1. Consider stochastic gradient descent method to learn the house price model

$$h(x_1, x_2) = \sigma(b + w_1x_1 + w_2x_2),$$

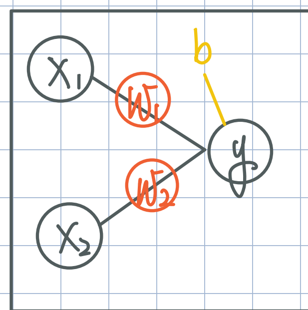
where σ is the sigmoid function.

Given one single data point $(x_1, x_2, y) = (1, 2, 3)$, and assuming that the current parameter is $\theta^0 = (b, w_1, w_2) = (4, 5, 6)$, evaluate θ^1.

> Just write the expression and substitute the numbers; no need to simplify or evaluate.

<sol>

$\theta^1 = \theta^0 - \eta \nabla_\theta L(\theta^0)$



For $\theta^0 = (b, W_1, W_2) = (4,5,6)$, $z^0 = 4 + 5\times 1 + 6 \times 2 = 21$

then, $y = \sigma(21)$, Moreover, $\sigma'(21) = \sigma(21)(1-\sigma(21))$

① $b^1 = 4 - \eta(\sigma(21)-3)\sigma(21)(1-\sigma(21))$;

② $W_1^{[1]} = 5 - \eta(\sigma(21)-3)\sigma(21)(1-\sigma(21))\cdot 1$;

③ $W_1^{[2]} = 6 - \eta(\sigma(21)-3)\sigma(21)(1-\sigma(21))\cdot 2$, where η is learning rate

Therefore, $\theta^1 = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} - \eta(\sigma(21)-3)(1-\sigma(21))\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ ∎

2. (a) Find the expression of $\frac{d^k}{dx^k}\sigma$ in terms of $\sigma(x)$ for $k = 1, \cdots, 3$ where σ is the sigmoid function.

(b) Find the relation between sigmoid function and hyperbolic function.

(a)

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

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

$= \sigma(x)(1-\sigma(x))$ ∎

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

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

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

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

③ $$\sigma'''(x) = \{[\sigma(x)(1-\sigma(x))(1-2\sigma(x))]\}'$$

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

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

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

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

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

(b) Compare $\tanh(x)$ & $\sigma(x)$

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

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

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

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

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

Therefore, $\tanh(x) = 2\sigma(2x) - 1 \quad \blacksquare$

3. There are unanswered questions during the lecture, and there are likely more questions we haven't covered. Take a moment to think about them and write them down here.

Loss function 可以確定收斂到 0，即收斂到我要的目標嗎？

Ans: No!

<explain 無法收斂到0>

就如同 Problem 1, $\nabla_\theta h = (h-y)\,h\,(1-h)\,[1, x_1, x_2]^T$

With $(x_1, x_2, y) = (1,2,3)$; $(b, w_1, w_2) = (4,5,6)$

Then $z = 4+5\cdot 1+6\cdot 2 = 21$, $h = \sigma(21) \approx 1$

$\boxed{L_{min} = \frac{1}{2}(1-y)^2 = \frac{1}{2}(1-3)^2 = 2}$, This means the loss function converge to 2, not to 0 ∎