神经常微分方程的理论推导——后向传播过程

郭晓波[1]

[1] 机电工程学院
池州学院

** 大学, 2022 年 7 月 1 日

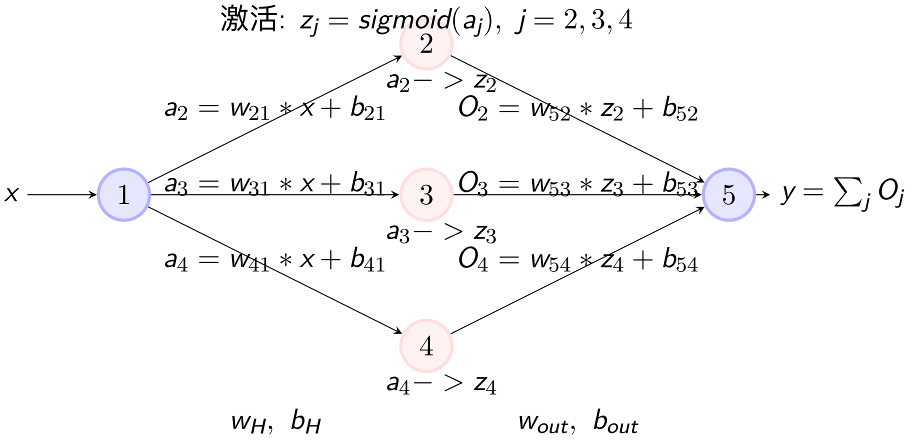
内容大纲

1 神经网络

内容大纲

1 神经网络

前向传播网络

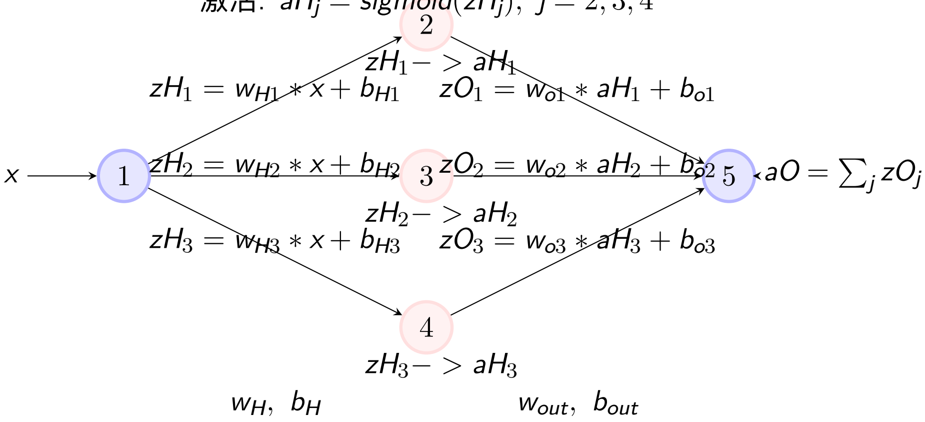


激活: $z_j = sigmoid(a_j),\ j = 2, 3, 4$
x
1
2
3
4
5
$a_2 = w_{21} * x + b_{21}$
$a_3 = w_{31} * x + b_{31}$
$a_4 = w_{41} * x + b_{41}$
$a_2 -> z_2$
$a_3 -> z_3$
$a_4 -> z_4$
$O_2 = w_{52} * z_2 + b_{52}$
$O_3 = w_{53} * z_3 + b_{53}$
$O_4 = w_{54} * z_4 + b_{54}$
$y = \sum_j O_j$
$w_H,\ b_H$
$w_{out},\ b_{out}$

前向传播网络

和前面网络标注进行对照，以符合我们第一个代码。

激活: $aH_j = sigmoid(zH_j)$, $j = 2, 3, 4$



后向传播公式

常微分方程：

$$\frac{dy}{dx} = f(x, y), \quad x \in [0, 1] \tag{1}$$

对 x 进行离散化：$x_1, x_2, ..., x_n$ 后，我们可以构建神经网络 $N(x, p)$，可以得到试探解，

$$y_{trial} = A_0 + x * N(x, p) \tag{2}$$

其中，A_0 是常数，p 是网络参数 w_{ij}, b_{ij}，$(w_{H,out}, b_{H,out})$，其中 b_{out} 为零。可以构建损失函数（误差函数）为，

$$E(w_{ij}, b_{ij}) = \sum_i \left(\frac{dy_{trial,i}}{dx} - f(x_i, y_{trial,i}) \right)^2 \tag{3}$$

