

1. 简述误差反向播算法的基本原理。

解：BP算法由信号的正向传播和误差的反向传播两个过程组成。

正向传播时，输入样本从输入层进入网络，经隐层逐层传递至输出层，如果输出层的实际输出与期望输出不同，则转至误差反向传播；如果输出层的实际输出与期望输出相同，结束学习算法。

反向传播时，将输出误差(期望输出与实际输出之差)按原通路反传计算，通过隐层反向，直至输入层，在反传过程中将误差分摊给各层的各个单元，获得各层各单元的误差信号，并将其作为修正各单元权值的根据。这一计算过程使用梯度下降法完成，在不停地调整各层神经元的权值和阈值后，使误差信号减小到最低限度。

权值和阈值不断调整的过程，就是网络的学习与训练过程，经过信号正向传播与误差反向传播，权值和阈值的调整反复进行，一直进行到预先设定的学习训练次数，或输出误差减小到允许的程度。



2. 给定数据 x 以及卷积核 w 如下，设步长为1，计算卷积输出。

$$x_{[:, :, 1]} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}, \quad x_{[:, :, 2]} = \begin{bmatrix} 2 & 1 & 2 & 2 \\ 0 & 2 & 1 & 1 \\ 1 & 2 & 0 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad x_{[:, :, 3]} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$
$$w_{[:, :, 1]} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad w_{[:, :, 2]} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad w_{[:, :, 3]} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

问题：卷积计算的时候出现错误

解：

$$y_{[:, :, 1]} = x_{[:, :, 1]} * w_{[:, :, 1]} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$y_{[:, :, 2]} = x_{[:, :, 2]} * w_{[:, :, 2]} = \begin{bmatrix} 3 & 3 & 4 \\ 2 & 3 & 2 \\ 3 & 2 & 2 \end{bmatrix}$$

$$y_{[:, :, 3]} = x_{[:, :, 3]} * w_{[:, :, 3]} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 3 \\ 2 & 2 & 3 \end{bmatrix}$$

3. 对于如下数据

$$x = \begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 2 \\ 1 & 3 & 1 & 4 & 1 & 3 \\ 2 & 3 & 2 & 3 & 1 & 5 \\ 6 & 2 & 2 & 5 & 4 & 1 \\ 1 & 2 & 5 & 4 & 1 & 3 \\ 4 & 4 & 3 & 1 & 6 & 7 \end{bmatrix}$$

1) 分别用 2×2 和 3×3 进行无重叠最大池化

2) 如果采用有重叠的池化且步长为1, 试求取最大池化结果。

解: 1) 采用 2×2 进行无重叠最大池化,

$$X(1,1) = \max\{1,2,1,3\} = 3$$

$$X(1,2) = \max\{1,2,1,4\} = 4$$

$$X(1,3) = \max\{1,2,1,3\} = 3$$

$$X(2,1) = \max\{2,3,6,2\} = 6$$

$$X(2,2) = \max\{2,3,2,5\} = 5$$

$$X(2,3) = \max\{1,5,4,1\} = 5$$

$$X(3,1) = \max\{1,2,4,4\} = 4$$

$$X(3,2) = \max\{5,4,3,1\} = 5$$

$$X(3,3) = \max\{1,3,6,7\} = 7$$

$$\text{所以 } X = \begin{bmatrix} 3 & 4 & 3 \\ 6 & 5 & 5 \\ 4 & 5 & 7 \end{bmatrix}$$



采用 3×3 进行无重叠最大池化,

$$X(1,1) = \max\{1,2,1,1,3,1,2,3,2\} = 3$$

$$X(1,2) = \max\{2,1,2,4,1,3,3,1,5\} = 5$$

$$X(2,1) = \max\{6,2,2,1,2,5,4,4,3\} = 6$$

$$X(2,2) = \max\{5,4,1,4,1,3,1,6,7\} = 7$$

$$\text{所以} X = \begin{bmatrix} 3 & 5 \\ 6 & 7 \end{bmatrix}$$

2) 采用 2×2 进行最大池化且步长为1,

$$X(1,1) = \max\{1,2,1,3\} = 3$$

$$X(1,2) = \max\{2,1,3,1\} = 3$$

$$X(1,3) = \max\{1,2,1,4\} = 4$$

....

