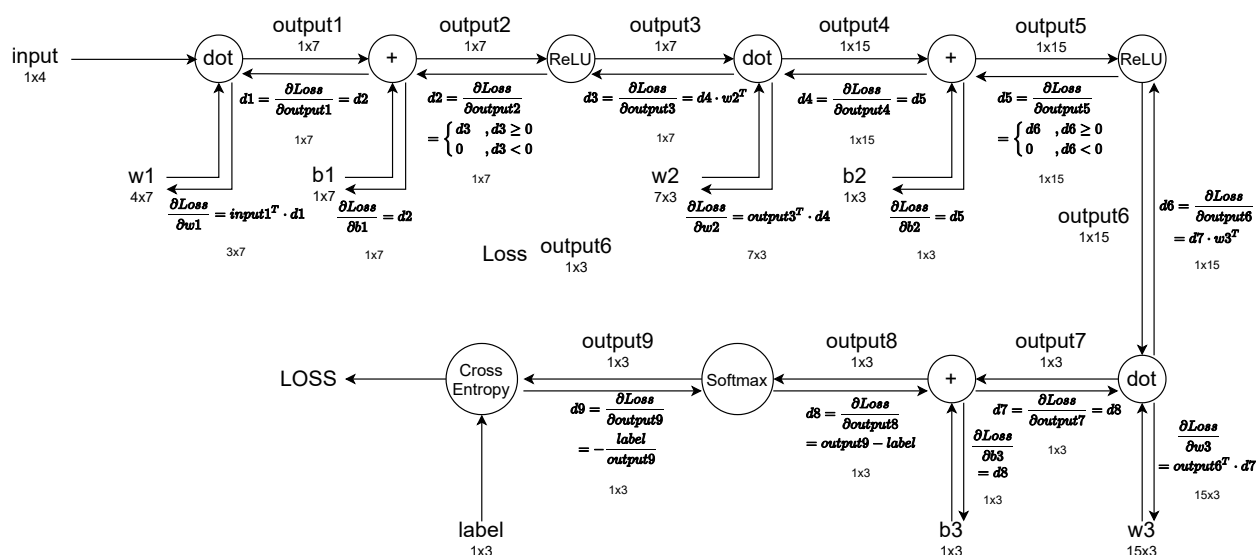
$$CrossEntropy(p, q) = - \sum_{i=1}^n p(i) \log q(i)$$

由于真实结果是one-hot形式的，所以交叉熵可以简化为 $-\log output[j]$ ， $j$ 为真实类别的编号。比如输出为 $[0.7 \ 0.1 \ 0.2]$ ，真实分布为 $[1 \ 0 \ 0]$ ，则交叉熵为 $-\log 0.7$ 。

## 6 优化方法

预先编写随机梯度下降、动量法、改进后的自适应梯度法RMSprop、Adam法，届时比较各个方法之间的优劣。

## 7 反向传播计算梯度



根据计算图计算反向传播后交叉熵损失对各个分量的导数，如上图所示。

## 8 训练过程

分别选择0.001，0.0001和0.00001作为学习率，并分别使用随机梯度下降、动量法、RMSprop法和Adam法进行优化，查看在不使用K折交叉验证的情况下的训练情况。

### 8.1 0.001

在学习率设为0.001且使用SGD的情况下，有时训练在2到3个epoch就进入过拟合，有时经过50到100个epoch后训练集上的损失也一直在下降，受不同初始化影响较大。

当使用其他方法的时候，训练会更早地进入过拟合。

无论使用什么优化方法，训练次数一多训练集和测试集上的损失和正确率都会开始振荡，说明此时学习率太大。

### 8.2 0.0001

当学习率设为0.0001时，使用SGD会在50轮训练内进入过拟合，使用其他优化方法时会更早地进入过拟合。

### 8.3 0.00001

SGD在200到300个epoch后才陷入过拟合，但是动量法和Adam法依然能在20到50个epoch内进入过拟合。RMSprop稍好些，会在80到200个epoch内进入过拟合。

## 9 总结

经过大量实验后得出以下结论：

1. 由于样本量太小，所以训练结果多数都很差，多数正确率都只在0.2到0.5之间，只有少数的几次训练中正确率能达到0.6到0.7，极少数能达到0.8以上。
2. 动量法、RMSprop、Adam都会比SGD更快地到达过拟合。其中动量法到达过拟合速度最快，无论lr设的有多小，它总能在10个epoch内进入过拟合；Adam法稍慢于动量法，RMSprop慢于Adam法。就本实验来看RMSprop训练得到的效果一般好于其他方法。
3. 当样本量更少时，哪怕使用K折交叉验证，依然无济于事，得到的效果会更差。可以考虑增加数据的一些办法，但在此问题中我找不到能增加样本的办法。
4. 可能我的网络结构有些复杂，但是我之前也试过只有一个隐层（含有7个神经元）的网络，效果也不是很好。所以我不确定究竟是我网络的问题还是样本量的问题。