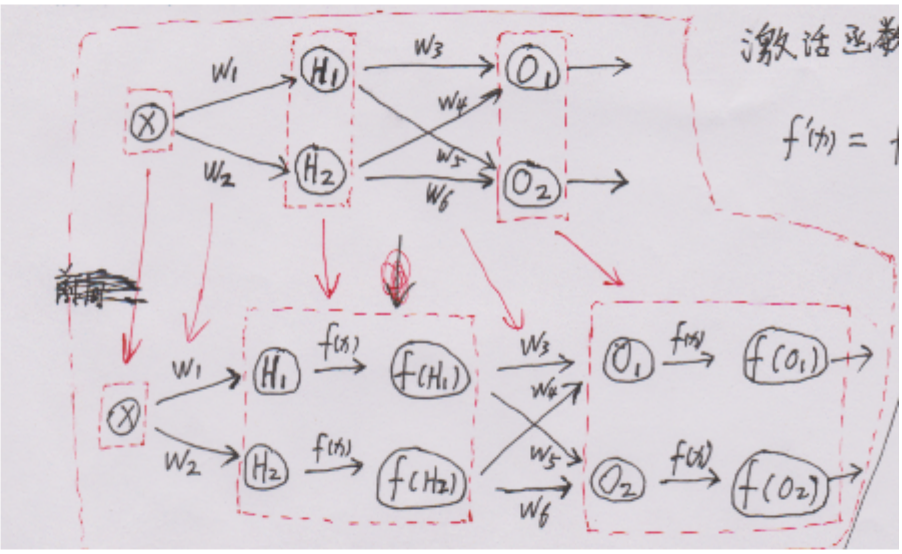
反向传播

1. 前向传播



$$
\begin{cases}
H_1 = w_1 x \\
f(H_1) = \dfrac{1}{1+e^{-H_1}} \\
O_1 = w_3 f(H_1) + w_4 f(H_2) \\
f(O_1) = \dfrac{1}{1+e^{-O_1}}
\end{cases}
\qquad
\begin{cases}
H_2 = w_2 x \\
f(H_2) = \dfrac{1}{1+e^{-H_2}} \\
O_2 = w_5 f(H_1) + w_6 f(H_2) \\
f(O_2) = \dfrac{1}{1+e^{-O_2}}
\end{cases}
$$

2. 误差

$$
E = -\sum \frac{1}{2}(f(O)-y)^2 = -\frac{1}{2}(f(O_1 - y_{O_1})) - \frac{1}{2}(f(O_2) - y_{O_2})^2
$$

3. 反向传播

$$
\frac{\partial E}{\partial w_3} = \frac{\partial E}{\partial f(O_1)} \frac{\partial f(O_1)}{\partial O_1} \frac{\partial O_1}{\partial w_3} = (y_{O_1} - f(O_1)) \cdot \underline{f(O_1) \cdot (1 - f(O_1))} \cdot f(H_1)
$$

$$\frac{\partial E}{\partial w_1}=\boxed{(\frac{\partial E}{\partial O_1}\frac{\partial O_1}{\partial f(H_1)}+\frac{\partial E}{\partial O_2}\frac{\partial O_2}{\partial f(H_1)})}\frac{\partial f(H_1)}{\partial H_1}\frac{\partial H_1}{\partial w_1}$$

4. 更新

$$w=w-\eta*\frac{\partial E}{\partial w}$$

5. 特殊环节的反向传播

- **ReLU**:

