1.

$$z^{[l]} = w^{[l]} a^{[l-1]} + b^{[l]} \quad , \quad a^{[l]} = \sigma^{[l]}(z^{[l]})$$

$$\text{其} \quad \phi \quad , \quad a^{[0]} = x \quad , \quad n_L = 1$$

$$\frac{\partial a^{[l]}}{\partial a^{[l-1]}} = \frac{\partial a^{[l]}}{\partial z^{[l]}} \frac{\partial z^{[l]}}{\partial a^{[l-1]}} = \text{diag}\left(\sigma'^{[l]}(z^{[l]})\right) w^{[l]}$$

$$J^{[l]} = \frac{\partial a^{[l]}}{\partial a^{[l-1]}} = \text{diag}\left(\sigma'^{[l]}(z^{[l]})\right) w^{[l]}$$

define 任一層 L 的 row 向量 $g^{[l]}$

for $l < L$, $\dfrac{\partial a^{[l]}}{\partial a^{[l]}} = \dfrac{\partial a^{[l]}}{\partial a^{[l+1]}} \dfrac{\partial a^{[l+1]}}{\partial a^{[l]}}$

$$g^{[l]} = g^{[l+1]} J^{[l+1]}$$

$$g^{[l-1]} = g^{[l]} J^{[l]} = g^{[l]} \, diag\left(\sigma'^{[l]}(z^{[l]})\right) W^{[l]}$$

最後 1 層. $g^{[L]} = \dfrac{\partial a^{[L]}}{\partial a^{[L]}} = 1$

$$\dfrac{\partial a^{[L]}}{\partial a^{[\ell]}} = J^{[L]} J^{[L-1]} \cdots J^{[\ell+1]} = \prod_{t=\ell+1}^{L} J^{[t]}$$

$$J^{[t]} = \text{diag}\left(\sigma'^{[t]}(z^{[t]}) \right) W^{[t]}$$

2.

(1) 激活函取的選擇.如何影響反向傳播的穩定性

(2) 在深度網路中,梯度下降的收斂有什麼理論論證

(3) 當資料集很小時,如何防止 overfitting .