$$\frac{\partial \mathcal{L}}{\partial W^{[3]}} = -\Big[y^{(i)} \frac{\partial}{\partial W^{(3)}} \big(\log \underbrace{\sigma(W^{(3)}a^{(2)} + b^{(3)})}_{=a^3}\big)$$

$$+ (1-y^{(i)}) \frac{\partial}{\partial W^{(3)}} \log\big(1-\underbrace{\sigma(W^{(3)}a^{(2)}+b^{(3)})}_{=a^3}\big)\Big]$$

$$= -\Big[y^{(i)} \frac{1}{a^{(3)}} \cdot a^{(3)}(1-a^{(3)}) \cdot a^{(2)T} + (1-y^{(i)}) \cdot \frac{1}{(1-a^{(3)})} \cdot (-1) \cdot a^{(3)}(1-a^{(3)})\, a^{(2)T}\Big]$$

$$\underset{(1,2)}{\frac{\partial}{\partial W^{(3)}}} \underbrace{(W^{(3)}a^{(2)} + b^{(3)})}_{(1,1)} \rightarrow \underset{(1,2)}{\square}$$

$$= -\Big[y^{(i)}(1-a^{(3)}) \cdot a^{(2)T} + (1-y^{(i)})(-1)\, a^{(3)} a^{(2)T}\Big]$$

$$= -\Big[y^{(i)} a^{(2)T} - y^{(i)} a^{(3)} a^{(2)T} + y^{(i)} a^{(3)} a^{(2)T} - a^{(3)} a^{(2)T}\Big]$$

$$= -\Big[y^{(i)} a^{(2)T} - a^{(3)} a^{(2)T}\Big] = \boxed{-(y^{(i)} - a^{(3)})\, a^{(2)T}}$$

$$\frac{\partial J}{\partial W^{(3)}} = \frac{\partial}{\partial W^{(3)}} \frac{1}{m} \sum_{i=1}^{m} \mathcal{L} = -\frac{1}{m} \sum_{i=1}^{m} (y^{(i)} - a^{(3)})\, a^{(2)T}$$

$$\Rightarrow ① \; W^{(3)} := W^{(3)} - \frac{\partial}{\partial W^{(3)}} \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}$$

$$:= W^{(3)} - \frac{1}{m} \sum_{i=1}^{m} (y^{(i)} - a^{(3)})\, a^{(2)T}$$

$$\frac{\partial \mathcal{L}}{\partial W^{(2)}} = \boxed{\frac{\partial J}{\partial a^{(3)}} \times \frac{\partial a^{(3)}}{\partial z^{(3)}}} \times \frac{\partial z^{(3)}}{\partial a^{(2)}} \times \frac{\partial a^{(2)}}{\partial z^{(2)}} \times \frac{\partial z^{(2)}}{\partial W^{(2)}}$$

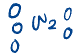
$$\frac{\partial}{\partial W^{(3)}} (\underbrace{W^{(3)}a^{(2)} + b^{(3)}}_{z^{(3)}}) = a^{(2)T} = \frac{\partial z^{(3)}}{\partial W^{(3)}}$$

$$\frac{\partial \mathcal{L}}{\partial W^{(3)}} = \boxed{\frac{\partial \mathcal{L}}{\partial a^{(3)}} \times \frac{\partial a^{(3)}}{\partial z^{(3)}}} \times \boxed{\frac{\partial z^{(3)}}{\partial W^{(3)}}} = \boxed{-(y^{(i)} - a^{(3)})}\boxed{a^{(2)T}}$$

$$\frac{\partial \mathcal{L}}{\partial W^{(2)}} = \underbrace{\frac{\partial J}{\partial a^{(3)}} \times \frac{\partial a^{(3)}}{\partial z^{(3)}}}_{-(y-a^{(3)})} \times \underbrace{\frac{\partial z^{(3)}}{\partial a^{(2)}}}_{W^{(3)T}} \times \underbrace{\frac{\partial a^{(2)}}{\partial z^{(2)}}}_{a^{(2)}(1-a^{(2)})} \times \underbrace{\frac{\partial z^{(2)}}{\partial W^{(2)}}}_{a^{(1)T}}$$

$$= (a^{(3)} - y) \times W^{(3)T} \times a^{(2)}(1-a^{(2)}) \times a^{(1)T}$$

$$\Rightarrow ② \; W^{(2)} := W^{(2)} - \frac{1}{m} \sum_{i=1}^{m} (a^{(3)} - y^{(i)})\, W^{(3)T} a^{(2)}(1-a^{(2)})\, a^{(1)T}$$

