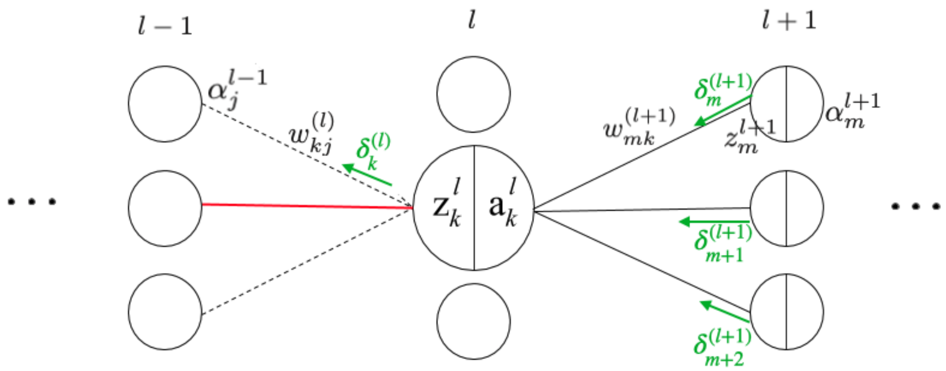

We will prove that the derivative of the cost function with respect to the weight $w_{kj}^{(l)}$ of the l^{th} hidden layer is $\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta w_{kj}^{(l)}} = \alpha_j^{(l-1)}\delta_k^{(l)} = \alpha_j^{(l-1)} \underbrace{f'(z_k^{(l)}) \sum_m \delta_m^{(l+1)} w_{mk}^{(l+1)}}_{\delta_k^{(l)}}$, where f is the activation function, i.e., $a_k^{(l)} = f(z_k^{(l)})$, and $\delta_m^{(l+1)}$ is the error propagated from layer $l+1$, to layer l, i.e., $\delta_m^{(l+1)} = \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta z_m^{(l+1)}}$.

In the following, we will assume zero bias term for the sake of simplicity.

$z_k^{(l)} = \sum_j w_{kj}^{(l)} \alpha_j^{(l-1)}$, $z_m^{(l+1)} = \sum_k w_{mk}^{(l+1)} \alpha_k^{(l)}$

$\alpha_k^{(l)} = f(z_k^{(l)})$

$z_m^{(l+1)} = \sum_k w_{mk}^{(l+1)} \alpha_k^{(l)}$

$$
\begin{aligned}
\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta w_{kj}^{(l)}} &= \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta z_k^{(l)}} \cdot \underbrace{\frac{\vartheta z_k^{(l)}}{\vartheta w_{kj}^{(l)}}}_{\alpha_j^{(l-1)}} \\
&= \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta \alpha_k^{(l)}} \cdot \underbrace{\frac{\vartheta \alpha_k^{(l)}}{\vartheta z_k^{(l)}}}_{f'(z_k^{(l)})} \cdot \alpha_j^{(l-1)} \\
&= \left(\sum_m \underbrace{\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta z_m^{(l+1)}}}_{\delta_m^{(l+1)}} \frac{\vartheta z_m^{(l+1)}}{\vartheta \alpha_k^{(l)}} \right) \cdot f'(z_k^{(l)}) \cdot \alpha_j^{(l-1)} \\
&= \left(\sum_m \delta_m^{(l+1)} w_{mk}^{(l+1)} \right) f'(z_k^{(l)}) \cdot \alpha_j^{(l-1)}
\end{aligned}
$$