

Homework 1 and 2

Question 1. 实现正向传播和反向传播的推导

在进行证明前，需要先定义一些符号：

表 1. 符号说明

符号含义	含义
$a^{(l)}$	第 l 层经过激活函数激活后的输出
$a^{(L)}$	最后一层输出，这时 $y = a^{(L)}$
x	输入数据，并且 $x = a^{(1)}$
$w_{ij}^{(l)}$	权重， l 代表第 l 层， i 代表第 l 层的第 i 个神经元， j 代表第 $l-1$ 层的第 j 个神经元
$b^{(l)}$	第 l 层的偏置
$z^{(l)}$	对第 $l-1$ 层的输出进行仿射变化得到结果，也就是 $z^{(l)} = w^{(l)}a^{(l-1)} + b^{(l)}$
$\sigma^{(l)}$	第 l 层的激活函数，并且有 $a^{(l)} = \sigma^{(l)}(z^{(l)})$
$\delta_j^{(l)}$	损失函数对于第 l 层第 j 个加权输出的偏导，也就是 $\delta_j^{(l)} = \frac{\partial E}{\partial z_j^{(l)}}$
t	输出的监督值
$E(y, t)$	损失函数，这里才用 L2 范数，也就是 $E(y, t) = \frac{\sum_k (y_k - t_k)^2}{2}$
θ	权重，就是 w 的总称
L	最后一层的神经网络的层数

正向传播和反向传播都是为了求解得到损失函数关于权重以及偏置的梯度，采用的都是链式法则。

(a) 反向传播公式推导：

由链式法则，可以得到：

$$(1a) \quad \left\{ \begin{array}{l} \frac{\partial E}{\partial w_{ij}^{(l)}} = \frac{\partial E}{\partial z^{(l)}} \frac{\partial z^{(l)}}{\partial w_{ij}^{(l)}} \\ \frac{\partial E}{\partial b_j^{(l)}} = \frac{\partial E}{\partial z^{(l)}} \frac{\partial z^{(l)}}{\partial b_j^{(l)}} \end{array} \right.$$

$$(1b) \quad \left\{ \begin{array}{l} \frac{\partial E}{\partial w_{ij}^{(l)}} = \frac{\partial E}{\partial z^{(l)}} \frac{\partial z^{(l)}}{\partial w_{ij}^{(l)}} \\ \frac{\partial E}{\partial b_j^{(l)}} = \frac{\partial E}{\partial z^{(l)}} \frac{\partial z^{(l)}}{\partial b_j^{(l)}} \end{array} \right.$$

对于 (1a) 和 (1b) 式子，可以化为：

$$(2a) \quad \left\{ \begin{array}{l} \frac{\partial E}{\partial w_{ij}^{(l)}} = \delta^{(l)} \frac{\partial z^{(l)}}{\partial w_{ij}^{(l)}} \\ \frac{\partial E}{\partial b_j^{(l)}} = \delta^{(l)} \frac{\partial z^{(l)}}{\partial b_j^{(l)}} \end{array} \right.$$

$$(2b) \quad \left\{ \begin{array}{l} \frac{\partial E}{\partial w_{ij}^{(l)}} = \delta^{(l)} \frac{\partial z^{(l)}}{\partial w_{ij}^{(l)}} \\ \frac{\partial E}{\partial b_j^{(l)}} = \delta^{(l)} \frac{\partial z^{(l)}}{\partial b_j^{(l)}} \end{array} \right.$$

所以，也就是求解 $\delta^{(l)}, \frac{\partial z^{(l)}}{\partial w_{ij}^{(l)}}, \frac{\partial z^{(l)}}{\partial b_j^{(l)}}$ 这三个变量。

对于该神经网络，假设我们使用的激活函数都为同一个激活函数，也就是 $\sigma = \sigma^{(l)}$ ，为了方便起见，后文中将激活函数写为： σ 。

首先先求解 $\delta^{(l)}$ ：

- 求解 $\delta^{(l)}$:

对于最后一层神经网络，我们从定义中知道：

$$a^{(L)} = \sigma(z^{(L)})$$

那么对于最后一层：

$$\delta^{(L)} = \frac{\partial E}{\partial z^{(L)}} = \frac{\partial \frac{\sum_k (y_k - a_k^{(L)})^2}{2}}{\partial a^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}}$$

也就是

$$\delta^{(L)} = \sum_k (a^{(L)} - y_k) \sigma'(z^{(L)})$$

当 $1 \leq l \leq L-1$ 时，我们这里采用链式法则来计算，由于是反向传播，所以我们选择用 $l+1$ 层的偏导来表示，也就是：

$$\delta^{(l)} = \frac{\partial E}{\partial z^{(l)}} = \frac{\partial E}{\partial z^{(l+1)}} \frac{\partial z^{(l+1)}}{\partial z^{(l)}} = \delta^{(l+1)} \frac{\partial z^{(l+1)}}{\partial z^{(l)}}$$

这是一个递推关系式，为了进一步求解，我们需要求解 $\frac{\partial z^{(l+1)}}{\partial z^{(l)}}$ 从定义中可以知道：

$$z^{(l+1)} = w^{(l+1)} a^{(l)} + b^{(l+1)} = w^{(l+1)} \sigma(z^{(l)}) + b^{(l+1)}$$

