1. $\theta^0 = (b, w_1, w_2) = (4,5,6)$, $\sigma(z) = \frac{1}{1+e^{-z}}$, $\sigma'(z) = \sigma(z)(1-\sigma(z))$

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

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

$\Rightarrow \theta^1 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + 2\alpha(3 - \sigma(21)) \cdot \sigma'(z) \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$

$\Rightarrow \begin{bmatrix} b^1 \\ w_1^1 \\ w_2^1 \end{bmatrix} = \begin{bmatrix} 4 + 2\alpha(3-\sigma(21))\sigma(21)(1-\sigma(21)) \\ 5 + 2\alpha(3-\sigma(21))\sigma(21)(1-\sigma(21)) \\ 6 + 2\alpha(3-\sigma(21))\sigma(21)(1-\sigma(21)) \cdot 2 \end{bmatrix}$ #

2. $\frac{d^k \sigma}{dx^k}(x)$

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

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

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

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

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

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

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

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

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

$= \sigma(x)(1-\sigma(x)) - 6\sigma(x)^2(1-\sigma(x)) + 6\sigma(x)^3(1-\sigma(x))$

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

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

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

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

3. The formulas mentioned in class have some similarity to ones I learned in PDE. There might be some connection between them, but I don't have enough understanding of this lesson. I need more time to figure them out.