$$a_0 = x$$
$$\left.\begin{aligned} z_i &= W_{i-1} a_{i-1} + b_{i-1} \\ a_i &= \sigma(z_i) \\ C &= \frac{1}{n} \| a_L - y \|_2^2 \end{aligned}\right\} i = 1, 2, 3, 4, \ldots, L$$

Numerator layout input output change

$$\left.\frac{\partial C}{\partial a_L} = \frac{2}{n}(a_L - y)^T\right\} \text{MSE}$$

$$\left.\frac{\partial C}{\partial z_i} = \frac{\partial C}{\partial a_i} \operatorname{diag}(\sigma'(z_i))\right\} \text{Element-wise activation}$$

$$\left.\begin{aligned} \frac{\partial C}{\partial a_{i-1}} &= \frac{\partial C}{\partial z_i} W_{i-1} \\ \frac{\partial C}{\partial b_{i-1}} &= \frac{\partial C}{\partial z_i} \\ \frac{\partial C}{\partial W_{i-1}} &= a_{i-1} \frac{\partial C}{\partial z_i} \end{aligned}\right\} \text{Linear Layer}$$

Denominator layout

$$\left.\frac{\partial C}{\partial a_L} = \frac{2}{n}(a_L - y)\right\} \text{MSE}$$

$$\left.\frac{\partial C}{\partial z_i} = \operatorname{diag}(\sigma'(z_i)) \frac{\partial C}{\partial a_i}\right\} \text{Element-wise activation}$$

$$\left.\begin{aligned} \frac{\partial C}{\partial a_{i-1}} &= W_{i-1}^T \frac{\partial C}{\partial z_i} \\ \frac{\partial C}{\partial b_{i-1}} &= \frac{\partial C}{\partial z_i} \\ \frac{\partial C}{\partial W_{i-1}} &= \frac{\partial C}{\partial z_i} a_{i-1}^T \end{aligned}\right\} \text{Linear layer}$$