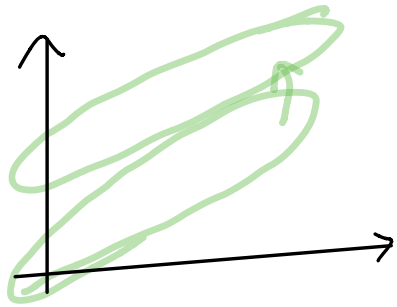
Bias term 추가.



$y = ax + b$.

(ex) 공부시간. 정수 에서 ⊕ 추가정수 1개. ⇒ 전체에 $+b$.

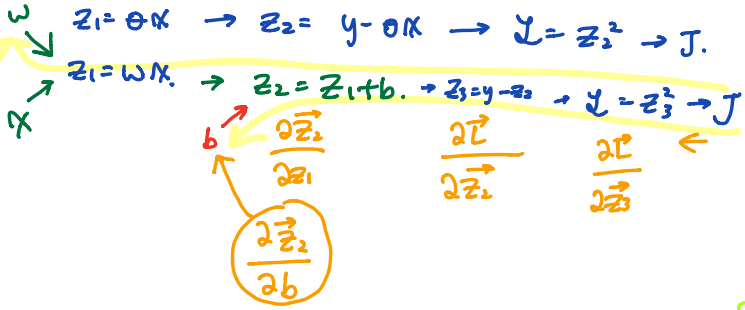
$y = ax + 1$. $y = \underbrace{a}_{\text{weight}}x + \underbrace{b}_{\text{bias}}$

data → modeling → parameter. → learning.

pred ⇒ $y^{(i)} = ax^{(i)} + b$.

$\mathcal{L}^{(i)} = (y^{(i)} - \hat{y}^{(i)})^2$

$J = \frac{1}{N}\sum_{i}^{N}\mathcal{L}^{(i)}$ 이부분의 편차값.



$w = w - \alpha \cdot \frac{\partial J}{\partial w}$ → $\frac{\partial J}{\partial \mathcal{L}} \cdot \frac{\partial \mathcal{L}}{\partial z_3} \cdot \frac{\partial z_3}{\partial z_2} \cdot \frac{\partial z_2}{\partial z_1} \cdot \frac{\partial z_1}{\partial w}$

$b = b - \alpha \cdot \frac{\partial J}{\partial b}$ → $= \cdot \frac{\partial z_2}{\partial b}$

※ w, b는 동시에 update? → NO.

back prop으로 update 후 pred할 시, bias가 learning이 안되면 실제 pred도 slow해짐.

즉. w와 b가 학습될 시, 증감하다가 멈추는 순간,
대충 pred로 망함.

pred ⇒ $\hat{y} = ax+b$

$\mathcal{L} = (y-\hat{y})^2$

$= (y-(ax+b))^2$ → 편미분. (stochastic gradd. descent 방법)

cost. $J = \frac{1}{N}\sum (y^{(i)} - w\cdot x^{(i)} - b)^2$

$\frac{\partial \mathcal{L}}{\partial w} = 2(y - wx^{(i)} - b)\cdot(-x^{(i)})$

$\frac{\partial \mathcal{L}}{\partial b} = 2(y - wx^{(i)} - b)\cdot(-1)$

$w = w - \alpha(-2x^{(i)})(y^{(i)} - w\cdot x^{(i)} - b)$ ($\hat{y}$)

$b = b - \alpha(-2)(y^{(i)} - w\cdot x^{(i)} - b)$ ($\hat{y}$)

x값이 클수록 증감차이가 큼.