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

1 概述 3

2 BP 算法推导 3

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 . . . 5

2.2.5 sigmoid 激活函数 . . . 5

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

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

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

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

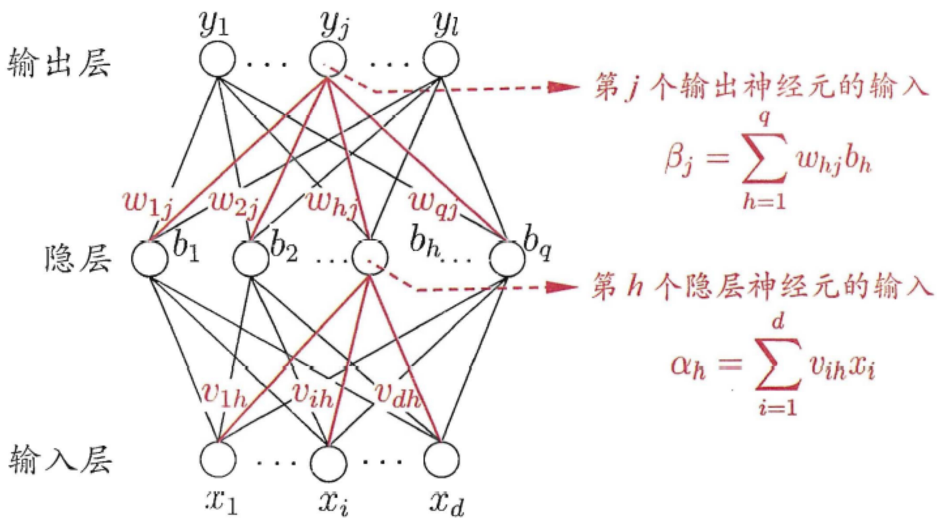

图 1: 含有单层隐含层的三层神经网络

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

2.1.2 阈值调整

类似于权重调整，也对 E 求阈值 θ 的导数。

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

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

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

每次训练后，朝误差减少的方向调整参数。其中 η 是学习率。

2.2 卷积神经网络

以 28*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 $y = -(\sum_i^n y_i log(\hat{y}_i) + (1-y_i)log(1-\hat{y}_i))$，因此其导数定义如公式5。

$$\frac{\Delta y}{\Delta \hat{y}_i} = \frac{1-y_i}{1-\hat{y}_i} - \frac{y_i}{\hat{y}_i} \tag{5}$$

2.2.2 softmax 激活函数

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

$$\frac{\Delta y}{\Delta x_i} = e^{x_i} \frac{\sum_{k=1}^n e^{x_k} - e^{x_i}}{(\sum_{k=1}^n e^{x_k})^2} \tag{6}$$

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7。

$$\begin{gathered}\frac{\Delta y}{\Delta w_i} = x_i \\ \frac{\Delta y}{\Delta b} = 1\end{gathered} \tag{7}$$

由于下面的卷积层仍然需要用到全连接层对输入的导数，因此求出全连接层对输入的导数如公式8。

$$\frac{\Delta y}{x_i} = w_i \tag{8}$$

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 $[x_1, x_2, \ldots, x_{2n^2}]$ 和 $[y_1, y_2, \ldots, y_{n^2}]$，其对应关系可以用公式9描述。

$$y_i = \frac{x_{2[\frac{i}{n}]\times 2n+2(i\%n)} + x_{2[\frac{i}{n}]\times 2n+2(i\%n+1)} + x_{2([\frac{i}{n}]+1)\times 2n+2(i\%n)} + x_{2([\frac{i}{n}]+1)\times 2n+2(i\%n+1)}}{4} \tag{9}$$

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 $y = \frac{1}{1+e^{-x}}$，sigmoid 的导函数如公式10。

$$\frac{\Delta y}{\Delta x} = y(1-y) \tag{10}$$

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11描述卷积层执行的操作。

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix} \tag{11}$$

因此，卷积层的导数如公式12。

$$\frac{\Delta y_i}{\Delta w_j} = x_{ij} \tag{12}$$

最后使用导数的求导法则，将这些导数相互嵌套，即可求出对权值和偏置的导数，并沿导数下降的方向更新权值和偏置。