$$X(5,3) = \max\{5,4,3,1\} = 5$$

$$X(5,4) = \max\{4,1,1,6\} = 6$$

$$X(5,5) = \max\{1,3,6,7\} = 7$$

$$\text{所以} X = \begin{bmatrix} 3 & 3 & 4 & 4 & 3 \\ 3 & 3 & 4 & 4 & 5 \\ 6 & 3 & 5 & 5 & 5 \\ 6 & 5 & 5 & 5 & 4 \\ 4 & 5 & 5 & 6 & 7 \end{bmatrix}$$



如果采 3×3 进行最大池化且步长为1,

$$X(1,1) = \max\{1,2,1,1,3,1,2,3,2\} = 3$$

$$X(1,2) = \max\{2,1,2,3,1,4,3,2,3\} = 4$$

$$X(1,3) = \max\{1,2,1,1,4,1,2,3,1\} = 4$$

$$X(1,4) = \max\{2,1,2,4,1,3,3,1,5\} = 5$$

... ..

$$X(4,1) = \max\{6,2,2,1,2,5,4,4,3\} = 6$$

$$X(4,2) = \max\{2,2,2,2,5,4,4,3,1\} = 5$$

$$X(4,3) = \max\{2,5,4,5,4,1,3,1,6\} = 6$$

$$X(4,4) = \max\{5,4,1,4,1,3,1,6,7\} = 7$$

$$\text{所以 } X = \begin{bmatrix} 3 & 4 & 4 & 5 \\ 6 & 5 & 5 & 5 \\ 6 & 5 & 5 & 5 \\ 6 & 5 & 6 & 7 \end{bmatrix}$$

4. 对于某一多分类问题及对应的网络对两个样本输出 $y_1 = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.7 \end{bmatrix}$ $y_2 = \begin{bmatrix} 0.6 \\ 0.2 \\ 0.2 \end{bmatrix}$, 对应的真实标签为 $\hat{y}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $\hat{y}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ 试计算该网络输出的交叉熵损失。

解：交叉熵损失： $L = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}))$

$$L(1) = -[\log(0.9) + \log(0.8) + \log(0.7)] = 0.297$$

$$L(2) = -[\log(0.6) + \log(0.8) + \log(0.8)] = 0.415$$

分析：部分同学代入的是 $L(Y, P) = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{K-1} y_{ik} \log(p_{ik})$

5. 如何理解Inception的结构设计，Inception中不同大小感受野是如何实现的？

答：使用一个密集成分来近似或者代替最优的局部稀疏结构。

a) 采用不同大小的卷积核意味着不同大小的感受野，最后拼接意味着不同尺度特征的融合；

b) 之所以卷积核大小采用1、3和5，主要是为了方便对齐；

c) 网络越到后面，特征越抽象，而且每个特征所涉及的感受野也更大了，因此随着层数的增加， 3×3 和 5×5 卷积的比例也要增加。



6. 设输入数据为

$$x = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 3 \\ 2 & 1 & 1 & 1 \\ 5 & 0 & 3 & 1 \end{bmatrix}$$

卷积核 $w = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, 在增加Padding操作以确保卷积输出一致的前提下, 请写出残差模块下的卷积输出。

解: 因为要通过增加padding操作以确保卷积输出一致, 因此 $Padding = 2, stride = 2$ 。经过padding后, 得到

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 5 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

