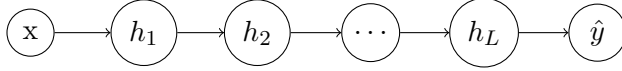


0.1 Notations

\odot is the Hadamard product

1 Forward propagation



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

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

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

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

2 Backward propagation

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

$$\begin{aligned} \nabla \frac{c}{w^l} &= (\sum_i \nabla \frac{a_i^l}{w^l}) [\nabla \frac{c}{a^l}]^T \odot \frac{\overrightarrow{\partial o^L}}{\partial w_i^L} = \frac{\overrightarrow{\partial c}}{\partial a^L} \odot \frac{\overrightarrow{\partial a^L}}{\partial o^L} \odot \frac{\overrightarrow{\partial o^L}}{\partial w_i^L} \\ \frac{\overrightarrow{\partial c}}{\partial a^L} &= \frac{\overrightarrow{\partial o^{L+1}}}{\partial a^L} \sum_i \frac{\partial c}{\partial o_i^{L+1}} = \sum_j \overrightarrow{w_{i,j}^{L+1}} \sum_i \frac{\partial c}{\partial o_i^{L+1}} = \sum_k \overrightarrow{w_{\nabla,k}^{L+1}} \frac{\partial c}{\partial o_k^{L+1}} \\ \frac{\overrightarrow{\partial o^L}}{\partial w_i^L} &= a^{L-1} \end{aligned}$$

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

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