后向传播公式

$$E(w,b)=\sum_i\left(\frac{dy_{trial}}{dx}-f(x,y_{trial})\right)^2 \tag{4}$$

我们下面分别求 $E(w,b)$ 对 $w_{H2}, b_{H2}, w_{o3}, b_{o3}$ 的导数，

$$\begin{aligned}\frac{\partial E(w,b)}{\partial w_{H2}}&=\frac{\partial}{\partial w_{H2}}\sum_i\left(\frac{dy_{trial,i}}{dx}-f(x_i,y_{trial,i})\right)^2\\&=\sum_i\frac{\partial}{\partial w_{H2}}\left(\frac{dy_{trial,i}}{dx}-f(x_i,y_{trial,i})\right)^2\\&=2\sum_i\left(\frac{dy_{trial,i}}{dx}-f(x_i,y_{trial,i})\right)\frac{\partial}{\partial w_{H2}}\left(\left.\frac{dy_{trial}}{dx}\right|_{x_i}-f(x_i,y_{trial,i})\right)\end{aligned} \tag{5}$$

后向传播公式

对于 $y_{trial}=A_0+x*N(x,p)$， $N(x_i,p)=aO(x_i,p)=a_{out}(x_i,p)=a_{out}(x_i)$，

$$\begin{aligned}\left.\frac{dy_{trial}}{dx}\right|_{x_i} &= \frac{d}{dx}(x*N(x,p)) \\ &= N(x_i,p)+x_i*\frac{dN(x_i,p)}{dx} \\ &= a_{out}(x_i)+x_i*\frac{da_{out}(x_i)}{dx}\end{aligned}$$

$$\begin{aligned}\frac{\partial E(w,b)}{\partial w_{H2}} &= 2\sum_i\left(\left.\frac{dy_{trial}}{dx}\right|_{x_i}-f(x_i,y_{trial,i})\right)\frac{\partial}{\partial w_{H2}}\left(\left.\frac{dy_{trial}}{dx}\right|_{x_i}-f(x_i,y_{trial,i})\right) \\ &= 2\sum_i\left(\left(a_{out}(x_i)+x_i*\frac{da_{out}(x_i)}{dx}\right)-f(x_i,y_{trial,i})\right) \\ &\quad \frac{\partial}{\partial w_{H2}}\left(\left(a_{out}(x_i)+x_i*\frac{da_{out}(x_i)}{dx}\right)-f(x_i,y_{trial,i})\right)\end{aligned}$$

后向传播公式

$$\frac{\partial}{\partial w_{H2}}\left(\left(a_{out}(x_i)+x_i*\frac{da_{out}(x_i)}{dx}\right)-f(x_i,y_{trial,i})\right)$$
$$=\frac{\partial}{\partial w_{H2}}\left(a_{out}(x_i,p)+x_i*\frac{da_{out}(x_i,p)}{dx}-f(x_i,A_0+x_i*a_{out}(x_i,p))\right)$$

下面，逐项进行计算（其中，$D\sigma(z)=d\sigma(z)/dz=\sigma(z)*(1-\sigma(z))$），

$$\frac{\partial}{\partial w_{H2}}a_{out}(x_i,p)=\sum_{j=1,2,3}\frac{\partial zO_j}{\partial wH_2}$$
$$=\sum_{j=1,2,3}wO_j\frac{\partial aH_j}{\partial wH_2}=\sum_{j=1,2,3}wO_j\frac{\partial\sigma(zH_j)}{\partial wH_2}$$
$$=\sum_{j=1,2,3}wO_j*D\sigma(zH_j)\frac{\partial zH_j}{\partial wH_2}=wO_j*D\sigma(zH_j)\big|_{j=2}*x_i$$

后向传播公式

(其中，$D\sigma(z) = d\sigma(z)/dz = \sigma(z) * (1 - \sigma(z))$)，另外，

$$\begin{aligned}
\frac{da_{out}(x_i)}{dx} &= \sum_j \frac{dzO_j}{dx} = \sum_j \frac{d}{dx}(wO_j \cdot aH_j + bO_j) \\
&= \sum_j \left(wO_j \cdot \frac{daH_j}{dx} + 0\right) = \sum_j \left(wO_j \cdot \frac{d\sigma(zH_j)}{dx}\right) \\
&= \sum_j \left(wO_j \cdot D\sigma(zH_j)\frac{dzH_j}{dx}\right) = \sum_j wO_j \cdot D\sigma(zH_j) \cdot wH_j
\end{aligned} \tag{6}$$

