1. Assumptions:

$$a^{[1]} = x \in \mathbb{R}^{n_1} \quad (3.1)$$

$$a^{[l]} = \sigma\left(W^{[l]} a^{[l-1]} + b^{[l]}\right) \in \mathbb{R}^{n_l} \text{ for } l = 1, 2, 3, \ldots, L$$

$$= \sigma\left(z^{[l]}\right) \quad (3.2)$$

Solution:

Let $C = \frac{1}{2}\| y - a^{[L]} \|_2^2$

Define $\delta_j^{[l]} = \frac{\partial C}{\partial z_j^{[l]}}$ for $1 \le j \le n_L$, $2 \le l \le L$

Use the Lemma 5.1, we have following results.

$$\delta^{[L]} = \frac{1}{2} \cdot 2 (y - a^{[L]})(-1) \circ \sigma'(z^{[L]})$$

$$= \sigma'(z^{[L]}) \circ (a^{[L]} - y) = \sigma'(z^{[L]}) \cdot (a^{[L]} - y)$$

$$\delta^{[l]} = \left[\frac{\partial C}{\partial z_1^{[l]}} \cdots \frac{\partial C}{\partial z_{n_l}^{[l]}}\right]^T, \quad 1 \le l \le L-1$$

$$= \frac{\partial C}{\partial z^{[l+1]}} \frac{\partial z^{[l+1]}}{\partial z^{[l]}}$$

$$= \sum_{k=1}^{n_{l+1}} \frac{\partial C}{\partial z_k^{[l+1]}} \frac{\partial z_k^{[l+1]}}{\partial z_j^{[l]}} \quad (\text{全微分})$$

$$= \sum_{k=1}^{n_{l+1}} \delta_k^{[l+1]} \frac{\partial z_k^{[l+1]}}{\partial z_j^{[l]}}$$

Since $z_k^{[l+1]} = \sum_{s=1}^{n_l} W_{ks}^{[l+1]} \sigma(z_s^{[l]}) + b_k^{[l+1]}$

$$\frac{\partial z_k^{[l+1]}}{\partial z_j^{[l]}} = w_{kj}^{[l+1]} \sigma'(z_j^{[l]})$$

Hence, $\delta_j^{[l]} = \sum_{k=1}^{n_{l+1}} \delta_k^{[l+1]} w_{kj}^{[l+1]} \sigma'(z_j^{[l]})$

$$= \sigma'(z_j^{[l]}) \left((W^{[l+1]})^T \delta^{[l+1]} \right)_j$$

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

$$\frac{\partial C}{\partial b_j^{[l]}} = \frac{\partial C}{\partial z_j^{[l]}} \frac{\partial z_j^{[l]}}{\partial b_j^{[l]}} = \delta_j^{[l]} \cdot 1 = \delta_j^{[l]}$$

$$\frac{\partial C}{\partial W_{jk}^{[l]}} = \sum_{s=1}^{n_l} \frac{\partial C}{\partial z_s^{[l]}} \frac{\partial z_s^{[l]}}{\partial W_{jk}^{[l]}} = \frac{\partial C}{\partial z_j^{[l]}} \cdot \frac{\partial z_j^{[l]}}{\partial W_{jk}^{[l]}} = \frac{\partial C}{\partial z_j^{[l]}} a_k^{[l-1]} = \delta_j^{[l]} a_k^{[l-1]}$$

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

Let $C = \frac{1}{2}\| y - a^{[L]}(x) \|_2^2$

$$\delta^{[L]} = \frac{\partial C}{\partial z^{[L]}} = \sigma'(z^{[L]}) \cdot (a^{[L]} - y)$$

$$\frac{\partial C}{\partial b_j^{[l]}} = \frac{\partial C}{\partial a^{[L]}(x)} \frac{\partial a^{[L]}}{\partial b_j^{[l]}} \Rightarrow \frac{\partial a^{[L]}}{\partial b_j^{[l]}} = \frac{1}{(a^{[L]}(x) - y)} \frac{\partial C}{\partial b_j^{[l]}}$$

$$\frac{\partial C}{\partial W_{jk}^{[l]}} = \frac{\partial C}{\partial a^{[L]}(x)} \frac{\partial a^{[L]}(x)}{\partial W_{jk}^{[l]}} \Rightarrow \frac{\partial a^{[L]}}{\partial W_{jk}^{[l]}} = \frac{1}{(a^{[L]}(x) - y)} \frac{\partial C}{\partial W_{jk}^{[l]}}$$

Hence,

$$\frac{\partial a^{[L]}}{\partial b_j^{[l]}} = \frac{1}{(a^{[L]}(x) - y)} \delta_j^{[l]}$$

$$= \frac{1}{(a^{[L]}(x) - y)} \left(\sigma'(z^{[l]}) \circ (W^{[l+1]})^T \delta^{[l+1]} \right)_j$$

$$= \frac{1}{a^{[L]}(x) - y} \prod_{s=l}^{L-1} \left(\sigma'(z^{[s]}) \circ (W^{[s+1]})^T \delta^{[s+1]} \right)_j$$

$$\frac{\partial a^{[L]}}{\partial W_{jk}^{[l]}} = \frac{1}{(a^{[L]}(x) - y)} \left(\delta_j^{[l]} a_k^{[l-1]} \right)$$

$$= \frac{1}{(a^{[L]}(x) - y)} \left(\sigma'(z^{[l]}) \circ (W^{[l+1]})^T \delta^{[l+1]} \right)_j a_k^{[l-1]}$$

$$= \frac{1}{(a^{[L]}(x) - y)} \prod_{s=l}^{L-1} \left(\sigma'(z^{[s]}) \circ (W^{[s+1]})^T \delta^{[s+1]} \right)_j a^{[s-1]}$$

So when we calculate the $\nabla a^{[L]}(x)$,

we need calculate the forward part, $z^{[l]}$, $a^{[l]}$

and calculate the backward part $\delta^{[L]}, \delta^{[L-1]} \ldots, \delta^{[l]}$

2. 在課堂上有提到 LWLR 的概念，是透過預測點附近的資料擬合一個線性迴歸，來預測該點的值，但是在時間序列的資料上，如果想要使用這個方法，那對於未來資料的預測就只能用當前資料來預測，更久之前的資料不會拿來使用，可能不太適用在有時序性且針對未來做預測的場景使用。