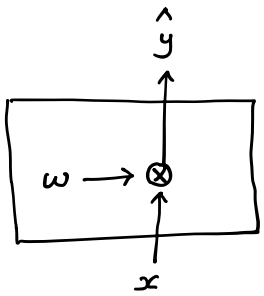
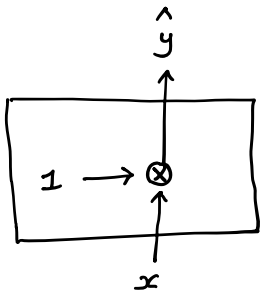
Math Description of Machine Learning

작성자 : 강재권 교수
jaekwon@yonsei.ac.kr
6G + NNV 연구실
Jan. 9 2026

$y = 2x + 1$ ← ideal parameters, unknown parameters



$\hat{y} = wx$ ← 출력 함수, NN architecture



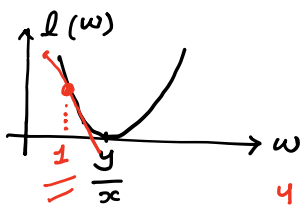
$w \leftarrow 1$ random initialization

$\{x = 1, y = 3\}$ ← training data, observation

$$\ell(w)\Big|_{w=1, x=1, y=3} = (1 \times 1 - 3)^2 = 4$$

loss fcn, 손실 함수

fcn of w → $\ell(w) = (\hat{y} - y)^2 = (wx - y)^2$



why not $w = \frac{y}{x} = \frac{3}{1}$?

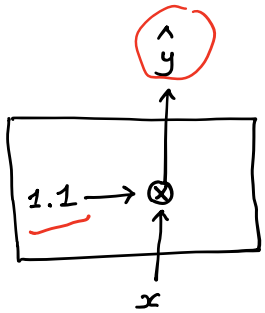
$$\Rightarrow \frac{d\ell}{dw} = 2(wx - y)x$$

$$\frac{d\ell}{dw}\Big|_{x=1, y=3, w=1} = 2(1 - 3) = -4$$ 의미?

learning rate η

$$w \leftarrow 1 - 0.1 \frac{-4}{|-4|} = 1 + 0.1 = 1.1$$

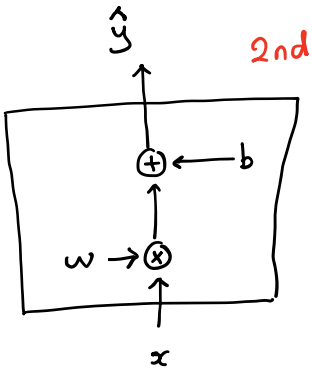
update 반대 방향
초기는 경험상 "1"



$\ell(w) = (1.1 - 3)^2 = (-1.9)^2 < 4$

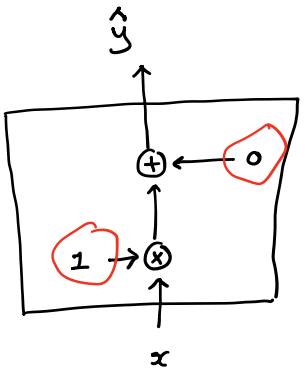
↑ cost reduced

Test $\{x = 2, y = 5\}$ $(\hat{y} - y)^2 = (2.2 - 5)^2 = (2.8)^2$



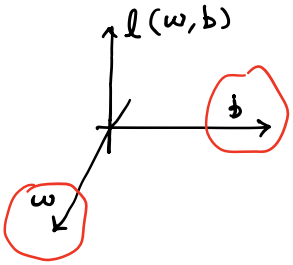
2nd NN architecture

$\hat{y} = wx + b$



$w \leftarrow 1$, $b \leftarrow 0$ } random initialization

$\{x = 1, y = 3\}$ ← training data



fcn of w,b → $l(w,b) = (\hat{y} - y)^2 = (wx + b - y)^2$

$= (1 \times 1 + 0 - 3)^2 = 4$

$$\nabla l = \begin{bmatrix} \frac{\partial l}{\partial w} \\ \frac{\partial l}{\partial b} \end{bmatrix} = \begin{bmatrix} 2(wx + b - y)x \\ 2(wx + b - y) \end{bmatrix}$$

$$\left.\frac{\partial l}{\partial w}\right|_{x=1,\, y=3,\, w=1,\, b=0} = 2(1-3) = -4$$

train data, 현위치

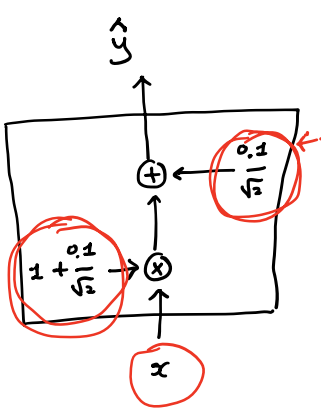
$$\left.\frac{\partial l}{\partial b}\right|_{x=1,\, y=3,\, w=1,\, b=0} = 2(1-3) = -4$$

$$\nabla l = \begin{bmatrix} -4 \\ -4 \end{bmatrix} \longrightarrow \frac{\nabla l}{\|\nabla l\|} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

← update 반대 방향, 초기는 경사 "1"

learning rate η

$$\begin{bmatrix} w \\ b \end{bmatrix} \leftarrow \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 + \frac{0.1}{\sqrt{2}} \\ \frac{0.1}{\sqrt{2}} \end{bmatrix}$$



fcn of x

$$l(w,b) = \left(\left(1 + \frac{0.1}{\sqrt{2}}\right) + \frac{0.1}{\sqrt{2}} - 3 \right)^2 = \left(1 + \frac{\sqrt{2}}{10} - 3\right)^2 \approx (1.14 - 3) < 4$$

↑ cost reduced

Test $\{x = 2, y = 5\}$ $(\hat{y} - y)^2 = \left(\left(1 + \frac{0.1}{\sqrt{2}}\right)2 + \frac{0.1}{\sqrt{2}} - 5 \right)^2 = \left(\frac{0.3}{\sqrt{2}} - 3 \right)^2 \approx (2.79)^2$