1. $Loss(b, w_1, w_2) = (y - h(x_1, x_2))^2$

$= (3 - h(1, 2))^2$

$= (3 - \sigma(b + w_1 + 2w_2))^2$

$\theta' = \theta^0 - \alpha \nabla_\theta Loss$

$= \theta^0 - 2\alpha(3 - \sigma(b + w_1 + 2w_2))\sigma'(b + w_1 + 2w_2)\nabla_\theta(b + w_1 + 2w_2)$

$= (4, 5, 6) - 2\alpha(3 - \sigma(4 + 5 + 2\times 6))\sigma'(4 + 5 + 2\times 6)(1, 1, 2)$

$= (4, 5, 6) - 2\alpha(3 - \sigma(21))\sigma'(21)(1, 1, 2)$

2.

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

$\frac{d^2\sigma}{dx^2} = \sigma'(x)(1 - \sigma(x)) - \sigma(x)\sigma'(x)$

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

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

$\frac{d^3\sigma}{dx^3} = \frac{d}{dx}\sigma'(x)(1 - 2\sigma(x))$

$= \sigma''(x)(1 - 2\sigma(x)) - 2\sigma'(x)^2$

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

$= \sigma(x)(1 - \sigma(x))(1 - 4\sigma(x) + 4\sigma(x)^2 - 2\sigma(x) + 2\sigma(x)^2)$

$= \sigma(x)(1 - \sigma(x))(6\sigma(x)^2 - 6\sigma(x) + 1)$

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

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

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

3. How to decide appropriate learning rate?