1. $\theta^1 = (\theta^0)^T - \frac{1}{2} \cdot \nabla (y - h(x_1, x_2))^2 \Big|_{\begin{bmatrix} x_1 \\ x_2 \\ y \end{bmatrix}} \eta$ where η is learning rate

find $\nabla (y - h(x_1, x_2))^2$:

$$\nabla (y - h(x_1, x_2))^2 = \left(\frac{\partial (y - h(x_1, x_2))^2}{\partial b}, \frac{\partial (y - h(x_1, x_2))^2}{\partial w_1}, \frac{\partial (y - h(x_1, x_2))^2}{\partial w_2} \right)$$

$$= 2(h(x_1, x_2) - y) \cdot (\sigma'(z), \sigma'(z) x_1, \sigma'(z) x_2) \text{ where } z = b + w_1 x_1 + w_2 x_2 = 4 + 5 + 12 = 21$$

$$= 2(\sigma(21) - 3) \cdot (\sigma(21)(1 - \sigma(21)), \quad \sigma(21)(1 - \sigma(21)), \sigma(21)(1 - \sigma(21)) \times 2)$$

Thus, $\theta^1 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}^T - \eta (\sigma(21) - 3) \sigma(21)(1 - \sigma(21)) \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}^T$

2.

(a)

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

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

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

$$= \frac{e^{-x}}{(1 + e^{-x})^3} \cdot e^{-x} - \frac{e^{-x}}{(1 + e^{-x})^3}$$

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

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

$$- 2\sigma(x)\sigma'(x)(1 - \sigma(x)) + [\sigma(x)]^2 \sigma'(x)$$

$$= \sigma'(x) \left[(1 - \sigma(x))^2 - 4(1 - \sigma(x))\sigma(x) + [\sigma(x)]^2 \right]$$

$$= \frac{e^{-x}}{(1 + e^{-x})^2} \left[\frac{e^{-2x}}{(1 + e^{-x})^2} - 4 \cdot \frac{e^{-x}}{(1 + e^{-x})^2} + \frac{1}{(1 + e^{-x})^2} \right]$$

(b)

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad \sigma(x) = \frac{1}{1 + e^{-x}}$$

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

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

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

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

Thus, $\tanh(x) = 2 \cdot \sigma(2x) - 1$ □