

# 机器学习第 7 次实验报告

李宜霖、苏家兴、窦志成

2021 年 06 月 16 日

## 目录

<b>1 概述</b>	<b>3</b>
<b>2 BP 算法推导</b>	<b>3</b>
2.1 全连接神经网络 . . . . .	3
2.1.1 权重调整 . . . . .	3
2.1.2 阈值调整 . . . . .	4
2.2 卷积神经网络 . . . . .	4
2.2.1 交叉熵损失函数 . . . . .	5
2.2.2 softmax 激活函数 . . . . .	5
2.2.3 全连接层 . . . . .	5
2.2.4 池化层 . . . . .	6
2.2.5 sigmoid 激活函数 . . . . .	6
2.2.6 卷积层 . . . . .	6

## 1 概述

这里是概述

- 第一条
- 第二条
- 第三条

Code 1: 这里是代码标题

```
1 def convolution_forward(X, W, b, stride = 1, pad = 0):
2     cal_X = im2col_forward(X, 5, 5, stride, pad)
3     result = np.dot(cal_X[0], W)
4     result += b
5     result = sigmoid(result)
6     return result
```

## 2 BP 算法推导

反向传播算法使用梯度下降策略对神经网络参数进行调整。

### 2.1 全连接神经网络

以如图1的三层全连接神经网络为例：

1. 含有  $d$  个神经元的输入层
2. 含有  $q$  个神经元的隐层
3. 含有  $l$  个神经元的输出层

假设网络的激活函数都是 Sigmoid 函数，即  $f(x) = \frac{1}{1+e^{-x}}$ ，每个神经元接受的输入如图1所示，定义误差函数为  $E = \frac{\sum_{i=1}^l (\hat{y}_i - y_i)^2}{2}$ ，即均方误差函数。以隐层到输出层的参数调整为例。

#### 2.1.1 权重调整

权重  $w_{ij}$  的调整方式如公式1。

$$\begin{aligned}\Delta w_{ij} &= -\eta \frac{\partial E}{\partial w_{ij}} \\ &= -\eta \frac{\partial E}{\partial \hat{y}_j} \frac{\partial \hat{y}_j}{\partial \hat{\beta}_j} \frac{\partial \hat{\beta}_j}{\partial w_{ij}}\end{aligned}\quad (1)$$

由于  $\frac{\partial E}{\partial \hat{y}_j} = \hat{y}_j - y_j$ ，Sigmoid 函数满足  $f'(x) = f(x)(1 - f(x))$ ， $\frac{\partial \hat{\beta}_j}{\partial w_{ij}} = b_i$ 。因此公式1可以变换为如公式2形式。

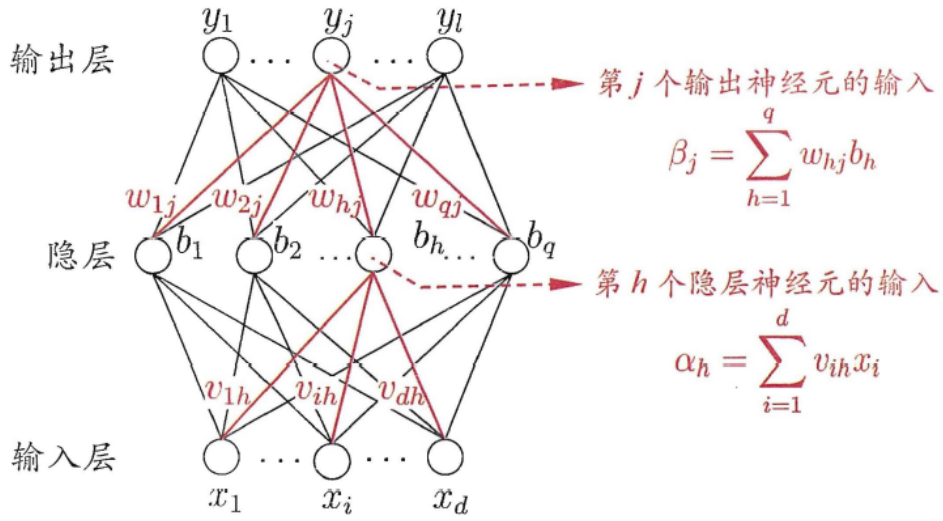


图 1: 含有单层隐含层的三层神经网络

$$\begin{aligned} \Delta w_{ij} &= -\eta(\hat{y}_j - y_j)(\hat{y}_j(1 - \hat{y}_j))(b_i) \\ &= -\eta b_i \hat{y}_j (\hat{y}_j - y_j)(1 - \hat{y}_j) \end{aligned} \quad (2)$$

