1.

MSE: $L = \frac{1}{2}|y-h|^2$

Let $z = b + w_1x_1 + w_2x_2$ and $s = \sigma(z)$

Then, $\frac{\partial L}{\partial b} = (s-y)s(1-s)$, $\frac{\partial L}{\partial w_1} = (s-y)s(1-s)x_1$, $\frac{\partial L}{\partial w_2} = (s-y)s(1-s)x_2$.

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

Given that $(x_1, x_2, y) = (1,2,3)$ and $\theta^0 = (b, w_1, w_2) = (4,5,6)$,

$z = 4 + 5x_1 + 6x_2$, $s = \sigma(21)$.

So, $\theta' = (4,5,6) - \alpha(s-3)s(1-s)\cdot(1,1,2)$, where $s = \sigma(21)$.

2.

(a)

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

$\sigma''(x) = \frac{d}{dx}(\sigma(x)(1-\sigma(x))) = \sigma'(x)(1-\sigma(x)) + \sigma(x)\cdot\frac{d}{dx}(1-\sigma(x))$

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

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

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

$= \sigma(x)(1-\sigma(x))(1-4\sigma(x)+4\sigma(x)^2-2\sigma(x)+2\sigma(x)^2) = \sigma(x)(1-\sigma(x))(1-6\sigma(x)+6\sigma(x)^2)$.

(b)

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

3.

In GD algorithm, what role does the learning rate α playing in the update, and how might its value affect convergence?