1、 $\theta^1 = \theta^0 - \alpha \nabla_\theta Loss$

$= \theta^0 + 2\alpha ((y - h(x_1, x_2)) \nabla_\theta h)$

$= \theta^0 + 2\alpha ((3 - \sigma(4 + 5x_1 + 6x_2)) \nabla_\theta h)$

$= \theta^0 + 2\alpha (3 - \sigma(21)) \left(\frac{\partial h}{\partial b}, \frac{\partial h}{\partial w_1}, \frac{\partial h}{\partial w_2}\right)$

$= (4, 5, 6) + 2\alpha (3 - \sigma(21)) \sigma'(21) (1, x_1, x_2)$

$y = (1, 1, 2)$

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

$k=1$: $\frac{d}{dx}\sigma = \frac{d\left(\frac{1}{1+e^{-x}}\right)}{d(1+e^{-x})} \cdot \frac{d(1+e^{-x})}{dx}$

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

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

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

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

$= \sigma'(x) \times (1 - \sigma(x)) + \sigma(x) \times (-\sigma'(x))$

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

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

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

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

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

(b) $\tanh(x) = \dfrac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$

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

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

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

3. What's the use of the relation between sigmoid function and hyperbolic function in Machine Learning?