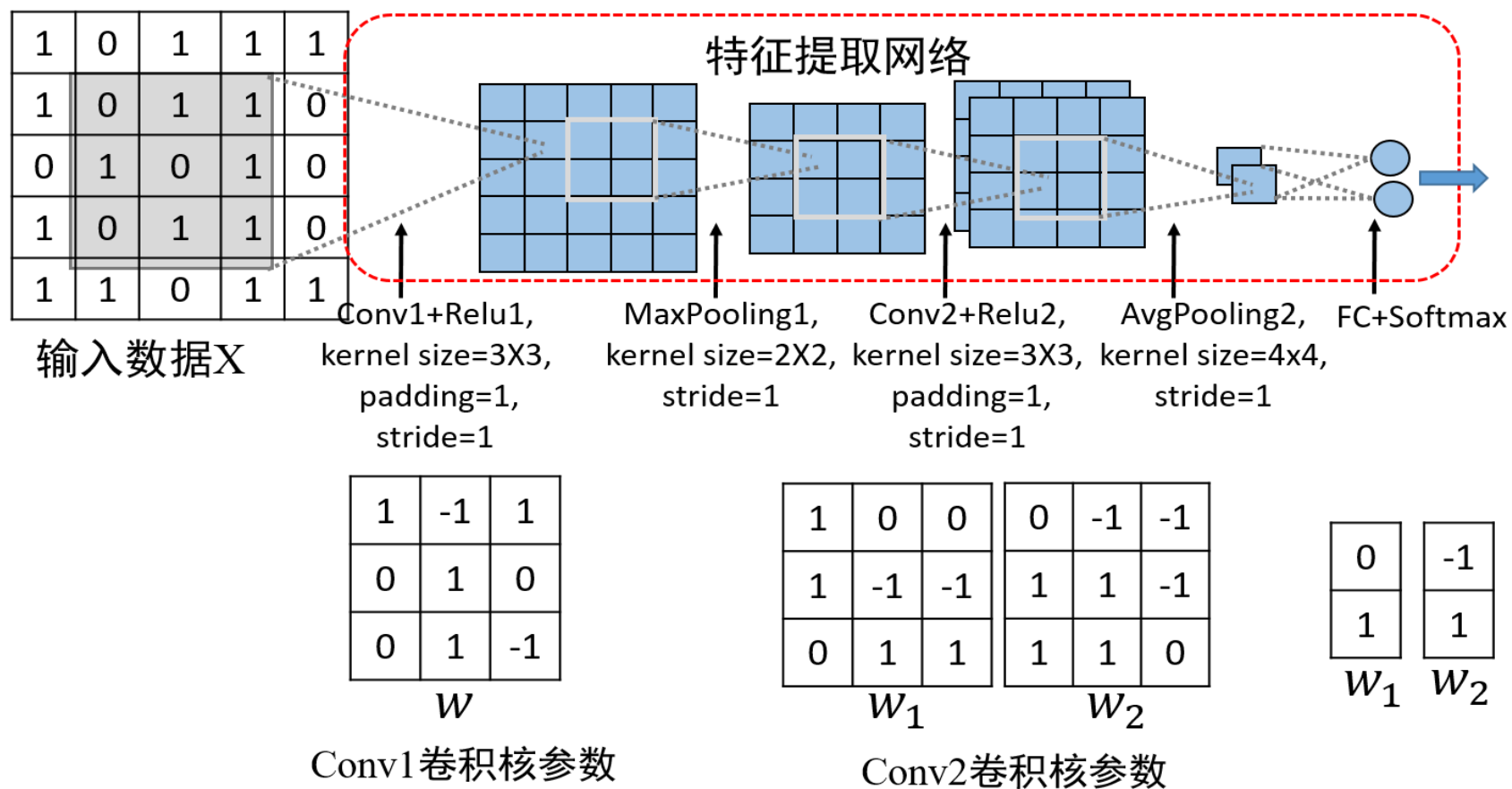
经过卷积后得到,

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

从而残差模块的卷积输出为

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 3 \\ 2 & 1 & 1 & 1 \\ 5 & 0 & 3 & 1 \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 2 & 2 & 1 & 3 \\ 2 & 3 & 2 & 1 \\ 5 & 0 & 3 & 1 \end{bmatrix}$$

