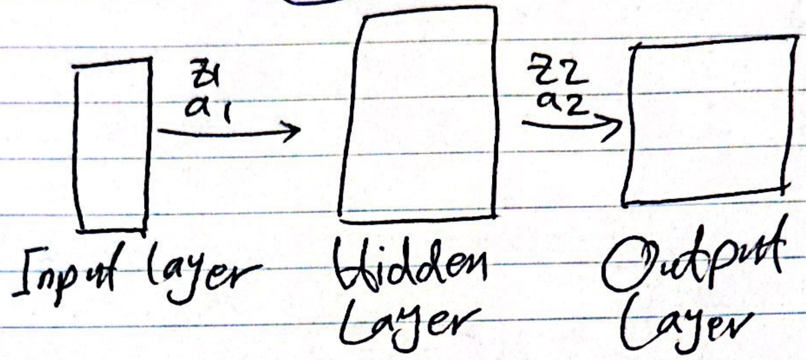
P1

assimilates in other weights and biases. So it is ok to ignore it!

$$L = \boxed{Loss\_MSE} = \boxed{\frac{1}{N}} \sum_{i=0}^{N} (y-a_2)^2$$



$z_1 = w_1 \cdot X + b_1$

$a_1 = \sigma(z_1)$ ↳ Sigmoid Activation Function

$z_2 = w_2 \cdot a_1 + b_2$ $a_2 = Id(z_2)$ ↳ Identity Function (Linear) for regression	$z_2 = w_2 \cdot a_1 + b_2$ $a_2 = \sigma(z_2)$ ↳ Sigmoid Function (for binary classification)

For updating w_1, w_2, b_1, b_2, we need:

$$\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \frac{\partial L}{\partial b_1}, \frac{\partial L}{\partial b_2}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial}{\partial w_2}\left[\sum (y-a_2)^2\right] =$$

can be assimilated in parameters (w_i, b_i)

$$\sum\left[\frac{\partial}{\partial w_2}(y-a_2)^2\right] = \sum\left[2(y-a_2)\frac{\partial}{\partial w_2}[y-a_2]\right]$$

Just ignore it!

$$= \sum\left[(y-a_2)\left(\cancel{\frac{\partial y}{\partial w_2}}_{\to 0} - \frac{\partial a_2}{\partial w_2}\right)\right] = \sum\left[-(y-a_2)\left(\frac{\partial a_2}{\partial w_2}\right)\right]$$

⟹ (next page)

Scanned with CamScanner

P2

$$\rightarrow \frac{\partial L}{\partial w_2} = \sum\left[(a_2-y)\boxed{\frac{\partial a_2}{\partial w_2}}\right]$$

$$\boxed{\frac{\partial a_2}{\partial w_2}} = \cancel{\frac{\partial a_2}{\partial z_2}} \cdot \underbrace{\frac{\partial z_2}{\partial w_2}}_{a_1} \Rightarrow$$

can be ignored now. ← constant

$$\boxed{\frac{\partial L}{\partial w_2} = \cancel{\sum}\left[(a_2-y)(a_1)\right]}$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial}{\partial b_2}\left[\sum (y-a_2)^2\right] = \sum\left[\frac{\partial}{\partial b_2}(y-a_2)^2\right] =$$

$$\sum\left[\cancel{2}(y-a_2)\left(\cancel{\frac{\partial y}{\partial b_2}} - \frac{\partial a_2}{\partial b_2}\right)\right] \Rightarrow$$

ignored

صفر

$$\frac{\partial L}{\partial b_2} = \sum\left[-(y-a_2)\boxed{\frac{\partial a_2}{\partial b_2}}\right]$$

$$\boxed{\frac{\partial a_2}{\partial b_2}} = \cancel{\frac{\partial a_2}{\partial z_2}} \cdot \underbrace{\frac{\partial z_2}{\partial b_2}}_{1} \Rightarrow$$

ignore

$$\boxed{\frac{\partial L}{\partial b_2} = \cancel{\sum}\left[(a_2-y)(1)\right] = \cancel{\sum}(a_2-y)}$$

Where activation function is linear (identity) and MSE as the loss function

Of course for the update rule, since we apply it elementwise, no Σ is existing.