### 2.1.2 阈值调整

类似于权重调整，也对  $E$  求阈值  $\theta$  的导数。

$$\Delta \theta_j = -\eta \frac{\partial E}{\partial \hat{y}_j} \frac{\partial \hat{y}_j}{\partial \theta_j} \quad (3)$$

由于 Sigmoid 函数满足  $f'(x) = f(x)(1 - f(x))$ ，因此公式3可以变换为如公式4形式。

$$\begin{aligned} \Delta \theta_j &= -\eta(\hat{y}_j - y_j)(-1)(\hat{y}_j(1 - \hat{y}_j)) \\ &= \eta \hat{y}_j (\hat{y}_j - y_j)(1 - \hat{y}_j) \end{aligned} \quad (4)$$

每次训练后，朝误差减少的方向调整参数。其中  $\eta$  是学习率。

## 2.2 卷积神经网络

以  $28 \times 28$  的输入图像，一层卷积、一层池化、一层全连接的神经网络为例，如图??。CNN 的模型参数只有卷积层的权值矩阵和偏置，全连接层的权值矩阵和偏置。因此求导顺序如下：

1. 交叉熵损失函数
2. softmax 激活函数
3. 全连接层
4. 池化层
5. sigmoid 激活函数

## 6. 卷积层

## 2.2.1 交叉熵损失函数

交叉熵定义为  $CELF = -(\sum_{k=1}^n y_k \log(\hat{y}_k))$ ，因此其导数定义如公式5。

$$\frac{\Delta CELF}{\Delta \hat{y}_k} = -\frac{y_k}{\hat{y}_k} \quad (5)$$

## 2.2.2 softmax 激活函数

softmax 激活函数定义为  $y_i = \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}$ ，其导函数定义如公式6。

$$\frac{\partial y_i}{\partial x_j} = \begin{cases} \frac{e^{x_j} (\sum_{k=1}^n e^{x_k} - e^{x_j})}{(\sum_{k=1}^n e^{x_k})^2} = y_i(1 - y_i), & i = j \\ \frac{0 - e^{x_i} e^{x_j}}{(\sum_{k=1}^n e^{x_k})^2} = -y_i y_j, & i \neq j \end{cases} \quad (6)$$

求 softmax 激活函数的输入  $x_i$  对于 CELF 的导数如公式7。

$$\begin{aligned} \frac{\partial CELF}{\partial x_i} &= \sum_{k=1}^n \frac{\partial CELF}{\partial y_k} \frac{\partial y_k}{\partial x_i} \\ &= -\frac{y_i}{\hat{y}_i} \hat{y}_i (1 - \hat{y}_i) + \sum_{k \neq i} \frac{y_k}{\hat{y}_k} \hat{y}_i \hat{y}_k \\ &= y_i \hat{y}_i + \sum_{k \neq i} y_k \hat{y}_i - y_i \\ &= \hat{y}_i \sum_{k=1}^n y_k - y_i \\ &= \hat{y}_i - y_i \end{aligned} \quad (7)$$

## 2.2.3 全连接层

全连接层可以由  $y = \sum_{k=1}^n x_k w_k + b$ ，其中  $w_k$  是第  $k$  个输入  $x_k$  的权重。因此其对  $w_k$  和  $b$  的导数定义为如公式8。

$$\begin{aligned} \frac{\Delta y}{\Delta w_i} &= x_i \\ \frac{\Delta y}{\Delta b} &= 1 \end{aligned} \quad (8)$$

