

HW1 2025/09/07

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 .

Solution:

Assume the loss function is $L(\theta) = \frac{1}{n} \sum_{i=1}^n \|y - h(x_1, x_2)\|_2^2 = (y - h(x_1, x_2))^2$.

Implement stochastic gradient descent method.

$$\theta^{1T} = \theta^{0T} - \alpha \nabla_{\theta} L,$$

where α is the learning rate.

$$\theta^{1T} = \begin{pmatrix} b^1 \\ w_1^1 \\ w_2^1 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + 2\alpha(y - h(x_1, x_2))h'(x_1, x_2)$$

$$\theta^{1T} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + 2\alpha(3 - \sigma(4 + 5 \cdot 1 + 6 \cdot 2))\sigma'(x_1, x_2, b) \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

Since the derivative of sigmoid function is $\sigma'(z) = -(1 + e^{-z})^{-2}e^{-z} \cdot (-1) = \frac{e^{-z}}{(1+e^{-z})^2} = \sigma(z)(1 - \sigma(z))$, then

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

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

Solution:

In problem 1, I have showed the derivative of sigmoid function.

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

$$\begin{aligned}
\frac{d^2}{dx^2}\sigma(x) &= \sigma'(x)(1 - \sigma(x)) - \sigma'(x)\sigma(x) \\
&= \sigma'(x) - 2\sigma(x)\sigma'(x) \\
&= (1 - 2\sigma(x))\sigma'(x) \\
&= (1 - 2\sigma(x))\sigma(x)(1 - \sigma(x))
\end{aligned}$$

$$\begin{aligned}
\frac{d^3}{dx^3}\sigma(x) &= \sigma'(x)(1 - \sigma(x))(1 - 2\sigma(x)) + \sigma(x)(-\sigma'(x))(1 - 2\sigma(x)) + \sigma(x)(1 - \sigma(x))(-2\sigma'(x)) \\
&= \sigma(x)(1 - \sigma(x))(6\sigma(x)^2 - 6\sigma(x) + 1)
\end{aligned}$$

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

Solution:

Let the logistic sigmoid be

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

Step 1 — Solve for the exponentials in terms of $\sigma(x)$:

$$e^{-x} = \frac{1}{\sigma(x)} - 1 = \frac{1 - \sigma(x)}{\sigma(x)}, \quad e^x = \frac{1}{e^{-x}} = \frac{\sigma(x)}{1 - \sigma(x)}.$$

Step 2 — Start from the definition of \tanh :

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

Step 3 — Substitute the expressions from Step 1:

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

Step 4 — Simplify the numerator and denominator:

$$\sigma^2 - (1 - \sigma)^2 = (\sigma - (1 - \sigma))(\sigma + (1 - \sigma)) = (2\sigma - 1) \cdot 1,$$

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

Result:

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

3.在課堂上有提到過 optimizer 的策略的問題，教授你有提到有論文發表了新的optimizer，它的想法是先走一大步，再走一小步，然後再走一大步，接著一小步，如此的循環反覆更新參數，最後就會到最佳解，那這個還蠻有趣的，因為目前optimizer 大家比較常用的是Adam，想知道說當把optimizer的更新過程給視覺化後，跟Adam的差別是如何？