P3

$$\frac{\partial L}{\partial w_1} = \frac{\partial}{\partial w_1}\left[\sum (y-a_2)^2\right] =$$

$$\sum\left[(y-a_2)\frac{\partial}{\partial w_1}(\overset{0}{\cancel{y}}-a_2)\right] = \sum (a_2-y)\boxed{\frac{\partial}{\partial w_1}(a_2)}$$

$$\boxed{\frac{\partial a_2}{\partial w_1}} = \cancel{\frac{\partial a_2}{\partial z_2}} \cdot \underbrace{\frac{\partial z_2}{\partial a_1}}_{w_2} \cdot \underbrace{\frac{\partial a_1}{\partial z_1}}_{g'(z_1)} \cdot \underbrace{\frac{\partial z_1}{\partial w_1}}_{X} \Rightarrow$$

$$\frac{\partial L}{\partial w_1} = \sum (a_2-y)\, w_2\, g'_{(z_1)}\, X \xrightarrow{\text{when } g_{(z_1)} = \operatorname{sig}(z_1)} g'_{(z_1)} = \underbrace{\operatorname{sig}_{(z_1)}}_{a_1}\underbrace{(1-\operatorname{sig}(z_1))}_{(1-a_1)}$$

$$\boxed{\frac{\partial L}{\partial w_1} = \sum (a_2-y) w_2\, a_1 (1-a_1)\, X}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial}{\partial b_1}\left(\sum (y-a_2)^2\right) = \sum\left((y-a_2)\frac{\partial}{\partial b_1}(\overset{0}{\cancel{y}}-a_2)\right)$$

$$\Rightarrow \frac{\partial L}{\partial b_1} = \sum\left[-(y-a_2)\boxed{\frac{\partial}{\partial b_1}(a_2)}\right]$$

$$\boxed{\frac{\partial a_2}{\partial b_1}} = \underbrace{\cancel{\frac{\partial a_2}{\partial z_2}}}_{1} \cdot \underbrace{\frac{\partial z_2}{\partial a_1}}_{w_2} \cdot \underbrace{\frac{\partial a_1}{\partial z_1}}_{g'(z_1)} \cdot \underbrace{\frac{\partial z_1}{\partial b_1}}_{1} \Rightarrow$$

$$\boxed{\frac{\partial L}{\partial b_1} = \sum (a_2-y) w_2\, a_1 (1-a_1)}$$

In the deriving process, I just ignored the constant coefficients, as they can be assimilated in the magnitudes of Ws, bs.

P4

update rule:

$$
\begin{cases}
w_1 = w_1 - lr \, \frac{\partial L}{\partial w_1} \\
w_2 = w_2 - lr \cdot \frac{\partial L}{\partial w_2} \\
b_1 = b_1 - \frac{\partial L}{\partial b_1} \\
b_2 = b_2 - \frac{\partial L}{\partial b_2}
\end{cases}
$$

For regression, I used the identity (linear) activation function in the output layer, while, for binary classification sigmoid activation function is used in the output layer. Thus, it results in a difference in the relation of a_2, z_2, and of course $\frac{\partial a_2}{\partial z_2}$.

Also, using different loss functions, makes the whole equation of L different.

Scanned with CamScanner

P5

(LogLoss) $L = -\left[(1-y)\log(1-a_2) + (y)\log(a_2)\right]$

$$\frac{\partial L}{\partial w_2} = -\left[(1-y)\frac{\partial}{\partial w_2}\{\log(1-a_2)\} + (y)\frac{\partial}{\partial w_2}\{\log a_2\}\right]$$

$$= -\left[\frac{1-y}{1-a_2}\cdot\frac{\partial(\overset{\text{zero } 0}{\cancel{1}}-a_2)}{\partial w_2} + \frac{y}{a_2}\cdot\frac{\partial a_2}{\partial w_2}\right] \Rightarrow$$

$$\frac{\partial L}{\partial w_2} = +\frac{1-y}{1-a_2}\cdot\frac{\partial a_2}{\partial w_2} - \frac{y}{a_2}\cdot\frac{\partial a_2}{\partial w_2} \Rightarrow$$

