Machine Learning Assignment 01

313652008 黃睿帆

September 06, 2025

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

Answer:

Set $z = b + w_1x_1 + w_2x_2$,

$$h = \sigma(z) = \sigma(4 + 5(1) + 6(2)) = \sigma(21)$$

and in MSE loss, we have the formula: (here, $m = 1$ and α is learning rate.)

$$\theta^{n+1} = \theta^n + 2\alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(y^i - h(x_1^i, x_2^i) \right) \nabla_\theta h \right] \tag{1}$$

Since $h = \sigma(z)$ and note that

$$\begin{aligned} \frac{\partial h}{\partial z} = \sigma'(z) = \frac{e^{-z}}{(1+e^{-z})^2} &= \left[\frac{1+e^{-z}}{(1+e^{-z})^2} \right] - \left[\frac{1}{(1+e^{-z})^2} \right] = \left[\frac{1}{1+e^{-z}} \right] - \left[\frac{1}{(1+e^{-z})} \right]^2 \\ &= \sigma(z) - \sigma(z)^2 = h - h^2 = h(1-h) \end{aligned}$$

Then we calculate $\nabla_\theta h = \left(\frac{\partial h}{\partial b}, \frac{\partial h}{\partial x_1}, \frac{\partial h}{\partial x_2} \right)$, where

$$\begin{aligned} \frac{\partial h}{\partial b} &= \frac{\partial h}{\partial z} \frac{\partial z}{\partial b} = h(1-h) \cdot 1 \\ \frac{\partial h}{\partial w_1} &= \frac{\partial h}{\partial z} \frac{\partial z}{\partial w_1} = h(1-h)x_1 = h(1-h) \cdot 1 \\ \frac{\partial h}{\partial w_2} &= \frac{\partial h}{\partial z} \frac{\partial z}{\partial w_2} = h(1-h)x_2 = h(1-h) \cdot 2. \end{aligned}$$

Hence, we get ($n = 1, m = 1$ in (1)):

$$\theta^1 = \theta^0 + 2\alpha(y - h)\nabla_\theta h$$

so $\theta^1 = (b^1, w_1^1, w_2^1)$,$(h = \sigma(21))$,where

$$b^1 = 4 + 2\alpha(3-h)h(1-h)\cdot 1$$

$$w_1^1 = 5 + 2\alpha(3-h)h(1-h)\cdot 1$$

$$w_2^1 = 6 + 2\alpha(3-h)h(1-h)\cdot 2$$

Problem 2

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

Answer:

(a) Like in Problem 1, we already have

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

Then, using $\sigma'(x) = \sigma - \sigma^2$

$$\sigma''(x) = \frac{\partial}{\partial x}(\sigma - \sigma^2) = \sigma'(x) - 2\sigma(x)\sigma'(x) = \sigma'(x)(1-2\sigma(x)) = \sigma(1-\sigma)(1-2\sigma)$$

For $k = 3$,

$$\begin{aligned}\sigma'''(x) &= \frac{\partial}{\partial x}[\sigma'(x)(1-2\sigma(x))] = \sigma''(x)(1-2\sigma(x)) - \sigma'(x)2\sigma'(x)\\ &= \sigma(x)(1-\sigma(x))(1-2\sigma(x))^2 - 2(\sigma(1-\sigma))^2 = \sigma(1-\sigma)[(1-4\sigma+4\sigma^2) - 2\sigma(1-\sigma)]\\ &= \sigma(1-\sigma)[1-6\sigma+6\sigma^2]\end{aligned}$$

Hence we have

$$\begin{cases}\sigma'(x) = \sigma(1-\sigma)\\ \sigma''(x) = \sigma(1-\sigma)(1-2\sigma)\\ \sigma'''(x) = \sigma(1-\sigma)(1-6\sigma+6\sigma^2)\end{cases}$$

(b) Recall that

$$\cosh(x) = \frac{e^x + e^{-x}}{2}; \sinh(x) = \frac{e^x - e^{-x}}{2}; \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

and since

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

Note that

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

hence

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

Problem 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.

Answer:

1. Like someone asked in the class, when using **Gradient Descent Algorithm**, how to deal with the situation that f is NOT differentiable (i.e. ∇f cannot be calculated)

2. In supervised learning, we commonly use mini-batch gradient descent, and choosing $m < N$, is there the best m to choose?