所以，$\frac{\partial}{\partial w_{H2}}\left(\frac{da_{out}(x_i)}{dx}\right) = wO_j * D\sigma(zH_j)|_{j=2}$，和下面的计算结果相同。

$$\begin{aligned}
&\frac{\partial}{\partial w_{H2}}\left(\frac{da_{out}(x_i)}{dx}\right) = \frac{d}{dx}\left(\frac{\partial a_{out}(x_i)}{\partial w_{H2}}\right) \\
=&\frac{d}{dx}\frac{\partial a_{out}(x_i)}{\partial w_{H2}} = \frac{d}{dx}\left(wO_j * D\sigma(zH_j)|_{j=2} * x_i\right) \\
=& wO_j * D\sigma(zH_j)|_{j=2}
\end{aligned} \tag{7}$$

后向传播公式

(其中，$y_{trial,i} = A_0 + x_i * a_{out}(x_i, p)$)，

$$
\begin{aligned}
&\frac{\partial}{\partial w_{H2}} f(x_i, A_0 + x_i * a_{out}(x_i, p)) = \frac{\partial}{\partial w_{H2}} f(x_i, y_{trial,i}) \\
=&\frac{\partial f}{\partial y_{trial,i}} \frac{\partial}{\partial wH_2} (A_0 + x_i * a_{out}(x_i, p)) = x_i * \frac{\partial f}{\partial y_{trial,i}} \frac{\partial a_{out}(x_i, p)}{\partial wH_2} \\
=&x_i * \frac{\partial f}{\partial y_{trial,i}} \left(wO_j * D\sigma(zH_j)|_{j=2} * x_i \right) \\
=&x_i^2 * \frac{\partial f}{\partial y_{trial,i}} \left(wO_j * D\sigma(zH_j)|_{j=2} \right)
\end{aligned}
\tag{8}
$$

后向传播公式

$$\frac{\partial}{\partial w_{H2}}\left(a_{out}(x_i,p)+x_i*\frac{da_{out}(x_i,p)}{dx}-f(x_i,A_0+x_i*a_{out}(x_i,p))\right)$$
$$=x_i\cdot wO_j*D\sigma(zH_j)\big|_{j=2}\cdot\left(2-x_i*\frac{\partial f}{\partial y_{trial,i}}\right) \tag{9}$$

后向传播公式

$$
\begin{aligned}
\frac{\partial E(w,b)}{\partial w_{H2}} =& 2\sum_i \left(\left(a_{out}(x_i) + x_i * \frac{da_{out}(x_i)}{dx}\right) - f(x_i, y_{trial,i})\right) \\
& \frac{\partial}{\partial w_{H2}}\left(\left(a_{out}(x_i) + x_i * \frac{da_{out}(x_i)}{dx}\right) - f(x_i, y_{trial,i})\right) \\
=& 2\sum_i \left(\left(a_{out}(x_i) + x_i * \frac{da_{out}(x_i)}{dx}\right) - f(x_i, y_{trial,i})\right) \\
& x_i \cdot wO_j * D\sigma(zH_j)|_{j=2} \cdot \left(2 - x_i * \frac{\partial f}{\partial y_{trial,i}}\right) \\
=& x_i \cdot wO_j \cdot D\sigma(zH_j)|_{j=2} \cdot 2 \cdot \sum_i \left(a_{out} + x_i * \frac{da_{out}}{dx} - f(x_i, y_{trial,i})\right) \\
& \cdot \left(2 - x_i * \frac{\partial f}{\partial y_{trial,i}}\right)
\end{aligned}
$$

