$f_i \longrightarrow$ função de ativação,

$w^i \longrightarrow$ pesos da i-ésima camada,

$b^i \longrightarrow$ bias da i-ésima camada

$a^i \longrightarrow$ resultado intermediário da i-ésima camada.

$\ell \to$ última camada (last)

$$C(f,w,b) = \frac{1}{n}\sum_{i=0}^{n}\Big(\underbrace{f\big(w_i^\ell a_i^{(\ell-1)} + b_i^\ell\big)}_{a_i^\ell \,\to\, \text{resultado da última camada, para cada entrada } i} - \underbrace{y}_{\text{saída esperada}}\Big)^2$$

(somatório sobre os dados de entrada, n samples.)

$$\frac{\partial C^\ell}{\partial w} = \frac{1}{n}\sum_{i=0}^{n}\frac{\partial}{\partial w}\left(a_i^\ell - y\right)^2$$

$$= \frac{1}{n}\sum_{i=0}^{n} 2\left(a_i^\ell - y\right)\frac{\partial\left(a_i^\ell - y\right)}{\partial w} \quad \Big| \quad \frac{\partial a_i^\ell}{\partial w}, \text{ onde } a_i^\ell = f\big(\overbrace{w_i^\ell \cdot a_i^{(\ell-1)} + b_i^\ell}^{\alpha}\big)$$

$$= \frac{1}{n}\sum_{i=0}^{n} 2\left(a_i^\ell - y\right)\cdot f'(\alpha)\cdot a_i^{(\ell-1)} \quad \Big| \quad \frac{\partial a_i^\ell}{\partial w} = f'(\alpha)\cdot\frac{\partial \alpha}{\partial w} = \underline{f'(\alpha)\cdot a_i^{(\ell-1)}}$$

$$\frac{\partial C^\ell}{\partial b} = \frac{1}{n}\sum_{i=0}^{n} 2\left(a_i^\ell - y\right)\cdot f'(\alpha)$$

$$\frac{\partial C^\ell}{\partial a^{(\ell-1)}} = \frac{1}{n}\sum_{i=0}^{n} 2\left(a_i^\ell - y\right)\cdot f'(\alpha)\cdot w_\ell^\ell$$

$$C^{\ell-1} = \frac{1}{n}\sum_{i=0}^{n}\left(a_i^{\ell-1} - e_i^{\ell-1}\right)^2$$, onde $e_i^{\ell-1}$ é o valor esperado da penúltima camada.

O valor esperado de uma camada é o valor atual da camada sofrendo uma correção, que foi propagada pelas próximas camadas para trás.

$$e_i^{\ell-1} = a_i^{\ell-1} - \frac{\partial C^\ell}{\partial a^{\ell-1}}$$

Ou seja, é o valor atual menos o quanto ele contribui para o custo do valor na próxima camada.

$$C^{\ell-1} = \frac{1}{n}\sum_{i=0}^{n}\left(\cancel{a^{(\ell-1)}} - \left(\cancel{a^{(\ell-1)}} - \frac{\partial C^\ell}{\partial a^{(\ell-1)}}\right)\right)^2 = \frac{1}{n}\sum_{i=0}^{n}\left(\frac{\partial C^\ell}{\partial a^{(\ell-1)}}\right)^2$$

$$\frac{\partial C^{\ell-1}}{\partial w} = \frac{1}{n}\sum_{i=0}^{n} 2\left(\frac{\partial C^\ell}{\partial a^{(\ell-1)}}\right)\cdot\frac{\partial\left(a_i^{\ell-1} - e_i^{\ell-1}\right)}{w}$$

O valor esperado é uma constante em relação ao w, visto que o custo da camada posterior não depende do w desta camada (diretamente).

$$= \frac{1}{n}\sum_{i=0}^{n} 2\left(\frac{\partial C^\ell}{\partial a^{(\ell-1)}}\right)\cdot\frac{\partial a_i^{\ell-1}}{w} \to f\big(\underbrace{w_i^{\ell-1} a_i^{\ell-2} + b_i^{\ell-1}}_{\beta}\big) = \frac{\partial a_i^{\ell-1}}{\partial w} = f'(\beta)\cdot a_i^{\ell-2}$$

$$\frac{\partial C^{\ell-1}}{\partial b} = \frac{1}{n}\sum_{i=0}^{n} 2(\;''\;)\cdot f'(\beta), \quad \frac{\partial C^{\ell-1}}{\partial a^{\ell-2}} = \frac{1}{n}\sum_{i=0}^{n} 2(\;''\;)\cdot f'(\beta)\cdot w^{\ell-1}$$