矩阵定义

$$X = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}, W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix}, B = \begin{bmatrix} b_1 & b_2 \end{bmatrix}, Y = \begin{bmatrix} y_1 & y_2 \end{bmatrix}, Z = \begin{bmatrix} z_1 & z_2 \end{bmatrix}$$

前向传播 ¶

$$y_1 = x_1 * w_{11} + x_2 * w_{21} + x_3 * w_{31} + b_1$$

$$y_2 = x_1 * w_{12} + x_2 * w_{22} + x_3 * w_{32} + b_2$$

$$z_1 = sigmoid(y_1)$$

$$z_2 = sigmoid(y_2)$$

$$Y = X * W + B = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} * \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix} + \begin{bmatrix} b_1 & b_2 \end{bmatrix} = \begin{bmatrix} y_1 & y_2 \end{bmatrix}$$

$$Z = sigmoid(Y)$$

如果Z不是输出层，那么Z是下一层的输入(X)，而且如果Z是输出层，一般不激活(直接输出Y)，或者理解为$(Z =$
用一个损失函数评价和标准值的距离，我们希望距离越小越好

$$损失函数(MSE) \quad loss = \sum_{i=1}^{n} \frac{(out_i - label_i)^2}{2}$$

可以理解为下面这个向量求和

$$\begin{bmatrix} loss_1 & loss_2 & \dots & loss_n \end{bmatrix} = \begin{bmatrix} out_1 - label_1 & out_2 - label_2 & \dots & out_n - label_n \end{bmatrix}$$

反向传播思路

$$目标求解每一层的\frac{\partial loss}{\partial w}和\frac{\partial loss}{\partial b}$$

$$迭代 \quad w - lr * \frac{\partial loss}{\partial w} \quad 和 \quad b - lr * \frac{\partial loss}{\partial b}$$

达到减少$loss$的效果

$$\frac{\partial loss}{\partial w_{ij}} = \frac{\partial loss}{\partial z_j} * \frac{\partial z_j}{\partial y_j} * \frac{\partial y_j}{\partial w_{ij}}$$

$$\frac{\partial loss}{\partial b_j} = \frac{\partial loss}{\partial z_j} * \frac{\partial z_j}{\partial y_j} * \frac{\partial y_j}{\partial b_j}$$

$$\frac{\partial loss}{\partial z_j^{-1}} = \frac{\partial loss}{\partial z_j} * \frac{\partial z_j}{\partial y_j} * \frac{\partial y_j}{\partial z_j^{-1}}$$

z_j^{-1}代表前一层的输出，即为这一层输入x_j，然后继续可以对前一层的误差进行求w，b，和前前一层的误差

$$\text{不激活时, 可以看做} z = y \quad \text{那么} \frac{\partial z_j}{\partial y_j} = 1$$

反向传播中的标量形式

$$\text{我们回顾一下，} loss = \sum_{i=1}^{n} \frac{(out_i - label_i)^2}{2} = \sum_{i=1}^{n} \frac{(z_i - label_i)^2}{2}$$

$$\frac{\partial loss}{\partial z_j} = z_j - label_j$$

$$\text{不激活的情况下} \frac{\partial z_j}{\partial y_j} = 1$$

$$\text{用} sigmoid \text{激活的情况下} \frac{\partial z_j}{\partial y_j} = (\ 1 - sigmoid(y_j)\ ) * sigmoid(\ y_j\ )$$

$$\text{这里说明一下函数} f(x) = sigmoid(x) \text{ 的导数为 } \frac{\partial f(x)}{\partial x} = sigmoid(x) * (1 - sigmoid(x))$$

$$\text{这样(以} sigmoid \text{激活为例子)} \frac{\partial loss}{\partial z_j} * \frac{\partial z_j}{\partial y_j} = (\ z_j - label_j\ ) * sigmoid(z_j) * (1 - sigmoid(z_j))$$

weight 和 bias 标量形式推导

$$\text{若计算} \frac{\partial loss}{\partial w_{ij}} = \frac{\partial loss}{\partial z_j} * \frac{\partial z_j}{\partial y_j} * \frac{\partial y_j}{\partial w_{ij}}, \text{只需求} \frac{\partial y_j}{\partial w_{ij}}$$

$$\text{若计算} \frac{\partial loss}{\partial b_j} = \frac{\partial loss}{\partial z_j} * \frac{\partial z_j}{\partial y_j} * \frac{\partial y_j}{\partial b_j}, \text{只需求} \frac{\partial y_j}{\partial b_j}$$

以上面为例

$$y_1 = x_1 * w_{11} + x_2 * w_{21} + x_3 * w_{31} + b_1$$

得到一般的算式

$$y_j = x_1 * w_{1j} + x_2 * w_{2j} \ + \ \dots \ + \ x_i * w_{ij} \ + \ \dots \ + \ x_n * w_{nj} + b_j$$

