HW 2

1. $a^{[l]} = \sigma(w^{[l]} a^{[l-1]} + b^{[l]}) = \sigma(z^{[l]})$, for $l = 2, 3, \dots, L$

Assume $z^{[l]} \in \mathbb{R}^n$, $z^{[l]} = \begin{bmatrix} z_1^{[l]} \\ z_2^{[l]} \\ \vdots \\ z_n^{[l]} \end{bmatrix}$, $\sigma(z^{[l]}) = \begin{bmatrix} \sigma_1(z_1^{[l]}) \\ \vdots \\ \sigma_n(z_n^{[l]}) \end{bmatrix}$

$$\frac{\partial(\sigma(z^{[l]}))}{\partial z^{[l]}} = \begin{bmatrix} \frac{\partial \sigma_1}{\partial z_1^{[l]}} & \cdots & \frac{\partial \sigma_1}{\partial z_n^{[l]}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma_n}{\partial z_1^{[l]}} & \cdots & \frac{\partial \sigma_n}{\partial z_n^{[l]}} \end{bmatrix} = \begin{bmatrix} \sigma'(z_1^{[l]}) & & 0 \\ & \sigma'(z_2^{[l]}) \ddots & \\ 0 & & \sigma'(z_n^{[l]}) \end{bmatrix} = \mathrm{diag}(\sigma'(z^{[l]}))$$

Let $\mathrm{diag}(\sigma'(z^{[l]})) = D^{[l]}$

$$\nabla a^{[L]}(x) = \frac{\partial a^{[L]}}{\partial x} = \frac{\partial a^{[L]}}{\partial z^{[L]}} \cdot \frac{\partial z^{[L]}}{\partial a^{[L-1]}} \cdot \cdots \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial a^{[1]}}$$

$$= D^{[L]} w^{[L]} \cdot D^{[L-1]} w^{[L-1]} \cdots D^{[2]} w^{[2]}$$

Algorithm:

$a^{[1]} = x$

for i in range $(2, L+1, 1)$:

  $z^{[i]} = w^{[i]} a^{[i-1]} + b^{[i]}$, saved $z^{[i]}$

  $a^{[i]} = \sigma(z^{[i]})$, saved $a^{[i]}$

$A = 1$ # since $n_L = 1$

for j in range $(L, 1, -1)$:

  $D^{[j]} = \mathrm{daig}(\sigma'(z^{[j]}))$

  $A = A\, D^{[j]} w^{[j]}$

print A # A is what we want

2.

$$\mathrm{Cost}\left(W^{[2]}, W^{[3]}, W^{[4]}, b^{[2]}, b^{[3]}, b^{[4]}\right) = \frac{1}{10} \sum_{i=1}^{10} \frac{1}{2} \| y(x^{\{i\}}) - F(x^{\{i\}}) \|_2^2 .$$

平方

$\|v\|_2 = \sqrt{v_1^2 + \cdots + v_n^2}$

Here, the factor $\frac{1}{2}$ is included for convenience; it simplifies matters when we start differentiating. We emphasize that Cost is a function of the weights and biases—the

微分的用意是什麼？是為了找cost的最小值嗎