0.1 Notations

$\odot$ is the Hadamart productsss

1 Forward propagation

$$x \longrightarrow h_1 \longrightarrow h_2 \longrightarrow \cdots \longrightarrow h_L \longrightarrow \hat{y}$$

$$a^l = f^l(o^l)$$

$$o^l = W^l a^{l-1} + b^l$$

Therefore, if we omit the bias terms, we have that

$$y = f^L(W^L f^{L-1}(W^{L-1}(\ldots)))$$

2 Backward propagation

2.1 How to compute the gradient slice of w_i^L recursively

$$\nabla \frac{c}{w^l} = (\sum_i \nabla \frac{a_i^l}{w^l})[\nabla \frac{c}{a^l}]^T \odot \frac{\overrightarrow{\partial o^L}}{\partial w_i^L} = \frac{\overrightarrow{\partial c}}{\partial a^L} \odot \frac{\overrightarrow{\partial a^L}}{\partial o^L} \odot \frac{\overrightarrow{\partial o^L}}{\partial w_i^L}$$

$$\frac{\overrightarrow{\partial c}}{\partial a^L} = \frac{\overrightarrow{\partial o^{L+1}}}{\partial a^L} \sum_i \frac{\partial c}{\partial o_i^{L+1}} = \overrightarrow{\sum_j w_{i,j}^{L+1}} \sum_i \frac{\partial c}{\partial o_i^{L+1}} = \overrightarrow{\sum_k w_{\forall,k}^{L+1} \frac{\partial c}{\partial o_k^{L+1}}}$$

$$\frac{\overrightarrow{\partial o^L}}{\partial w_i^L} = a^{L-1}$$

If we pose $\lambda = \frac{\partial c}{\partial o}$ and use the previous results we can obtain a compact recursive formulation of the gradient slice w_i^L

$$\frac{\overrightarrow{\partial c}}{\partial w_i^L} = \lambda^L a^{L-1},$$