我们可以一并得到

$$\frac{\partial y_j}{\partial w_{ij}} = x_i$$

$$\frac{\partial y_j}{\partial b_j} = 1$$

所以得出

$$\frac{\partial loss}{\partial w_{ij}} = \frac{\partial loss}{\partial z_j} * \frac{\partial z_j}{\partial y_j} * \frac{\partial y_j}{\partial w_{ij}} = (\ z_j - label_j\ ) * sigmoid(z_j) * (1 - sigmoid(z_j)) * x_i \quad [1]$$

$$\frac{\partial loss}{\partial b_j} = \frac{\partial loss}{\partial z_j} * \frac{\partial z_j}{\partial y_j} * \frac{\partial y_j}{\partial b_j} = (\ z_j - label_j\ ) * sigmoid(z_j) * (1 - sigmoid(z_j)) * 1 \quad [2]$$

前一层误差推导

$$\text{这里推导}\frac{\partial loss}{\partial z_j^{-1}} = \frac{\partial loss}{\partial x_j}\text{其中前一层的输出就是这一层的输入}$$

$$loss = \sum_{i=1}^{n} \frac{(out_i - label_i)^2}{2} = \sum_{i=1}^{n} \frac{(z_i - label_i)^2}{2}$$

从刚才的正项传播:

$$y_1 = x_1 * w_{11} + x_2 * w_{21} + x_3 * w_{31} + b_1$$

$$y_2 = x_1 * w_{12} + x_2 * w_{22} + x_3 * w_{32} + b_2$$

$$\text{可以发现}x_i\text{和 }y_j,\ j \in 1,\ 2,\ 3,\ \ldots,n\ \text{ 有关，就是和}y\text{中每一项有关}$$

$$\text{与上边}weight\text{与}bias\text{不同,}weight\text{和}bias\text{仅与}y\text{中某一项有关}$$

$$y\text{到}z\text{是单射函数,所以}x_i\text{和 }z_j\ ,j \in 1,\ 2,\ 3,\ \ldots,n\ \text{ 有关,}$$

$$\text{如果计算 }\frac{\partial loss}{\partial z_j^{-1}} = \frac{\partial loss}{\partial z_j} * \frac{\partial z_j}{\partial y_j} * \frac{\partial y_j}{\partial z_j^{-1}}\text{则需要计算 }\frac{\partial y_j}{\partial z_j^{-1}}$$

求和

$$\text{注意，这里的}z_j^{-1}\text{就相当于这一层的}x_j$$

所以得到

$$\frac{\partial loss}{\partial z_j^{-1}} = \sum_{i=1}^{n} \frac{\partial loss}{\partial z_i} * \frac{\partial z_i}{\partial y_i} * \frac{\partial y_i}{\partial z_j^{-1}} = \sum_{i=1}^{n} \frac{\partial loss}{\partial z_i} * \frac{\partial z_i}{\partial y_i} * \frac{\partial y_i}{\partial x_j}$$

得到一般的算式

$$y_i = x_1 * w_{1i} + x_2 * w_{2i}\ +\ \ldots\ +\ x_j * w_{ji}\ +\ \ldots\ +\ x_m * w_{mi} + b_i$$

$$\frac{\partial y_i}{\partial z_j^{-1}} = w_{ji}$$

前一层误差推导

$$\text{这里推导}\frac{\partial loss}{\partial z_j^{-1}} = \frac{\partial loss}{\partial x_j}\text{其中前一层的输出就是这一层的输入}$$

$$loss = \sum_{i=1}^{n} \frac{(out_i - label_i)^2}{2} = \sum_{i=1}^{n} \frac{(z_i - label_i)^2}{2}$$

从刚才的正项传播:

$$y_1 = x_1 * w_{11} + x_2 * w_{21} + x_3 * w_{31} + b_1$$

$$y_2 = x_1 * w_{12} + x_2 * w_{22} + x_3 * w_{32} + b_2$$

$$\text{可以发现}x_i\text{和 }y_j,\ j \in 1,\ 2,\ 3,\ \ldots,n\ \text{ 有关，就是和}y\text{中每一项有关}$$

$$\text{与上边}weight\text{与}bias\text{不同,}weight\text{和}bias\text{仅与}y\text{中某一项有关}$$

$$y\text{到}z\text{是单射函数,所以}x_i\text{和 }z_j\ ,j \in 1,\ 2,\ 3,\ \ldots,n\ \text{ 有关,}$$

$$\text{如果计算 }\frac{\partial loss}{\partial z_j^{-1}} = \frac{\partial loss}{\partial z_j} * \frac{\partial z_j}{\partial y_j} * \frac{\partial y_j}{\partial z_j^{-1}}\text{则需要计算 }\frac{\partial y_j}{\partial z_j^{-1}}$$

求和

$$注意，这里的 z_j^{-1} 就相当于这一层的 x_j$$

所以得到

$$\frac{\partial loss}{\partial z_j^{-1}} = \sum_{i=1}^{n} \frac{\partial loss}{\partial z_i} * \frac{\partial z_i}{\partial y_i} * \frac{\partial y_i}{\partial z_j^{-1}} = \sum_{i=1}^{n} \frac{\partial loss}{\partial z_i} * \frac{\partial z_i}{\partial y_i} * \frac{\partial y_i}{\partial x_j}$$

