1.

define a loss function: $\ell(\theta) = \frac{1}{2}\left(y - h(x_1, x_2)\right)^2$

$\hat{\leq}$ $h(x_1, x_2) = \sigma(b + w_1 x_1 + w_2 x_2)$

$\theta^0 = (b, w_1, w_2)$

新參數: $\theta^{n+1} = \theta^n - \alpha \nabla \ell(\theta)$

$\nabla h = \left(\frac{\partial}{\partial b}\sigma, \frac{\partial}{\partial w_1}\sigma, \frac{\partial}{\partial w_2}\sigma\right)$

$$\nabla h = \begin{bmatrix} \sigma'(b + w_1 x_1 + w_2 x_2) \times 1 \\ \sigma'(b + w_1 x_1 + w_2 x_2) \times x_1 \\ \sigma'(b + w_1 x_1 + w_2 x_2) \times x_2 \end{bmatrix}$$

$$\theta^0 = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \rightarrow \theta' = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + \alpha(y - \sigma(21)) \begin{bmatrix} \sigma'(21) \\ \sigma'(21) \times 1 \\ \sigma'(21) \times 2 \end{bmatrix}$$

$(x_1, x_2, y) = (1, 2, 3)$

$b + w_1 x_1 + w_2 x_2 = 4 + 5 \times 1 + 2 \times 6 = 21$

$$\begin{bmatrix} \sigma'(z_3) \times 2 \\ \sigma'(z_2) \times 1 \\ \sigma'(z_1) \end{bmatrix} \left((\sigma'(z_1), \sigma'(z_2), \sigma'(z_3)) - y \right) + \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} = 0 \quad \therefore$$

$$\sigma'(z) = \sigma(z) \times (1 - \sigma(z))$$

$$\sigma(z) = \frac{1}{1+e^{-z}} \qquad \sigma' = \frac{e^{-z}}{(1+e^{-z})^2}$$

2. (a)

$$\sigma(x) = \frac{1}{1+e^{-x}} = s$$

for $k=1$

$$s' = -1(1+e^{-x})^{-2}(-e^{-x}) = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= s(1-s)$$

for $k=2$

$$s'' = [s(1-s)]' = s'(1-s) + s(-s')$$

$$= s'(1-s-s) = s'(1-2s) = s(1-s)(1-2s)$$

for $k=3$

$$\sigma''(x) = \frac{ds}{dx} \cdot \frac{d}{ds}\left[s(1-s)(1-2s) \right]$$

$$= \frac{ds}{dx} \cdot \frac{d}{ds}\left(s - 3s^2 + 2s^3 \right)$$

$$= s(1-s) \cdot \left(1 - 6s + 6s^2\right)$$

2. (b)

$$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^{\frac{x}{2}}}{e^{\frac{x}{2}} + e^{-\frac{x}{2}}}$$

$$\tanh\left(\frac{x}{2}\right) = \frac{e^{\frac{x}{2}} - e^{\frac{x}{2}}}{e^{\frac{x}{2}} + e^{\frac{x}{2}}}$$

$$\sigma(x) = \frac{1}{2}\left(1 + \tanh\left(\frac{x}{2}\right)\right)$$

$$\tanh(x) = 2\sigma(2x) - 1$$

3.

1. 如何選擇 η or α (學習率)

2. 如何選擇 batch size

3. 6121 會不會導致模型學不到東西