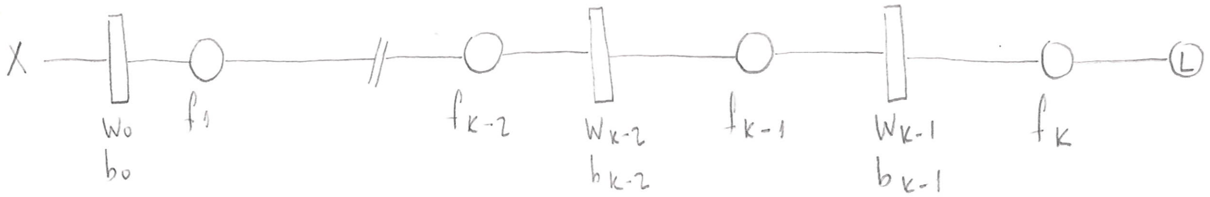
AMIA TF (TEÓRICO) CARLOS STRUMIA

Utilizo la notación de la bibliografía de referencia.



$$\begin{cases} f_0 = X \\ f_i = f(w_{i-1} y_{i-1} + b_{i-1}) \\ L = \frac{1}{2}(y - y_K)^2 \end{cases}$$

Solo hago el desarrollo para las últimas dos capas:

Backpropagation (capa de salida K)

$$\frac{\partial J}{\partial w_{K-1}} = \frac{\partial J}{\partial f_K}\frac{\partial f_K}{\partial w_{K-1}} = (y - y_K)\, f'(y_{K-1})\, y_{K-1}$$

$$\frac{\partial J}{\partial b_{K-1}} = \frac{\partial J}{\partial f_K}\frac{\partial f_K}{\partial b_{K-1}} = (y - y_K)\, f'(y_{K-1})$$

$$\frac{\partial J}{\partial y_{K-1}} = \frac{\partial J}{\partial f_K}\frac{\partial f_K}{\partial y_{K-1}} = (y - y_K)\, f'(y_{K-1})\, w_{K-1}$$

$(y - y_K) \in \mathbb{R}^{m_K}$ $\quad y_{K-1} \in \mathbb{R}^{m_{K-1}}$

$f'(y_{K-1}) \in \mathbb{R}^{m_K}$

$w_{K-1} \in \mathbb{R}^{m_K \times m_{K-1}}$

$\frac{\partial J}{\partial w_{K-1}} \in \mathbb{R}^{m_K \times m_{K-1}}$

$\frac{\partial J}{\partial b_{K-1}} \in \mathbb{R}^{m_K}$ $\quad \frac{\partial J}{\partial y_{K-1}} \in \mathbb{R}^{m_{K-1}}$

Capa (K-1)

$$\frac{\partial J}{\partial w_{K-2}} = \frac{\partial J}{\partial f_K}\frac{\partial f_K}{\partial f_{K-1}}\frac{\partial f_{K-1}}{\partial w_{K-2}} = (y - y_K).\, f'(y_{K-1})\, w_{K-1} .\, f'(y_{K-2})\, y_{K-2}$$

$$\frac{\partial J}{\partial b_{K-2}} = \frac{\partial J}{\partial f_K}\frac{\partial f_K}{\partial f_{K-1}}\frac{\partial f_{K-1}}{\partial b_{K-2}} = (y - y_K)\, f'(y_{K-1})\, w_{K-1}\, f'(y_{K-2})$$

$$\frac{\partial J}{\partial y_{K-2}} = \frac{\partial J}{\partial f_K}\frac{\partial f_K}{\partial f_{K-1}}\frac{\partial f_{K-1}}{\partial y_{K-2}} = (y - y_K)\, f'(y_{K-1})\, w_{K-1}\, f'(y_{K-2})\, w_{K-2}$$

$f'(y_{K-2}) \in \mathbb{R}^{m_{K-1}}$

$y_{K-2} \in \mathbb{R}^{m_{K-2}}$

$w_{K-2} \in \mathbb{R}^{m_{K-1} \times m_{K-2}}$

$\frac{\partial J}{\partial w_{K-2}} \in \mathbb{R}^{m_{K-1} \times m_{K-2}}$

$\frac{\partial J}{\partial b_{K-2}} \in \mathbb{R}^{m_{K-1}}$

$\frac{\partial J}{\partial y_{K-2}} \in \mathbb{R}^{m_{K-2}}$

Implementación (Layer.backwards)

Se define dY (out gradient). Para la última capa vale $dY = y - y_K$. Pero cuando hace la primer pasa, se actualiza dY con el valor que devuelve la función (dX).

Por lo tanto, se tiene:

$$dx_{j-1} = (y - y_K) \prod_{i=K-1}^{j-1} f'(y_i)\, w_i \qquad j = 1, \dots, K \qquad d_{j-1} \in \mathbb{R}^{m_{j-1}}$$

Se define dZ como:

$dZ_j = dY_j * \text{activación.df}(j)$ $\qquad dZ_j \in \mathbb{R}^{n_j}$

Con dZ podemos reescribir dX como:

$dX_{j-1} = W_{j-1}^T @ dZ_j \qquad dX_{j-1} \in \mathbb{R}^{n_{j-1}}$

Con dZ escribimos dW:

$dW_{j-1} = dZ_j @ y_{j-1}^T \quad (y_{j-1} = \text{last.input}) \quad dW_{j-1} \in \mathbb{R}^{n_j \times n_{j-1}}$

Por último, db_{j-1}:

$db_{j-1} = \text{sum}(dZ, \text{axis}=1)$ para que quede como vector columna $\mathbb{R}^{n_j}$