HW 2

1. Read Deep Learning: An Introduction for Applied Mathematicians. Consider a network as defined in (3.1) and (3.2). Assume that $n_L = 1$, find an algorithm to calculate $\nabla a^{[L]}(x)$.

2. There are unanswered questions during the lecture, and there are likely more questions we haven't covered. Take a moment to think about them and write them down here.

(3.1) $a^{[1]} = x \in \mathbb{R}^{n_1}$, (3.2) $a^{[l]} = \sigma(W^{[l]} a^{[l-1]} + b^{[l]}) \in \mathbb{R}^{n_l}$

for $l \in [2, 3, \dots L]$

1. $$\nabla a^{[L]}(x) = \left(\frac{\partial a^{[L]}}{\partial x}\right)^T$$

$$= \left(\frac{\partial a^{[L]}}{\partial a^{[L-1]}} \cdot \frac{\partial a^{[L-1]}}{\partial a^{[L-2]}} \cdots \frac{\partial a^{[2]}}{\partial a^{[1]}}\right)^T$$

Let $J_l = \dfrac{\partial a^{[l]}}{\partial a^{[l-1]}} = \mathrm{diag}(\sigma'(z^{[l]}))\, W^{[l]}$

$$\nabla a^{[L]}(x) = (J_L J_{L-1} \cdots J_2)^T = (J_2^T J_3^T \cdots J_L^T) \in \mathbb{R}^{n_1}$$

2. 問題：用神經網路擬合函數，神經元數量是否和誤差成反比