后向传播公式

把本公式和代码中的计算做对比，

$$
\begin{aligned}
\frac{\partial E(w,b)}{\partial w_{H2}} =& x_i \cdot\, wO_j \cdot D\sigma(zH_j)\big|_{j=2} \cdot 2 \cdot \sum_i \left(a_{out} + x_i * \frac{da_{out}}{dx} - f(x_i, y_{trial,i}) \right) \\
& \cdot \left(2 - x_i * \frac{\partial f}{\partial y_{trial,i}} \right) \\
=& \frac{\partial}{\partial w_{H2}} a_{out}(x_i, p) \cdot 2 \cdot \sum_i \left(a_{out} + x_i * \frac{da_{out}}{dx} - f(x_i, y_{trial,i}) \right) \\
& \cdot \left(2 - x_i * \frac{\partial f}{\partial y_{trial,i}} \right) \\
=& \frac{\partial}{\partial w_{H2}} a_{out}(x_i, p) \sum_i \frac{\partial}{\partial a_{out}} \left(a_{out} + x_i * \frac{da_{out}}{dx} - f(x_i, y_{trial,i}) \right)^2
\end{aligned}
$$

后向传播公式

上式中，如果 $\frac{\partial}{\partial a_{out}} da_{out}/dx = 0$

$$
\begin{aligned}
\frac{\partial E(w,b)}{\partial w_{H2}} &= \frac{\partial}{\partial w_{H2}} a_{out}(x_i,p) \sum_i \frac{\partial}{\partial a_{out}} \left(a_{out} + x_i * \frac{da_{out}}{dx} - f(x_i, y_{trial,i}) \right)^2 \\
&= \frac{\partial}{\partial w_{H2}} a_{out}(x_i,p) \cdot 2 \cdot \sum_i \left(a_{out} + x_i * \frac{da_{out}}{dx} - f(x_i, y_{trial,i}) \right) \\
&\quad \cdot \left(1 + 0 - \frac{\partial f}{\partial y_{trial,i}} \frac{\partial y_{trial,i}}{\partial a_{out}} \right)
\end{aligned}
$$

其中，

$$
\frac{\partial y_{trial,i}}{\partial a_{out}} = \frac{\partial}{\partial a_{out}} \left(A_0 + x_i \cdot a_{out}(x_i) \right) = x_i \tag{10}
$$

这个差异可以和前面做对比，这里是对的，前面是错的，关键在于：

$$
\frac{\partial}{\partial a_{out}} \left(\frac{da_{out}}{dx} \right) = 0
$$

后向传播公式

但是和代码中的计算公式依然有区别：

$$
\begin{aligned}
\frac{\partial E(w,b)}{\partial w_{H2}} =& \frac{\partial}{\partial w_{H2}} a_{out}(x_i,p) \cdot 2 \cdot \sum_i \left(a_{out} + x_i * \frac{da_{out}}{dx} - f(x_i, y_{trial,i}) \right) \\
& \cdot \left(1 + 0 - \frac{\partial f}{\partial y_{trial,i}} \frac{\partial y_{trial,i}}{\partial a_{out}} \right) \\
=& \frac{\partial}{\partial w_{H2}} a_{out}(x_i,p) \cdot 2 \cdot \sum_i \left(a_{out} + x_i * \frac{da_{out}}{dx} - f(x_i, y_{trial,i}) \right) \\
& \cdot \left(1 - \frac{\partial f}{\partial y_{trial,i}} x_i \right) \\
=& \frac{\partial}{\partial w_{H2}} a_{out}(x_i,p) \cdot grad_N \\
=& x_i \cdot \ wO_j * D\sigma(zH_j)\big|_{j=2} \cdot grad_N
\end{aligned} \tag{11}
$$

后向传播公式

代码中 ΔwH_2 计算公式如下：

$$\begin{aligned}\Delta wH_2 &= \frac{\partial E(w,b)}{\partial w_{H2}} \\ &= x_i \cdot wO_j * D\sigma(zH_j)|_{j=2} \cdot D\sigma(z_{out}) \cdot grad_N \\ &= x_i \cdot wO_j * D\sigma(zH_j)|_{j=2} \cdot D\sigma(aO) \cdot grad_N \\ &= x_i \cdot wO_j * D\sigma(zH_j)|_{j=2} \cdot D\sigma(\sum_i zO_i) \cdot grad_N \end{aligned} \tag{12}$$

我们上面推导的结果如下，

$$\begin{aligned}\Delta wH_2 &= \frac{\partial E(w,b)}{\partial w_{H2}} \\ &= x_i \cdot wO_j * D\sigma(zH_j)|_{j=2} \cdot \cdot grad_N \end{aligned} \tag{13}$$

比较后，发现我们的结果中少了一项：$D\sigma(\sum_i zO_i)$

后向传播公式

代码中计算公式：$dw_H = x(i) * w_{out} * dsig(z_H) * dsig(z_{out}) * grad_N$，比我们多了倒数第二项。
但是，前向传播过程中，$a_out = z_out$，输出层并没有使用激活函数。