Connects 3 neurons to 2 neurons 

$$\underbrace{\frac{\partial \mathcal{L}}{\partial W^{(2)}}}_{(2,3)} = \boxed{\underbrace{(a^{(3)}-y)}_{(1,1)}\, \underbrace{W^{(3)T}}_{(2,1)} * \underbrace{a^{(2)}(1-a^{(2)})}_{(2,1)}\, \underbrace{a^{(1)T}}_{(1,3)}}$$

$*$: element-wise product

Shape doesn't match!

$$= \underbrace{\underbrace{\underbrace{W^{(3)T}}_{(2,1)} * \underbrace{a^{(2)}(1-a^{(2)})}_{(2,1)}}_{(2,1)}\, \underbrace{(a^{(3)}-y)}_{(1,1)}\, \underbrace{a^{(1)T}}_{(1,3)}}_{(2,3)}$$

$$\frac{\partial J}{\partial W^{(2)}} = \frac{1}{m}\sum_{i=1}^{m} W^{(3)T} * a^{(2)}(1-a^{(2)})(a^{(3)}-y)\, a^{(1)T}$$

$$\boxed{\frac{\partial \mathcal{L}}{\partial z^{(2)}}} \times \boxed{\frac{\partial z^{(2)}}{\partial W^{(2)}}} = \underbrace{\boxed{W^{(3)T} * a^{(2)}(1-a^{(2)})(a^{(3)}-y)}}_{(2,1)}\, \underbrace{\boxed{a^{(1)T}}}_{(1,3)}$$

$$\underbrace{\frac{\partial \mathcal{L}}{\partial W^{[1]}}}_{(3,n)} = \boxed{\frac{\partial \mathcal{L}}{\partial z^{[2]}}} \times \frac{\partial z^{[2]}}{\partial a^{[1]}} \times \frac{\partial a^{[1]}}{\partial z^{[1]}} \times \frac{\partial z^{[1]}}{\partial W^{[1]}}$$

$$= \underbrace{\boxed{\quad}}_{(2,1)} \times \underbrace{\underbrace{W^{[2]T}}_{(3,2)} * \underbrace{a^{[1]}(1-a^{[1]})}_{(3,2)}}_{(3,2)} \times \underbrace{X^T}_{(1,n)}$$

$$= \underbrace{\underbrace{W^{[2]T} * a^{[1]}(1-a^{[1]})}_{(3,2)} \times \underbrace{\boxed{\quad}}_{(2,1)} \times \underbrace{X^T}_{(1,n)}}_{(3,n)}$$

$$= W^{[2]T} * a^{[1]}(1-a^{[1]}) \times W^{(3)T} * a^{(2)}(1-a^{(2)})(a^{(3)}-y)$$

$$\frac{\partial J}{\partial W^{[1]}} = \frac{1}{m}\sum_{i=1}^{m} W^{[2]T} * a^{[1]}(1-a^{[1]}) \times W^{(3)T} * a^{(2)}(1-a^{(2)})(a^{(3)}-y)$$

$$\Rightarrow ③\ W^{(1)} := W^{(1)} - \frac{1}{m}\sum_{i=1}^{m} W^{[2]T} * a^{[1]}(1-a^{[1]}) \times W^{(3)T} * a^{(2)}(1-a^{(2)})(a^{(3)}-y)$$

$$\frac{\partial \mathcal{L}}{\partial b^{(3)}} = -\Big[y^{(i)} \frac{\partial}{\partial b^{(3)}} \Big(\log \underbrace{\sigma\big(w^{(3)} a^{(2)} + b^{(3)}\big)}_{=a^3}\Big)$$

$$+ (1-y^{(i)}) \frac{\partial}{\partial b^{(3)}} \log\Big(1-\underbrace{\sigma\big(w^{(3)} a^{(2)} + b^{(3)}\big)}_{=a^3}\Big) \Big]$$

$$= -\Big[y^{(i)} \frac{1}{a^{(3)}} a^{(3)} (1-a^{(3)}) \cdot 1 + (1-y^{(i)}) \cdot \frac{1}{(1-a^{(3)})} (-1)(a^{(3)})(1-a^{(3)}) \cdot 1 \Big]$$