1、

$$a^{[1]} = x \in \mathbb{R}^{n_1}, \quad z^{[\ell]} = W^{[\ell]} a^{[\ell-1]} + b^{[\ell]}, \quad a^{[\ell]} = \sigma^{[\ell]}(z^{[\ell]}), \quad \ell = 2, \dots, L$$

let $u(x) = a^{[L]}(x) \in \mathbb{R}$, $\delta^{[\ell]} := \dfrac{\partial u}{\partial z^{[\ell]}} \in \mathbb{R}^{n_\ell}$

Step 1: Forward pass

Calculate $z^{[\ell]}, a^{[\ell]}$, $\ell = 2, \dots, L$

Step 2: Backward pass

① $\ell = L$

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

② $\ell = L-1, \dots, 2$

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

where $\circ$ is the componentwise product

Step 3: 輸出對輸入的梯度

$\because$ $z^{[2]} = W^{[2]} x + b^{[2]} \Rightarrow \dfrac{\partial z^{[2]}}{\partial x} = W^{[2]}$

$\therefore$ $\nabla a^{[L]}(x) = (W^{[2]})^T \delta^{[2]} \in \mathbb{R}^{n_1}$ ☆

2、

1、About Backpropagation：在不同 activation function 下，梯度傳播會有什麼差異？

2、About Approximation theory：若換成 ReLU or sigmoid，還能保證 universal approximation 嗎？

3、如果誤差分布不是 Gaussian，MSE 仍然適用嗎？

4、如何選擇 LWLR 中 weight 函數中的參數 τ

5、LWLR 在高維資料下是否仍然有效？