Assignment 3: Neural Network ~~By~~: By Hand!

PART 1: Derivation of Backpropagation.

recall: $\forall \quad l = 1, \dots, N^{layer}$,

- $Z^{(i,l)} = \left(W^{(l)}\right)^T a^{(i,l-1)} + b^{(l)}$ ⊛

- $a^{(i,l)} = f\left(Z^{(i,l)}\right)$ ⊛⊛

with,
$$\begin{cases} W^{(l)} = \left(w_{jk}\right)_{\substack{1 \le j \le n_{l-1} \\ 1 \le k \le n_l}} \; ; \quad W^{(l)} \in \mathbb{R}^{n_{l-1} \times n_l} \\ a^{(i,l-1)} \in \mathbb{R}^{n_{l-1}} \\ b^{(l)} \in \mathbb{R}^{n_l} \quad , \; b^{(l)} = \left(b_k^{(l)}\right)_{1 \le k \le n_l} \end{cases}$$

therefore, ~~$\forall$~~

$$\begin{cases} Z^{(i,l)} = \left(Z_k^{(i,l)}\right)_{1 \le k \le n_l} \in \mathbb{R}^{n_l} \quad \cancel{\text{with } Z_k^{(i,l)}} \quad \forall \; l = 1, \dots, N^{layer} \\ \text{with} \quad Z_k^{(i,l)} = \sum_{j=1}^{n_{l-1}} w_{kj}\, a_j^{(i,l-1)} + b_k^{(l)} \quad \forall \quad 1 \le k \le n_l \end{cases} \quad (1)$$

and

$$a^{(i,l)} = \left(a_k^{(i,l)}\right)_{1 \le k \le n_l} \in \mathbb{R}^{n_l} \quad \text{with} \quad a_k^{(i,l)} = f\left(Z_k^{(i,l)}\right) \quad (2)$$

Scanné avec CamScanner

Now let's derive the equations :

→ Equation (108): Error at the Last layer:

$$\delta^{(i,L)} = f'(Z^{(i,L)}) \odot L'(a^{(i,L)}, y^{(i)})$$

$\delta^{(i,L)} = \left(\delta_k^{(i,L)}\right)_{1 \leq k \leq m_L} \in \mathbb{R}^{m_L}$, with

$$\delta_k^{(i,L)} = f'(z_k^{(i,L)}) L'(a_k^{(i,L)}, y^{(i)}) \quad (3)$$

where, $z_k^{(i,L)}$ and $a_k^{(i,L)}$ are given by (1) and (2)

→ Equation (100): Derivative of the Cost at the Last layer:

$$\frac{\partial \, Cost}{\partial W^{(L)}} = \frac{1}{m} \sum_{i=1}^{m} a^{(i,L-1)} \otimes \delta^{(i,L)}$$

$$= \frac{1}{m} \sum_{i=1}^{m} a^{(i,L-1)} \left(\delta^{(i,L)}\right)^T.$$

$$\frac{\partial \, Cost}{\partial W^{(L)}} = \left(\frac{1}{m}\sum_{i=1}^{m} c_{jk}^{(i,L)}\right)_{\substack{1 \leq j \leq m_{L-1} \\ 1 \leq k \leq m_L}} \in \mathbb{R}^{m_{L-1} \times m_L}$$

with $c_{jk}^{(i,L)} = a_j^{(i,L-1)} \times \delta_k^{(i,L)}$ wher $\delta_k^{(i,L)}$ is given by (3)

Scanné avec CamScanner

→ Equation (110) :

$$\delta^{(i,L-1)} = f'(Z^{(i,L-1)}) \odot [W^{(L)} \delta^{(i,L)}]$$

$$\delta^{(i,L-1)} = \left(\delta_j^{(i,L-1)}\right)_{1 \leq j \leq m_{L-1}} \in \mathbb{R}^{m_{L-1}}$$

$$\text{with} \quad \delta_j^{(i,L-1)} = f(Z_j^{(i,L-1)}) \times \sum_{k=1}^{m_L} W_{jk}\, \delta_k^{(i,L)} \qquad (4)$$

→ Equation (111) :

$$\frac{\partial\, Cost}{\partial W^{(L-1)}} = \frac{1}{m} \sum_{i=1}^{m} a^{(i,L-2)} \otimes [\delta^{(i,L-1)}]$$

$$= \frac{1}{m} \sum_{i=1}^{m} a^{(i,L-2)} \left(\delta^{(i,L-1)}\right)^T$$

$$\frac{\partial\, Cost}{\partial W^{(L-1)}} = \left(\frac{1}{m} \sum_{i=1}^{m} C_{jk}^{(i,L-1)}\right)_{\substack{1 \leq j \leq m_{L-2} \\ 1 \leq k \leq m_{L-1}}} \in \mathbb{R}^{m_{L-2} \times m_{L-1}}$$

$$\text{with,} \quad C_{jk}^{(i,L-1)} = a_j^{(i,L-2)} \times \delta_k^{(i,L-1)}$$

where $\delta_k^{(i,L-1)}$ is given by (4).

Scanné avec CamScanner

→ Equation (112):

$$\delta^{(i,l)} = f'(Z^{(i,l)}) \odot [W^{(l+1)} \delta^{(i,l+1)}]$$

$$\delta^{(i,l)} = (\delta_j^{(i,l)})_{1 \leq j \leq n_l} \in \mathbb{R}^{n_l}$$

$$\text{with } \delta_j^{(i,l)} = f'(z_j^{(i,l)}) \sum_{k=1}^{n_{l+1}} w_{jk} \delta_k^{(i,l+1)}, \quad \forall 1 \leq j \leq n_l \quad (5)$$

→ Equation (113):

$$\frac{\partial Cost}{\partial W^{(l)}} = \frac{1}{m} \sum_{i=1}^{m} a^{(i,l-1)} \otimes [\delta^{(i,l)}]$$

$$= \frac{1}{m} \sum_{i=1}^{m} a^{(i,l-1)} (\delta^{(i,l)})^T$$

$$= \left(\frac{1}{m} \sum_{i=1}^{m} c_{jk}^{(i,l)} \right)_{\substack{1 \leq j \leq n_{l-1} \\ 1 \leq k \leq n_l}} \in \mathbb{R}^{n_{l-1} \times n_l}$$

$$\text{with, } c_{jk}^{(i,l)} = a_j^{(i,l-1)} \times \delta_k^{(i,l)}$$

Scanné avec CamScanner