后向传播公式

(其中，$y_{trial,i} = A_0 + x_i * a_{out}(x_i, p)$)，代码中 ΔwH_2 计算公式如下：

$$\begin{aligned}
\Delta wH_2 = \frac{\partial E(w,b)}{\partial wH_2} &= \frac{\partial E(w,b)}{\partial a_{out}} \frac{\partial a_{out}}{\partial wH_2} \\
=& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i}) \right) \frac{\partial}{\partial a_{out}} \left(a_{out} + x_i \cdot \frac{da_{out}}{dx} - f(x_i, y_{t,i}) \right) \frac{\partial a_{out}}{\partial wH_2} \\
=& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i}) \right) \left(1 - \frac{\partial f(x_i, y_{t,i})}{\partial a_{out}} \right) \frac{\partial a_{out}}{\partial wH_2} \\
=& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i}) \right) \left(1 - \frac{\partial f(x_i, y_{t,i})}{\partial y_{t,i}} \frac{\partial y_{t,i}}{\partial a_{out}} \right) \frac{\partial a_{out}}{\partial wH_2} \\
=& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i}) \right) \left(1 - \frac{\partial f(x_i, y_{t,i})}{\partial y_{t,i}} \cdot x_i \right) \frac{\partial a_{out}}{\partial wH_2}
\end{aligned} \tag{14}$$

根据高博的推导，这个推导还需要修正，下面重新推导。

后向传播公式

(其中，$y_{trial,i} = A_0 + x_i * a_{out}(x_i, p)$)，

$$
\begin{aligned}
E(w,b) &= \sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i}) \right)^2 \\
&= \sum_i \left(a_{out} + x_i \cdot \frac{da_{out}}{dx} - f(x_i, y_{t,i}) \right)^2
\end{aligned} \tag{15}
$$

上面方程 $E(w,b)$ 是关于，$a_{out}, \frac{da_{out}}{dx}$ 的函数。另外，$f(x_i, y_{t,i})$ 是关于 x_i, a_{out} 的函数。$a_{out}, \frac{da_{out}}{dx}$ 包含了所有的参数 w, b。

代码中 ΔwH_2 计算公式如下：

$$\begin{aligned}
\Delta wH_2 = \frac{\partial E(w,b)}{\partial wH_2} &= \frac{\partial E(w,b)}{\partial a_{out}}\frac{\partial a_{out}}{\partial wH_2} + \frac{\partial E(w,b)}{\partial \frac{da_{out}}{dx}}\frac{\partial \frac{da_{out}}{dx}}{\partial wH_2} \\
=& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)\frac{\partial}{\partial a_{out}}\left(a_{out} + x_i \cdot \frac{da_{out}}{dx} - f(x_i, y_{t,i})\right)\frac{\partial a_{out}}{\partial wH_2} \\
&+ 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)\frac{\partial}{\partial \frac{da_{out}}{dx}}\left(a_{out} + x_i \cdot \frac{da_{out}}{dx} - f(x_i, y_{t,i})\right)\frac{\partial \frac{da_{out}}{dx}}{\partial wH_2} \\
=& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)\left(1 - \frac{\partial f(x_i, y_{t,i})}{\partial a_{out}}\right)\frac{\partial a_{out}}{\partial wH_2} \\
+& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)\left(x_i - \frac{\partial f(x_i, y_{t,i})}{\partial \frac{da_{out}}{dx}}\right)\frac{\partial \frac{da_{out}}{dx}}{\partial wH_2}
\end{aligned} \tag{16}$$

其中，$\frac{\partial f(x_i, y_{t,i})}{\partial \frac{da_{out}}{dx}} = 0$

后向传播公式

代码中 ΔwH_2 计算公式如下：

$$
\begin{aligned}
\Delta wH_2 = \frac{\partial E(w,b)}{\partial wH_2} &= \frac{\partial E(w,b)}{\partial a_{out}} \frac{\partial a_{out}}{\partial wH_2} + \frac{\partial E(w,b)}{\partial \frac{da_{out}}{dx}} \frac{\partial \frac{da_{out}}{dx}}{\partial wH_2} \\
=& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)\left(1 - \frac{\partial f(x_i, y_{t,i})}{\partial a_{out}}\right)\frac{\partial a_{out}}{\partial wH_2} \\
+& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)\left(x_i - \frac{\partial f(x_i, y_{t,i})}{\partial \frac{da_{out}}{dx}}\right)\frac{\partial \frac{da_{out}}{dx}}{\partial wH_2} \\
=& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)\left(1 - \frac{\partial f(x_i, y_{t,i})}{\partial a_{out}}\right)\frac{\partial a_{out}}{\partial wH_2} \\
+& 2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)(x_i - 0)\frac{\partial}{\partial wH_2}\left(\frac{da_{out}}{dx}\right)
\end{aligned} \tag{17}
$$

