HW 1 11/6-2015 曹晉嘉

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

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

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.

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.

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.

1. $h(x_1, x_2) = \sigma(b + w_1 x_1 + w_2 x_2)$, $(x_1, x_2, y) = (1, 2, 3)$,

$\theta^0 = (b, w_1, w_2) = (4, 5, 6)$

Let Loss function $L = \frac{1}{2}(y - h(x_1, x_2))^2$

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

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial b} = -(y - h(x_1, x_2))\,\sigma'(b + w_1 x_1 + w_2 x_2)$$

$$= -(y - h(x_1, x_2))\,\sigma(b + w_1 x_1 + w_2 x_2)\left[1 - \sigma(b + w_1 x_1 + w_2 x_2)\right]$$

$$= -(y - h(x_1, x_2))\,h(x_1, x_2)(1 - h(x_1, x_2))$$

$$\frac{\partial L}{\partial w_1} = -(y - h(x_1, x_2))\,\frac{\partial}{\partial x_1}\sigma(b + w_1 x_1 + w_2 x_2) = x_1 \cdot \frac{\partial L}{\partial b}$$

$$\frac{\partial L}{\partial w_2} = x_2 \cdot \frac{\partial L}{\partial b}$$

$h(1, 2) = \sigma(4 + 5 \cdot 1 + 6 \cdot 2) = \sigma(21)$. $y = 3$

$\theta^1 = \theta^0 - \alpha \nabla_\theta L = (\theta_b^1, \theta_{w1}^1, \theta_{w2}^1)$

$$\Rightarrow \begin{cases} \theta_b^1 = 4 - \alpha\{-[3-\sigma(z_1)]\sigma(z_1)\cdot[1-\sigma(z_1)]\} \\ \theta_{w1}^1 = 5 - \alpha\{-[3-\sigma(z_1)]\sigma(z_1)\cdot[1-\sigma(z_1)]\cdot 1\} \\ \theta_{w2}^1 = 6 - \alpha\{-[3-\sigma(z_1)]\sigma(z_1)\cdot[1-\sigma(z_1)]\cdot 2\} \end{cases}$$

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

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

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

$$= \sigma(x) - \sigma^2(x) \quad \#$$

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

$$= 3\sigma^3(x) - 3\sigma^2(x) + \sigma(x) \quad \#$$

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

$$= 9\sigma^2(x)\cdot\sigma'(x) - 9\sigma(x)\cdot\sigma'(x) + \sigma'(x)$$

$$= \sigma(x)[1-\sigma(x)][9\sigma^2(x) - 9\sigma(x) + 1] \quad \#$$

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

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

3. 問題：為什麼又需要少量的神經元(可能十幾個.)就能有那麼大的作用(例如擬合某個函數)？