1. $(x_1, x_2, y) = (1,2,3)$, $\theta^0 = (b, w_1, w_2) = (4,5,6)$, let $z = b + w_1 x_1 + w_2 x_2$

$$h(x_1, x_2) = \sigma(b + w_1 x_1 + w_2 x_2) = \sigma(z) = \frac{1}{1+e^{-z}}$$

$$L(b, w_1, w_2) = (y-h)^2$$

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

$$\nabla_\theta L = \left(\frac{\partial L}{\partial b}, \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}\right)$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial z} \cdot \frac{\partial z}{\partial b} = 2(y-h)\cdot(-1)\cdot \frac{e^{-z}}{(1+e^{-z})^2} \cdot 1 = 2\left(\frac{1}{1+e^{-z}} - 3\right) \cdot \frac{e^{-z}}{(1+e^{-z})^2} \cdot 1$$

$$\frac{\partial L}{\partial w_1} = 2(h-y)\frac{e^{-z}}{(1+e^{-z})^2} \cdot x_1 = 2\left(\frac{1}{1+e^{-z}} - 3\right) \cdot \frac{e^{-z}}{(1+e^{-z})^2} \cdot x_1$$

$$\frac{\partial L}{\partial w_2} = 2\left(\frac{1}{1+e^{-z}} - 3\right) \cdot \frac{e^{-z}}{(1+e^{-z})^2} \cdot x_2$$

$$\theta' = (4,5,6) - \alpha\left(\frac{\partial L}{\partial b}, \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}\right)$$

$$= \left(4 - 2\alpha\left(\frac{1}{1+e^{-21}} - 3\right)\cdot\frac{e^{-21}}{(1+e^{-21})^2},\ 5 - 2\alpha\left(\frac{1}{1+e^{-21}} - 3\right)\cdot\frac{e^{-21}}{(1+e^{-21})^2},\ 6 - 4\alpha\left(\frac{1}{1+e^{-21}} - 3\right)\cdot\frac{e^{-21}}{(1+e^{-21})^2}\right)_{\#}$$

2. (a) $\sigma(x) = \dfrac{1}{1+e^{-x}}$, $k=1$

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

$k=2$,

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

$$= \sigma(x)(1-\sigma(x))(1-2\sigma(x))_{\#}$$

$k=3$,

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

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

$$= \sigma(x)(1-\sigma(x))\left(1 - 6\sigma(x) + 6[\sigma(x)^2]\right)_{\#}$$

(b) simoid: $\sigma(x) = \dfrac{1}{1+e^{-x}}$, hyperbolic: $\tanh(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$

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

$$= 1 - 2\cdot(1-\sigma(2x)) = 2\sigma(2x) - 1_{\#}$$

3. To ensure a positive prediction, we can introduce a sigmoid activation function

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

and use assume that

$$h(x_1, x_2) = \sigma(b + w_1 x_1 + w_2 x_2).$$

為什麼要用sigmoid函數，Sigmoid函數有什麼特性及好處？

For now, let's take the choice of g as given. Other functions that smoothly increase from 0 to 1 can also be used, but for a couple of reasons that we'll see later (when we talk about GLMs, and when we talk about generative learning algorithms), the choice of the logistic function is a fairly natural one. Before

節錄自Stanford CS229 lecture notes, 後續會解釋，目前還沒有看到原因