# Machine Learning
# Assignment 1

CHANG Yung-Hsuan (張永璿)

111652004

eiken.sc11@nycu.edu.tw

September 4, 2025

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

$$h(x_1, x_2) = \sigma(b + \omega_1 x_1 + \omega_2 x_2),$$

where $\sigma$ is the sigmoid function.

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

**Solution**. The algorithm for the stochastic gradient descent method is

$$\theta^{i+1} := \theta^i - \alpha \cdot \left(\nabla_\theta L\left(\theta^i; (x_1, x_2)\right)\right)$$

with

$$L(\theta) = (y - h_\theta(x_1, x_2))^2$$

and $\alpha$ the learning rate.

Although the label $y = 3$ of the data point $(1, 2, 3)$ is out of the range of the sigmoid function, I write it as is anyway as this problem is just to test whether I can do differentiation properly or not. We have

$$\theta^1 = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + 2 \cdot \alpha \cdot \left(3 - h_{(4,5,6)}(1, 2)\right) \cdot h_{(4,5,6)}(1, 2) \cdot \left(1 - h_{(4,5,6)}(1, 2)\right) \cdot \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

$$= \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + 2 \cdot \alpha \cdot (3 - \sigma(4 + 1 \cdot 5 + 2 \cdot 6)) \cdot \sigma(4 + 1 \cdot 5 + 2 \cdot 6) \cdot (1 - \sigma(4 + 1 \cdot 5 + 2 \cdot 6)) \cdot \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

$$= \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + 2 \cdot \alpha \cdot (3 - \sigma(21)) \cdot \sigma(21) \cdot (1 - \sigma(21)) \cdot \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

$$\Longrightarrow \begin{pmatrix} b^1 \\ \omega_1^1 \\ \omega_2^1 \end{pmatrix} = \begin{pmatrix} 4 + 2 \cdot \alpha \cdot (3 - \sigma(21)) \cdot \sigma(21) \cdot (1 - \sigma(21)) \cdot 1 \\ 5 + 2 \cdot \alpha \cdot (3 - \sigma(21)) \cdot \sigma(21) \cdot (1 - \sigma(21)) \cdot 1 \\ 6 + 2 \cdot \alpha \cdot (3 - \sigma(21)) \cdot \sigma(21) \cdot (1 - \sigma(21)) \cdot 2 \end{pmatrix}.$$

2. Find the expression of

$$\frac{\mathrm{d}^k \sigma}{\mathrm{d}x^k}(x)$$

in terms of $\sigma(x)$ for $k = 1, 2, 3$, where $\sigma$ is the sigmoid function.

Find the relation between the sigmoid function and the hyperbolic function.

**Solution**. For $k = 1$, we have

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x}(x) = \frac{\mathrm{d}\sigma}{\mathrm{d}x}\left(\frac{1}{1 + \exp(-x)}\right)$$

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

$$= \frac{1}{1 + \exp(-x)} \cdot \frac{\exp(-x)}{1 + \exp(-x)}$$

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

For $k = 2$, we have

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

$$= \frac{\mathrm{d}}{\mathrm{d}x}\left(\sigma(x) - (\sigma(x))^2\right)$$

$$= \frac{\mathrm{d}}{\mathrm{d}x}(\sigma(x)) - \frac{\mathrm{d}}{\mathrm{d}x}\left((\sigma(x))^2\right)$$

$$
\begin{aligned}
&= \sigma(x) \cdot (1 - \sigma(x)) - 2 \cdot \sigma(x) \cdot \frac{\mathrm{d}}{\mathrm{d}x}(\sigma(x)) \\
&= \sigma(x) \cdot (1 - \sigma(x)) - 2 \cdot \sigma(x) \cdot \sigma(x) \cdot (1 - \sigma(x)) \\
&= \sigma(x) \cdot (1 - \sigma(x)) \cdot (1 - 2 \cdot \sigma(x)).
\end{aligned}
$$

For $k = 3$, we have

$$
\begin{aligned}
\frac{\mathrm{d}^3\sigma}{\mathrm{d}x^3}(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(\sigma(x) \cdot (1 - \sigma(x)) \cdot (1 - 2 \cdot \sigma(x))) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\left(\sigma(x) - 3 \cdot (\sigma(x))^2 + 2 \cdot (\sigma(x))^3\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}(\sigma(x)) - 3 \cdot \frac{\mathrm{d}}{\mathrm{d}x}\left((\sigma(x))^2\right) + 2 \cdot \frac{\mathrm{d}}{\mathrm{d}x}\left((\sigma(x))^3\right) \\
&= \sigma(x) \cdot (1 - \sigma(x)) \\
&\qquad - 3 \cdot 2 \cdot \sigma(x) \cdot \sigma(x) \cdot (1 - \sigma(x)) \\
&\qquad + 2 \cdot 3 \cdot (\sigma(x))^2 \cdot \sigma(x) \cdot (1 - \sigma(x)) \\
&= \sigma(x) \cdot (1 - \sigma(x)) \cdot \left(1 - 6 \cdot \sigma(x) + 6 \cdot (\sigma(x))^2\right).
\end{aligned}
$$

For the relation between $\sigma$ and tanh, we have

$$
\begin{aligned}
\sigma(x) &= \frac{1}{1 + \exp(-x)} \\
&= \frac{\exp\left(\frac{x}{2}\right)}{\exp\left(\frac{x}{2}\right) + \exp\left(-\frac{x}{2}\right)} \\
&= \frac{\exp\left(\frac{x}{2}\right) - \exp\left(-\frac{x}{2}\right)}{\exp\left(\frac{x}{2}\right) + \exp\left(-\frac{x}{2}\right)} + \frac{\exp\left(-\frac{x}{2}\right)}{\exp\left(\frac{x}{2}\right) + \exp\left(-\frac{x}{2}\right)} \\
&= \tanh\left(\frac{x}{2}\right) + \sigma(-x) \\
&= \tanh\left(\frac{x}{2}\right) + \frac{1}{1 + \exp(x)}
\end{aligned}
$$

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

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

and hence

$$2 \cdot \sigma(x) - 1 = \tanh\left(\frac{x}{2}\right).$$