

1、

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

$$\text{let } u(x) = a^{[L]}(x) \in \mathbb{R}, \quad \delta^{[l]} := \frac{\partial u}{\partial z^{[l]}} \in \mathbb{R}^{n_l}$$

Step 1: Forward pass

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

Step 2: Backward pass

① $l=L$

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

② $l=L-1, \dots, 2$

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

where \circ is the componentwise product

Step 3: 輸出對輸入的梯度

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

$$\therefore \nabla a^{[2]}(x) = \left(\frac{\partial z^{[2]}}{\partial x} \right)^T \delta^{[2]} = (W^{[2]})^T \delta^{[2]} \in \mathbb{R}^{n_1}$$

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 LWR 中 weight 函數中的參數 τ
5. LWR 在高維資料下是否仍然有效?

Sorry 我上禮拜身體不舒服沒有去上課, 是根據老師上傳的 note 問的.