得到一般的算式

$$y_i = x_1 * w_{1i} + x_2 * w_{2i} \ + \ \ldots \ + \ x_j * w_{ji} \ + \ \ldots \ + \ x_m * w_{mi} + b_i$$

$$\frac{\partial y_i}{\partial z_j^{-1}} = w_{ji}$$

$$\frac{\partial loss}{\partial z_j^{-1}} = \sum_{i=1}^{n} \frac{\partial loss}{\partial z_i} * \frac{\partial z_i}{\partial y_i} * \frac{\partial y_i}{\partial z_j^{-1}} = \sum_{i=1}^{n} \frac{\partial loss}{\partial z_i} * \frac{\partial z_i}{\partial y_i} * \frac{\partial y_i}{\partial x_j} = \sum_{i=1}^{n}(z_i - label_i) * sigmoid(z_i) * (1 - sigm$$

矩阵形式的梯度

我们转换为矩阵形式，我们一开始定义的矩阵如下:

$$X = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}, W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix}, B = \begin{bmatrix} b_1 & b_2 \end{bmatrix}, Y = \begin{bmatrix} y_1 & y_2 \end{bmatrix}, Z = \begin{bmatrix} z_1 & z_2 \end{bmatrix}$$

$$Y \ = \ XW \ + B$$

$$Z = sigmoid(Y)$$

$$\frac{\partial loss}{\partial Z} = \ Z - Label \ = \begin{bmatrix} z_1 - label_1 & z_2 - label_2 \end{bmatrix}$$

$$\frac{\partial z_j}{\partial y_j} = sigmoid(Y) * (1 \ - \ sigmoid(Y)) \quad * 为点乘(对应元素相乘)$$

$$\frac{\partial loss}{\partial W} = X^T \ @ \ \frac{\partial loss}{\partial Z} * \frac{\partial Z}{\partial Y} * \frac{\partial Y}{\partial W} = (\ Z - Label \ ) * sigmoid(Z) * (1 - sigmoid(Z)) \quad [4] \quad @ 为矩阵乘法$$

$$\frac{\partial loss}{\partial B} = \frac{\partial loss}{\partial Z} * \frac{\partial Z}{\partial Y} * \frac{\partial Y}{\partial B} = (\ Z - Label \ ) * sigmoid(Z) * (1 - sigmoid(Z)) \quad [5]$$

$$\frac{\partial loss}{\partial X} = \frac{\partial loss}{\partial Z} * \frac{\partial Z}{\partial Y} * \frac{\partial Y}{\partial B} = (\ Z - Label \ ) * sigmoid(Z) * (1 - sigmoid(Z)) \ @ \ W^T \quad [6]$$

推导 [4] 的细节

[4]式是由 [1] 式得出

$$不妨设 A = \frac{\partial loss}{\partial Z} * \frac{\partial Z}{\partial Y} = (\ Z - Label \ ) * sigmoid(Z) * (1 - sigmoid(Z))$$

由[1]可以得出

$$\begin{bmatrix} \frac{\partial loss}{\partial w_{11}} & \frac{\partial loss}{\partial w_{12}} & \ldots & \frac{\partial loss}{\partial w_{1n}} \\ \frac{\partial loss}{\partial w_{21}} & \frac{\partial loss}{\partial w_{22}} & \ldots & \frac{\partial loss}{\partial w_{2n}} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial loss}{\partial w_{m1}} & \frac{\partial loss}{\partial w_{m2}} & \ldots & \frac{\partial loss}{\partial w_{mn}} \end{bmatrix} = \begin{bmatrix} x_1 * a_1 & x_1 * a_2 & \ldots & x_1 * a_n \\ x_2 * a_1 & x_2 * a_2 & \ldots & x_2 * a_n \\ \ldots & \ldots & \ldots & \ldots \\ x_m * a_1 & x_m * a_2 & \ldots & x_m * a_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_m \end{bmatrix} \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix}$$

即为

$$\frac{\partial loss}{\partial W} = X^t @ A$$

[4]公式成立

推导 [6] 的细节

[6]式是由 [3] 式 得出

$$\text{不妨设}A = \frac{\partial loss}{\partial Z} * \frac{\partial Z}{\partial Y} = (\ Z - Label\ ) * sigmoid(Z) * (1 - sigmoid(Z))$$

$$\begin{bmatrix} \frac{\partial loss}{\partial x_1} & \frac{\partial loss}{\partial x_2} & \dots & \frac{\partial loss}{\partial x_m} \end{bmatrix} = \begin{bmatrix} w_{11} * a_1 + w_{12} * a_2 + \dots + x_{1n} * a_n & \dots & w_{m1} * a_1 + w_{m2} * a_2 + \dots + \end{bmatrix}$$

$$= \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \begin{bmatrix} w_{11} & w_{21} & \dots & w_{m1} \\ w_{12} & w_{22} & \dots & w_{m2} \\ \dots & \dots & \dots & \dots \\ w_{1n} & w_{2n} & \dots & w_{mn} \end{bmatrix}$$

即为

$$\frac{\partial loss}{\partial X} = A\ @\ W^T$$

[6]公式成立