所以有：

$$\frac{\partial z^{(l+1)}}{\partial z^{(l)}} = w^{(l+1)} \sigma'(z^{(l)})$$

那么这样我们便可以得到 $\delta^{(l)}$ 的表达式为：

$$\delta^{(l)} = \sigma'(z^{(l)}) (w^{(l+1)})^T \delta^{(l+1)}$$

至此我们完成了第一步，也就是求解 $\delta^{(l)}$ 然后我们求解 $\frac{\partial z^{(l)}}{\partial w_{ij}^{(l)}}$, $\frac{\partial z^{(l)}}{\partial b_j^{(l)}}$ 这两个即可，便可以完成反向传播的推导。

- 求解 $\frac{\partial z^{(l)}}{\partial w_{ij}^{(l)}}$

从定义中可以知道：

$$z^{(l+1)} = w^{(l+1)} a^{(l)} + b^{(l+1)}$$

所以有：

$$\frac{\partial z^{(l)}}{\partial w_{ij}^{(l)}} = a^{(l-1)}$$

- 求解 $\frac{\partial z^{(l)}}{\partial b_j^{(l)}}$

从定义中可以知道：

$$z^{(l+1)} = w^{(l+1)} a^{(l)} + b^{(l+1)}$$

所以有：

$$\frac{\partial z^{(l)}}{\partial b_j^{(l)}} = 1$$

至此，我们便完成了反向传播的推导，也就是：

$$\begin{aligned}
(3a) \quad & \frac{\partial E}{\partial w_{ij}^{(l)}} = \delta^{(l)} a^{(l-1)} \\
(3b) \quad & \frac{\partial E}{\partial b_j^{(l)}} = \delta^{(l)} \\
(3c) \quad & \text{其中: } \delta^{(l)} = \begin{cases} \sigma'(z^{(l)})(w^{(l+1)})^T \delta^{(l+1)}, & l = 1 \dots L-1 \\ \sum_k (a^{(L)} - y_k) \sigma'(z^{(L)}), & l = L \end{cases}
\end{aligned}$$

(b) 正向传播公式推导:

所谓正向传播, 与上面不同的地方仅仅存在于关于 $\delta^{(l)}$ 求解过程中的链式法则使用上, 这里将使用 1-1 层的偏导来表示, 所以被称为正向传播。所以我们只需要改写其中关于 $\delta^{(l)}$ 的求解部分即可, 其他部分与反向传播相同。

- 对于第一层神经网络, 我们从定义中知道:

$$a^{(1)} = x$$

那么对于第一层:

$$\delta^{(L)} = \frac{\partial E}{\partial z^{(L)}} = \frac{\partial E}{\partial a^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}} = \frac{\partial E}{\partial x} \frac{\partial x}{\partial z^{(L)}}$$

由于 x 是常数, 所以 $\frac{\partial x}{\partial z^{(L)}} = 0$ 也就是:

$$\delta^{(L)} = 0$$

- 当 $2 \leq l \leq L$ 时, 我们这里采用链式法则来计算, 由于是正向传播, 所以我们选择用 1-1 层的偏导来表示, 也就是:

$$\delta^{(l)} = \frac{\partial E}{\partial z^{(l)}} = \frac{\partial E}{\partial z^{(l-1)}} \frac{\partial z^{(l-1)}}{\partial z^{(l)}} = \delta^{(l-1)} \frac{\partial z^{(l-1)}}{\partial z^{(l)}}$$

这是一个递推关系式, 为了进一步求解, 我们需要求解 $\frac{\partial z^{(l+1)}}{\partial z^{(l)}}$ 从定义中可以知道:

$$z^{(l+1)} = w^{(l+1)} a^{(l)} + b^{(l+1)} = w^{(l+1)} \sigma(z^{(l)}) + b^{(l+1)}$$

所以有:

$$z^{(l-1)} = \sigma^{-1}\left(\frac{z^{(l)} - b^{(l)}}{w^{(l)}}\right)$$

那么可以得到:

$$\frac{\partial z^{(l-1)}}{\partial z^{(l)}} = (\sigma^{-1}\left(\frac{z^{(l)} - b^{(l)}}{w^{(l)}}\right))'$$

那么这样我们便可以得到 $\delta^{(l)}$ 的表达式为:

$$\delta^{(l)} = \delta^{(l-1)} (\sigma^{-1}\left(\frac{z^{(l)} - b^{(l)}}{w^{(l)}}\right))'$$

至此, 我们便完成了正向传播的推导, 也就是:

$$\begin{aligned}
(4a) \quad & \frac{\partial E}{\partial w_{ij}^{(l)}} = \delta^{(l)} a^{(l-1)} \\
(4b) \quad & \frac{\partial E}{\partial b_j^{(l)}} = \delta^{(l)} \\
(4c) \quad & \text{其中: } \delta^{(l)} = \begin{cases} \delta^{(l-1)} (\sigma^{-1}\left(\frac{z^{(l)} - b^{(l)}}{w^{(l)}}\right))', & l = 2 \dots L \\ 0, & l = 1 \end{cases}
\end{aligned}$$