7. 给定输入数据X，特征提取网络结构和参数，输出数据？



解：步骤一：padding操作（在x 周围一圈补0），卷积，Relu

$$\text{padding}(x) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{padding}(x) * w = \begin{bmatrix} 2 & -1 & 1 & 2 & 1 \\ -1 & 3 & 0 & 3 & 0 \\ 0 & 2 & 0 & 2 & 1 \\ 2 & 0 & 2 & 0 & 2 \\ 0 & 3 & 0 & 1 & 2 \end{bmatrix}$$

$$X_1 = \text{Relu}(\text{padding}(x) * w) = \begin{bmatrix} 2 & 0 & 1 & 2 & 1 \\ 0 & 3 & 0 & 3 & 0 \\ 0 & 2 & 0 & 2 & 1 \\ 2 & 0 & 2 & 0 & 2 \\ 0 & 3 & 0 & 1 & 2 \end{bmatrix}$$

步骤二：最大池化操作MaxPooling

$$x_{11} = \max(2, 0, 0, 3) = 3$$

$$x_{12} = \max(0, 3, 0, 1) = 3$$

...

$$x_{43} = \max(2, 0, 0, 1) = 2$$

$$x_{44} = \max(0, 1, 2, 2) = 2$$

$$X_2 = \text{MaxPooling}(X_1) = \begin{bmatrix} 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 2 & 2 \end{bmatrix}$$

步骤三： padding操作，卷积，Relu

$$\text{padding}(X_2) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 3 & 3 & 3 & 0 \\ 0 & 3 & 3 & 3 & 3 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 3 & 3 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{padding}(X_2) * w_1 = \begin{bmatrix} 0 & 3 & 3 & 3 \\ -2 & 4 & 4 & 5 \\ 2 & 6 & 5 & 5 \\ -6 & 0 & 1 & 2 \end{bmatrix}$$

$$X_3^{(1)} = \text{Relu}(\text{padding}(X_2) * w_1) = \begin{bmatrix} 0 & 3 & 3 & 3 \\ 0 & 4 & 4 & 5 \\ 2 & 6 & 5 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

$$\text{padding}(X_2) * w_2 = \begin{bmatrix} 3 & 9 & 9 & 12 \\ -4 & 1 & 1 & 7 \\ -3 & 2 & 1 & 5 \\ -4 & 0 & -1 & 2 \end{bmatrix}$$

$$X_3^{(2)} = \text{Relu}(\text{padding}(X_2) * w_2) = \begin{bmatrix} 3 & 9 & 9 & 12 \\ 0 & 1 & 1 & 7 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

$$X_3 = \text{concat}(X_3^{(1)}, X_3^{(2)}) = \left[\begin{bmatrix} 0 & 3 & 3 & 3 \\ 0 & 4 & 4 & 5 \\ 2 & 6 & 5 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad \begin{bmatrix} 3 & 9 & 9 & 12 \\ 0 & 1 & 1 & 7 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 0 & 2 \end{bmatrix} \right]$$

步骤四：平均池化操作，由于池化核大小是 4×4 ，故为全局平均池化

$$X_4^{(1)} = \text{AvgPooling}(X_3^{(1)}) = \text{mean}(X_3^{(1)}.\text{value}) = 2.6875$$

$$X_4^{(2)} = \text{AvgPooling}(X_3^{(2)}) = \text{mean}(X_3^{(2)}.\text{value}) = 3.2500$$

$$X_4 = \begin{bmatrix} X_4^{(1)} \\ X_4^{(2)} \end{bmatrix} = \begin{bmatrix} 2.6875 \\ 3.2500 \end{bmatrix}$$

步骤五：全连接层

$$X_5 = [w_1 \quad w_2]^T X_4 = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2.6875 \\ 3.2500 \end{bmatrix} = \begin{bmatrix} 3.2500 \\ 0.5625 \end{bmatrix}$$

步骤六：softmax归一化为概率向量

$$\begin{aligned} X_6 &= \begin{bmatrix} \frac{e^{3.2500}}{e^{3.2500} + e^{0.5625}} & \frac{e^{0.5625}}{e^{3.2500} + e^{0.5625}} \end{bmatrix}^T \\ &= \begin{bmatrix} 0.9363 \\ 0.0637 \end{bmatrix} \end{aligned}$$