其中，$\frac{\partial f(x_i,y_{t,i})}{\partial \frac{da_{out}}{dx}} = \frac{\partial f(x_i,y_{t,i})}{\partial y_{t,i}} \frac{\partial y_{t,i}}{\partial \frac{da_{out}}{dx}} = 0$

后向传播公式

$$
\begin{aligned}
&\frac{\partial}{\partial w_{H2}} a_{out}(x_i, p) = \sum_{j=1,2,3} \frac{\partial zO_j}{\partial wH_2} = \sum_{j=1,2,3} wO_j \frac{\partial aH_j}{\partial wH_2} \\
&= \sum_{j=1,2,3} wO_j \frac{\partial \sigma(zH_j)}{\partial wH_2} = \sum_{j=1,2,3} wO_j * D\sigma(zH_j) \frac{\partial zH_j}{\partial wH_2} \\
&= \sum_{j=2} wO_j \cdot D\sigma(zH_j) \cdot x_i
\end{aligned}
$$

后向传播公式

$$\frac{da_{out}(x_i)}{dx} = \sum_j \left(wO_j \cdot D\sigma(zH_j) \frac{dzH_j}{dx} \right) = \sum_j wO_j \cdot D\sigma(zH_j) \cdot wH_j \quad (18)$$

$$\begin{aligned}
\frac{\partial}{\partial wH_2} \frac{da_{out}(x_i)}{dx} &= \sum_j \frac{\partial}{\partial wH_2} \left(wO_j \cdot D\sigma(zH_j) \cdot wH_j \right) \\
&= \sum_{j=2} \left(wO_j \cdot D\sigma(zH_j) \right) + \sum_j \left(wO_j \cdot D^2\sigma(zH_j) \cdot wH_j \cdot \frac{\partial zH_j}{\partial wH_2} \right) \\
&= \sum_{j=2} \left(wO_j \cdot D\sigma(zH_j) \right) + \sum_{j=2} \left(wO_j \cdot D^2\sigma(zH_j) \cdot wH_j \cdot x_i \right) \\
&= \sum_{j=2} \left(wO_j \cdot D\sigma(zH_j) + wO_j \cdot D^2\sigma(zH_j) \cdot wH_j \cdot x_i \right) \quad (19)
\end{aligned}$$

后向传播公式

代码中 ΔwH_2 计算公式如下：

$$
\begin{aligned}
\Delta wH_2 = \frac{\partial E(w,b)}{\partial wH_2} &= \frac{\partial E(w,b)}{\partial a_{out}}\frac{\partial a_{out}}{\partial wH_2} + \frac{\partial E(w,b)}{\partial \frac{da_{out}}{dx}}\frac{\partial \frac{da_{out}}{dx}}{\partial wH_2} \\
&=2\sum_i\left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)\left(1 - \frac{\partial f(x_i, y_{t,i})}{\partial y_{t,i}}\frac{\partial y_{t,i}}{\partial a_{out}}\right)\frac{\partial a_{out}}{\partial wH_2} \\
&+2\sum_i\left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right) x_i \cdot \frac{\partial}{\partial wH_2}\left(\frac{da_{out}}{dx}\right) \\
&=2\sum_i\left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)\left(1 - x_i \cdot \frac{\partial f(x_i, y_{t,i})}{\partial y_{t,i}}\right)\frac{\partial a_{out}}{\partial wH_2} \\
&+2\sum_i\left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right) x_i \cdot \frac{\partial}{\partial wH_2}\left(\frac{da_{out}}{dx}\right)
\end{aligned}
\tag{20}
$$