$$\left.\begin{aligned}\frac{\partial L}{\partial w_2} &= \left(\frac{1-y}{1-a_2} - \frac{y}{a_2}\right)\cdot\boxed{\frac{\partial a_2}{\partial w_2}} \\ \boxed{\frac{\partial a_2}{\partial w_2}} &= \frac{\partial a_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial w_2} = a_2(1-a_2)\cdot a_1\end{aligned}\right\} \Rightarrow$$

$$\frac{\partial L}{\partial w_2} = \underbrace{\left(\frac{1-y}{1-a_2} - \frac{y}{a_2}\right)}_{\frac{a_2-y}{\cancel{a_2(1-a_2)}}}\cdot\cancel{a_2(1-a_2)}\,a_1 \Rightarrow$$

$$\boxed{\frac{\partial L}{\partial w_2} = (a_2-y)\,a_1}$$

where activation function is σ, and Log as the Loss function.

Similarly:

$$\left.\begin{aligned}\frac{\partial L}{\partial b_2} &= \boxed{\frac{a_2-y}{a_2(1-a_2)}}\cdot\frac{\partial a_2}{\partial b_2} \\ \frac{\partial a_2}{\partial b_2} &= \frac{\partial a_2}{\partial z_2}\cdot\underbrace{\cancel{\frac{\partial z_2}{\partial b_2}}}_{1}\end{aligned}\right\} \Rightarrow \frac{\partial L}{\partial b_2} = \frac{a_2-y}{a_2(1-a_2)}\frac{\partial a_2}{\partial z_2}$$

Scanned with CamScanner

P6

$$\sigma : \frac{\partial a_2}{\partial z_2} = a_2(1-a_2) \Longrightarrow \frac{\partial L}{\partial b_2} = \frac{a_2 - y}{\cancel{a_2(1-a_2)}} \cdot \cancel{a_2(1-a_2)}$$

$$\Rightarrow \boxed{\frac{\partial L}{\partial b_2} = (a_2 - y)}$$

Similarly:

$$\frac{\partial L}{\partial w_1} = \frac{a_2 - y}{a_2(1-a_2)} \cdot \boxed{\frac{\partial a_2}{\partial w_1}}$$

$$\boxed{\frac{\partial a_2}{\partial w_1}} = \underbrace{\frac{\partial a_2}{\partial z_2}}_{g'(z_2)} \cdot \underbrace{\frac{\partial z_2}{\partial a_1}}_{w_2} \cdot \underbrace{\frac{\partial a_1}{\partial z_1}}_{g'(z_1)} \cdot \underbrace{\frac{\partial z_1}{\partial w_1}}_{X} \Bigg\} \Rightarrow$$

$$\frac{\partial L}{\partial w_1} = \frac{a_2 - y}{\cancel{a_2(1-a_2)}} \cdot \cancel{a_2(1-a_2)}\, w_2 (a_1)(1-a_1) X$$

$$\Rightarrow \boxed{\frac{\partial L}{\partial w_1} = (a_2 - y)\, w_2\, (a_1)(1-a_1)\, X}$$

Similarly:

$$\frac{\partial L}{\partial b_1} = \frac{a_2 - y}{a_2(1-a_2)} \cdot \boxed{\frac{\partial a_2}{\partial b_1}}$$

$$\boxed{\frac{\partial a_2}{\partial b_1}} = \underbrace{\frac{\partial a_2}{\partial z_2}}_{g'(z_2)} \cdot \underbrace{\frac{\partial z_2}{\partial a_1}}_{w_2} \cdot \underbrace{\frac{\partial a_1}{\partial z_1}}_{g'(z_1)} \cdot \underbrace{\frac{\partial z_1}{\partial b_1}}_{1} \Bigg\} \Rightarrow$$

$$\frac{\partial L}{\partial b_1} = \frac{a_2 - y}{\cancel{a_2(1-a_2)}} \cdot \cancel{a_2(1-a_2)} \cdot w_2 (a_1)(1-a_1)$$

$$\Rightarrow \boxed{\frac{\partial L}{\partial b_1} = (a_2 - y)\, w_2 (a_1)(1-a_1)}$$

Scanned with CamScanner