由此得到 CELF 对权重  $w_k$  和偏置  $b$  的导数如公式9。其中  $x_i$  为 softmax 激活函数的输入， $w_j$  为全连接层第  $j$  输入  $a_j$  的权重

$$\begin{aligned}
\frac{\partial CELF}{\partial w_j} &= \frac{\partial CELF}{\partial x_i} \frac{\partial x_i}{\partial w_j} \\
&= (\hat{y}_i - y_i) a_j \\
\frac{\partial CELF}{\partial b} &= \frac{\partial CELF}{\partial x_i} \frac{\partial x_i}{\partial b} \\
&= (\hat{y}_i - y_i)
\end{aligned} \tag{9}$$

由于下面的卷积层仍然需要用到全连接层对输入的导数，因此求出 CELF 对输入  $a_j$  的导数如公式10。

$$\begin{aligned}
\frac{\partial CELF}{\partial a_j} &= \frac{\partial CELF}{\partial x_i} \frac{\partial x_i}{\partial a_j} \\
&= (\hat{y}_i - y_i) w_j
\end{aligned} \tag{10}$$

#### 2.2.4 池化层

池化层对全连接层的输入进行了重新排列组合并加权平均。池化层的输入和输出分别是  $[c_1, c_2, \dots, c_{2n^2}]$  和  $[a_1, a_2, \dots, a_{n^2}]$ ，其对应关系可以用公式11描述。

$$\begin{aligned}
w &= i \% n \\
h &= \left\lfloor \frac{i}{n} \right\rfloor \\
a_i &= \frac{c_{2h \times 2n + 2w} + c_{2h \times 2n + 2w + 1} + c_{(2h+1) \times 2n + 2w} + c_{(2h+1) \times 2n + 2w + 1}}{4}
\end{aligned} \tag{11}$$

池化层的导数恒为  $\frac{1}{4}$ ，因此对于梯度下降并没有影响。

#### 2.2.5 sigmoid 激活函数

sigmoid 激活函数定义为  $c = \frac{1}{1+e^{-d}}$ ，sigmoid 的导函数如公式12。

$$\frac{\partial c}{\partial d} = c(1 - c) \tag{12}$$

#### 2.2.6 卷积层

卷积层对输入的图像执行了卷积操作，可以用如公式13描述卷积层执行的操作。

$$\begin{bmatrix} d_1 \\ d_2 \\ \dots \\ d_m \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \dots & \dots & & \dots \\ u_{m1} & u_{m2} & \dots & u_{mn} \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix} + \begin{bmatrix} q \\ q \\ \dots \\ q \end{bmatrix} \tag{13}$$

因此，卷积层的导数如公式14。

$$\begin{aligned}\frac{\partial d_i}{\partial v_j} &= u_{ij} \\ \frac{\partial d_i}{\partial q} &= 1\end{aligned}\tag{14}$$

最终推出 CELF 对权重  $v_k$  和偏置  $q$  的导数公式如公式15，其中  $\frac{\partial CELF}{\partial a_j}$  是 CELF 对全连接层输入  $a_j$  的导数， $\frac{\partial a_j}{\partial c_l} \frac{\partial c_l}{\partial d_l}$  是  $a_j$  对卷积层输出  $d_l$  的导数， $\frac{\partial d_l}{\partial v_k}$  是卷积层输出  $d_l$  对权重  $v_k$  的导数。

$$\begin{aligned}\frac{\partial CELF}{\partial v_k} &= \sum_j \frac{\partial CELF}{\partial a_j} \sum_l \frac{\partial a_j}{\partial c_l} \frac{\partial c_l}{\partial d_l} \frac{\partial d_l}{\partial v_k} \\ &= \sum_j (\hat{y}_i - y_i) w_j \sum_l \frac{1}{4} c_l (1 - c_l) u_{lk} \\ \frac{\partial CELF}{\partial b} &= \sum_j \frac{\partial CELF}{\partial a_j} \sum_l \frac{\partial a_j}{\partial c_l} \frac{\partial c_l}{\partial d_l} \frac{\partial d_l}{\partial b} \\ &= \sum_j (\hat{y}_i - y_i) w_j \sum_l \frac{1}{4} c_l (1 - c_l)\end{aligned}\tag{15}$$