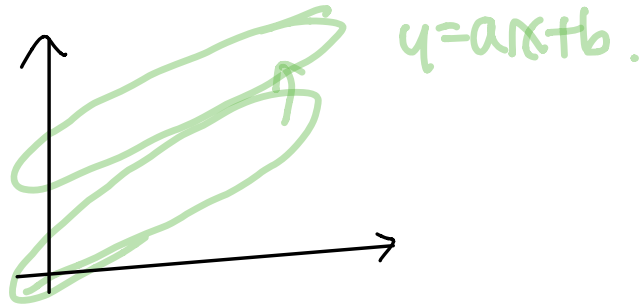


Bias term 추가.



(ex) 공복시간: 점수 α 에 \oplus 추가점수 1점 \Rightarrow 전체점 + b.

$$y = \alpha x + b$$

weight bias.

data \rightarrow modeling \rightarrow parameter \rightarrow learning.

pred $\Rightarrow y^{(w)} = \alpha x^{(i)} + b.$

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

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

이항변의 평균.

$z_1 = \alpha x \rightarrow z_2 = y - \alpha x \rightarrow \mathcal{L} = z_2^2 \rightarrow J.$
 $z_1 = \omega x \rightarrow z_2 = z_1 + b. \rightarrow z_3 = y - z_2 \rightarrow \mathcal{L} = z_3^2 \rightarrow J$

$\frac{\partial \mathcal{L}}{\partial z_2}$
 $\frac{\partial \mathcal{L}}{\partial z_1}$
 $\frac{\partial \mathcal{L}}{\partial b}$

* ω, b 는 학습률 update? \rightarrow NO.

$$\omega = \omega - \alpha \cdot \frac{\partial J}{\partial \omega}$$

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

backpropoz
update &
pred 할시, bias가
learning 이 느리면
실제 pred 느 slow해짐.

즉, w or b 가 학습될 시, 중간하나가 발생하는 순간,
다음 pred는 망함.

$$\text{pred} \Rightarrow \hat{y} = ax + b$$

$$\mathcal{L} = (y - \hat{y})^2 \quad \text{← 평미분. (stochastic grad. descent 방법)}$$

$$= (y - (ax + b))^2$$

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

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

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

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

x와 y가
 클 경우
 학습 차이가 큼.

\hat{y}