后向传播公式

代码中 ΔwH_2 计算公式如下：

$$\Delta wH_2 = \frac{\partial E(w,b)}{\partial wH_2} = \frac{\partial E(w,b)}{\partial a_{out}}\frac{\partial a_{out}}{\partial wH_2} + \frac{\partial E(w,b)}{\partial \frac{da_{out}}{dx}}\frac{\partial \frac{da_{out}}{dx}}{\partial wH_2}$$

$$=2\sum_i \left(\frac{dy_{t,i}}{dx} - f(x_i, y_{t,i})\right)\left(\left(1 - x_i\frac{\partial f(x_i, y_{t,i})}{\partial y_{t,i}}\right)\frac{\partial a_{out}}{\partial wH_2} + x_i\frac{\partial}{\partial wH_2}\frac{da_{out}}{dx}\right)$$

$$\frac{\partial}{\partial w_{H2}} a_{out}(x_i, p) = \sum_{j=2} wO_j \cdot D\sigma(zH_j) \cdot x_i$$

$$\frac{\partial}{\partial wH_2}\frac{da_{out}(x_i)}{dx} = \sum_j \frac{\partial}{\partial wH_2}\left(wO_j \cdot D\sigma(zH_j) \cdot wH_j\right)$$

$$= \sum_{j=2}\left(wO_j \cdot D\sigma(zH_j) + wO_j \cdot D^2\sigma(zH_j) \cdot wH_j \cdot x_i\right) \tag{21}$$

后向传播公式

$$
\begin{aligned}
\sigma'(x) =& \sigma(1-\sigma) \\
\sigma''(x) =& 2\sigma^3 - 3\sigma^2 + \sigma = \sigma(1-\sigma)(1-2\sigma)
\end{aligned}
\tag{22}
$$

后向传播公式

代码中 ΔwH_2 最后一项计算如下：

$$
\begin{aligned}
&\left(\left(1 - x_i \frac{\partial f(x_i, y_{t,i})}{\partial y_{t,i}}\right) \frac{\partial a_{out}}{\partial wH_2} + x_i \frac{\partial}{\partial wH_2} \frac{da_{out}}{dx}\right) \\
&= \left(1 - x_i \frac{\partial f(x_i, y_{t,i})}{\partial y_{t,i}}\right) \sum_{j=2} x_i \cdot wO_j \cdot D\sigma(zH_j) \\
&+ \sum_{j=2} \left(x_i \cdot wO_j \cdot D\sigma(zH_j) + x_i^2 \cdot wO_j \cdot D^2\sigma(zH_j) \cdot wH_j\right) \\
&= \left(1 - x_i \frac{\partial f(x_i, y_{t,i})}{\partial y_{t,i}}\right) \sum_{j=2} x_i \cdot wO_j \cdot D\sigma(zH_j) \\
&+ \sum_{j=2} (1 - x_i(1 - 2\sigma(zH_j)) \cdot wH_j) \cdot x_i \cdot wO_j \cdot D\sigma(zH_j)
\end{aligned} \tag{23}
$$

后向传播公式

代码中 ΔwH_2 计算公式如下：

$$
\begin{aligned}
\Delta wH_2 = \frac{\partial E(w,b)}{\partial wH_2} &= \frac{\partial E(w,b)}{\partial a_{out}}\frac{\partial a_{out}}{\partial wH_2} + \frac{\partial E(w,b)}{\partial \frac{da_{out}}{dx}}\frac{\partial \frac{da_{out}}{dx}}{\partial wH_2} \\
=& 2\sum_i\left(\frac{dy_{t,i}}{dx} - f(x_i,y_{t,i})\right)\left(\left(1 - x_i\frac{\partial f(x_i,y_{t,i})}{\partial y_{t,i}}\right)\frac{\partial a_{out}}{\partial wH_2} + x_i\frac{\partial}{\partial wH_2}\frac{da_{out}}{dx}\right) \\
=& 2\sum_i\left(\frac{dy_{t,i}}{dx} - f(x_i,y_{t,i})\right)\left(\left(1 - x_i\frac{\partial f(x_i,y_{t,i})}{\partial y_{t,i}}\right)\sum_{j=2} x_i\cdot wO_j\cdot D\sigma(zH_j)\right) \\
+& 2\sum_i\left(\frac{dy_{t,i}}{dx} - f(x_i,y_{t,i})\right)\left(\sum_{j=2}(1 - x_i(1-2\sigma)\cdot wH_j)\cdot x_i\cdot wO_j\cdot D\sigma(zH_j)\right)
\end{aligned} \tag{24}
$$

此处 $\sigma = \sigma(zH_j)$

谢谢!