Consider a network as defined in (3.1) and (3.2).

Assume that $n_L = 1$, find an algorithm to calculate $\nabla a^{[L]}(x)$.

Some element: $\{n_i\}_{i=1}^{L}$ where $n_i \in N$ for all $i = 1, 2, \cdots, L$ and $n_L = 1$. $W^{[i]} \in \mathbb{R}^{n_i \times n_{i-1}}$, $b^{[i]} \in \mathbb{R}^{n_i}$.

$a^{[1]} = x \in \mathbb{R}^{n_1}$ and $a^{[i]} \in \mathbb{R}^{n_i}$ for all $i = 1, 2, \cdots, L$.

by (3.2), we have $a^{[l]} = \sigma\left(W^{[l]} a^{[l-1]} + b^{[l]}\right) \in \mathbb{R}^{n_l}$ for $l = 1, 2, \cdots, L$.

Then we can construct a algorithm by following to get $\nabla a^{[L]}(x)$.

Input: $\{n_i\}_{i=1}^{L}$ in N, $x \in \mathbb{R}^{n_1}$, $\{W^{[i]}\}_{i=2}^{L}$ the collection of weight, $\{b^{[i]}\}_{i=2}^{L}$ the collection of bias, $\sigma(x)$.

Output: $\nabla a^{[L]}(x)$.

$a[1] = x$.

for i from 2 to L;

$\quad z[i] = W[i]\, a[i-1] + b[i]$

$\quad a[i] = \sigma(z[i])$.

$\delta[L] = \sigma'(z[L])$

for i from $L-1$ to 2;

$\quad \delta[i] = (W[i+1])^T \delta[i+1] \cdot \sigma'(z[i])$

$\nabla a^{[L]}(x) = (W^{[2]})^T \delta^{[2]}$. return.