1.

$\nabla a^{[L]}(x) = \left(\frac{\partial a^{[L]}}{\partial x_1} \cdots \frac{\partial a^{[L]}}{\partial x_{n_1}}\right)$, $z^{[1]} = (x_1, \ldots, x_{n_1})$

let $\delta^{[L]} = \frac{\partial a^{[L]}}{\partial z^{[L]}} = \sigma'(z^{[L]})$

then $\delta^{[l]} = \frac{\partial a^{[L]}}{\partial z^{[l]}} = \frac{\partial a^{[L]}}{\partial z^{[l+1]}} \cdot \frac{\partial z^{[l+1]}}{\partial z^{[l]}} = \sigma'(z^{[l]}) \circ \left[(W^{[l+1]})^T \delta^{[l+1]}\right]$, $z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$
$= W^{[l]} \sigma(z^{[l-1]}) + b^{[l]}$

prove it:

when $l = L-1$:

$\delta^{[L-1]} = \frac{\partial a^{[L]}}{\partial z^{[L-1]}} = \frac{\partial a^{[L]}}{\partial z^{[L]}} \cdot \frac{\partial z^{[L]}}{\partial z^{[L-1]}} = \delta^{[L]} \cdot \left(\frac{\partial z^{[L]}}{\partial z_1^{[L-1]}} \cdots \frac{\partial z^{[L]}}{\partial z_{n_{L-1}}^{[L-1]}}\right) = \delta^{[L]} \left(w_1^{[L]} \sigma'(z_1^{[L-1]}), \ldots, w_{n_{L-1}}^{[L]} \sigma'(z_{n_{L-1}}^{[L-1]})\right)$

$= \left(\sigma'(z_1^{[L-1]})\left[(W^{[L]})^T \delta^{[L]}\right]_1, \ldots, \sigma'(z_{n_{L-1}}^{[L-1]})\left[(W^{[L]})^T \delta^{[L]}\right]_{n_{L-1}}\right)$

$= \sigma'(z^{[L-1]}) \circ \left[(W^{[L]})^T \delta^{[L]}\right]$

Suppose when $l = k$:

$\delta^{[k]} = \sigma'(z^{[k]}) \circ \left[(W^{[k+1]})^T \delta^{[k+1]}\right]$ holds

when $l = k-1$:

$\delta^{[k-1]} = \frac{\partial a^{[L]}}{\partial z^{[k-1]}} = \frac{\partial a^{[L]}}{\partial z^{[k]}} \cdot \frac{\partial z^{[k]}}{\partial z^{[k-1]}} = \left(\delta_1^{[k]}, \ldots, \delta_{n_k}^{[k]}\right) \underbrace{\begin{bmatrix} \frac{\partial z_1^{[k]}}{\partial z_1^{[k-1]}} & \cdots & \frac{\partial z_1^{[k]}}{\partial z_{n_{k-1}}^{[k-1]}} \\ \vdots & & \vdots \\ \frac{\partial z_{n_k}^{[k]}}{\partial z_1^{[k-1]}} & & \frac{\partial z_{n_k}^{[k]}}{\partial z_{n_{k-1}}^{[k-1]}} \end{bmatrix}}_{n_k \times n_{k-1}}$

$= \left(\sum_{t=1}^{n_k} \delta_t^{[k]} w_{t1}^{[k]} \sigma'(z_1^{[k-1]}), \ldots, \sum_{t=1}^{n_k} \delta_t^{[k]} w_{t n_{k-1}} \sigma'(z_{n_{k-1}}^{[k-1]})\right)$

$= \left(\sigma'(z_1^{[k-1]})\left[(W^k)^T \delta^k\right]_1, \ldots, \sigma'(z_{n_{k-1}}^{[k-1]})\left[(W^k)^T \delta^k\right]_{n_{k-1}}\right)$

$= \sigma'(z^{[k-1]}) \circ (W^k)^T \delta^k$ also holds.

Thus, for $l = 1, \ldots, L-1$, $\delta^{[l]} = \sigma'(z^{[l]}) \circ (W^{[l+1]})^T \delta^{[l+1]}$

Algorithm:

① Forward pass:

- $a^{[1]} = x$
- For $l = 2, \ldots, L$:

$z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$, $a^{[l]} = \sigma(z^{[l]})$.

save each $z^{[l]}$ and $a^{[l]}$ for the backward pass.

② Backward pass:

- Compute $\delta^{[L]} = \frac{\partial a^{[L]}}{\partial z^{[L]}} = \sigma'(z^{[L]})$
- For $l = L, L-1, \ldots, 3, 2$:

$\delta^{[l-1]} = (W^{[l]})^T \delta^{[l]} \circ \sigma'(z^{[l-1]})$

- When we reach layer 1 (the input layer), note $a^{[1]} = x$ has no activation derivative. So

$\nabla_x a^{[L]}(x) = \delta^{[1]} = (W^{[2]})^T \delta^{[2]}$.

This produces $\nabla_x a^{[L]}(x) \in \mathbb{R}^{n_1}$.