$$\mathrm{Re}LU(x)=\begin{cases}x,x>0\\0,x\le 0\end{cases}$$

ReLU 在 x=0 处不可微，直接将其在 x=0 处的导数置为 1，

$$\mathrm{Re}LU'(x)=\begin{cases}1,x\ge 0\\0,x<0\end{cases}$$

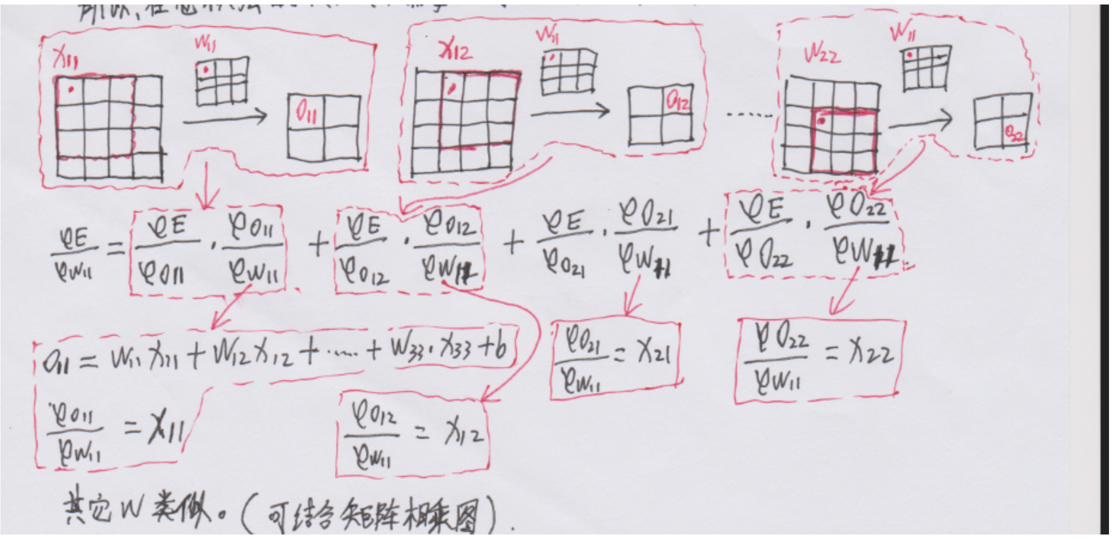
- **Pooling：**

把一个梯度传递给 nxn 个像素，需要保证梯度传递的总和不变；

$$\begin{cases}mean\ pooling:\ 平分n分进行传递，每个像素\dfrac{1}{n}\\ \max pooling:\ 前向时保持pooling位置，反向时根据位置传递给前一层，其他位置为0\end{cases}$$

- **卷积层**

在反向传播过程中，若第 L 层的 a 节点通过权值 w 对 L+1 层的 b 节点有贡献，则在方向过程中，梯度通过权值 w 从 b 节点传播回 a 节点。都遵循这个规律，所以，在卷积层的反向中，需要找到卷积层 L 中的每个单元和 L+1 层中哪些单元相关联。



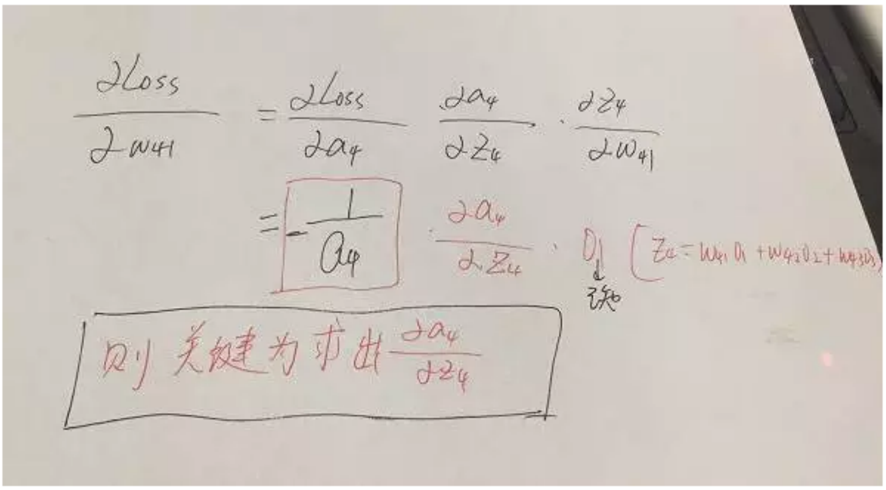
6. 核心思想

链式求导法则，动态规划

7. Softmax 反向传播

https://mp.weixin.qq.com/s/MS8h8BUv1BC3Ql9w2oxmJg

- 因为只有一个 y=1，其他都为 0，所以这里导数可以写成下面形式。



if $j = i$:

$$\frac{\partial a_j}{\partial z_i} = \frac{\partial}{\partial z_i}\left(\frac{e^{z_j}}{\sum_k e^{z_k}}\right) = \frac{(e^{z_j})' \cdot \sum_k e^{z_k} - e^{z_j} \cdot e^{z_j}}{\left(\sum_k e^{z_k}\right)^2} = \frac{e^{z_j}}{\sum_k e^{z_k}} - \frac{e^{z_j}}{\sum_k e^{z_k}} \cdot \frac{e^{z_j}}{\sum_k e^{z_k}} = a_j(1-a_j)$$

if $j \neq i$:

$$\frac{\partial a_j}{\partial z_i} = \frac{\partial}{\partial z_i}\left(\frac{e^{z_j}}{\sum_k e^{z_k}}\right) = \frac{0 \cdot \sum_k e^{z_k} - e^{z_j} \cdot e^{z_i}}{\left(\sum_k e^{z_k}\right)^2} = -\frac{e^{z_j}}{\sum_k e^{z_k}} \cdot \frac{e^{z_i}}{\sum_k e^{z_k}} = -a_j a_i$$

将这三个式子组合得到神奇的效果是：

如果输出是：[0.0903,0.2447,0.665]

那么梯度就是[0.0903,0.2447-1,0.665]=[0.0903,-0.7553,0.665]