

AI and Deep Learning

선형회귀

오류 그래프와 기울기, 그리고 경사하강(2/2)

제주대학교

변영철

<http://github.com/yungbyun/mllecture>

공부할 내용

- 회귀의 의미
- 뉴런의 출력과 절대값 오류
- 기울기와 경사하강
- 절대값 오류와 평균 제곱오류
- 기울기를 구하는 방법
- 기울기가 갖는 의미



어떻게 w 를 ‘**자동으로**’
조절할 것인가?

(학습, Learning)

- 오류 E 가 줄어들도록 w 를 조절
→ 그러면 뉴런 대답이 정답에 가까워 짐.

www.desmos.com

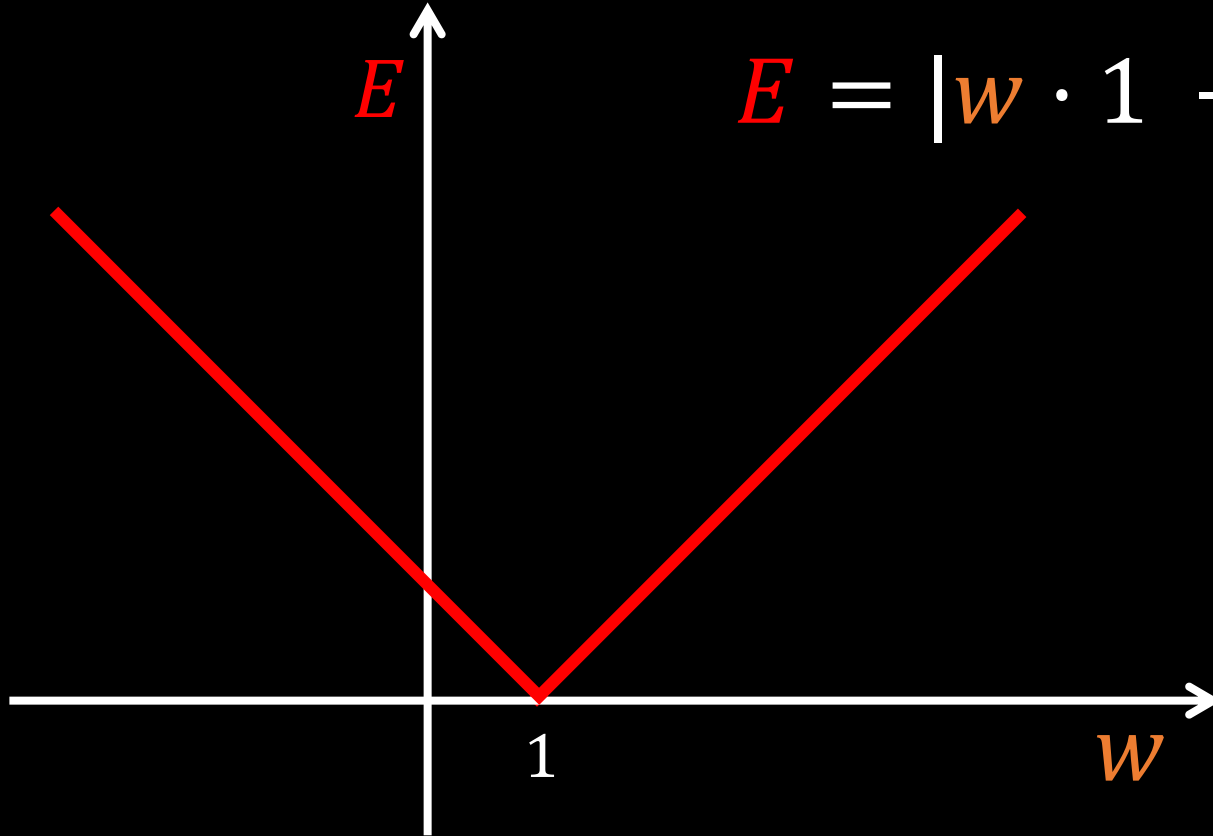
1. 데이터 $(1, 1)$ 표시
2. $E = |w \cdot 1 - 1|$
3. (w, E)



절대값 오류 함수

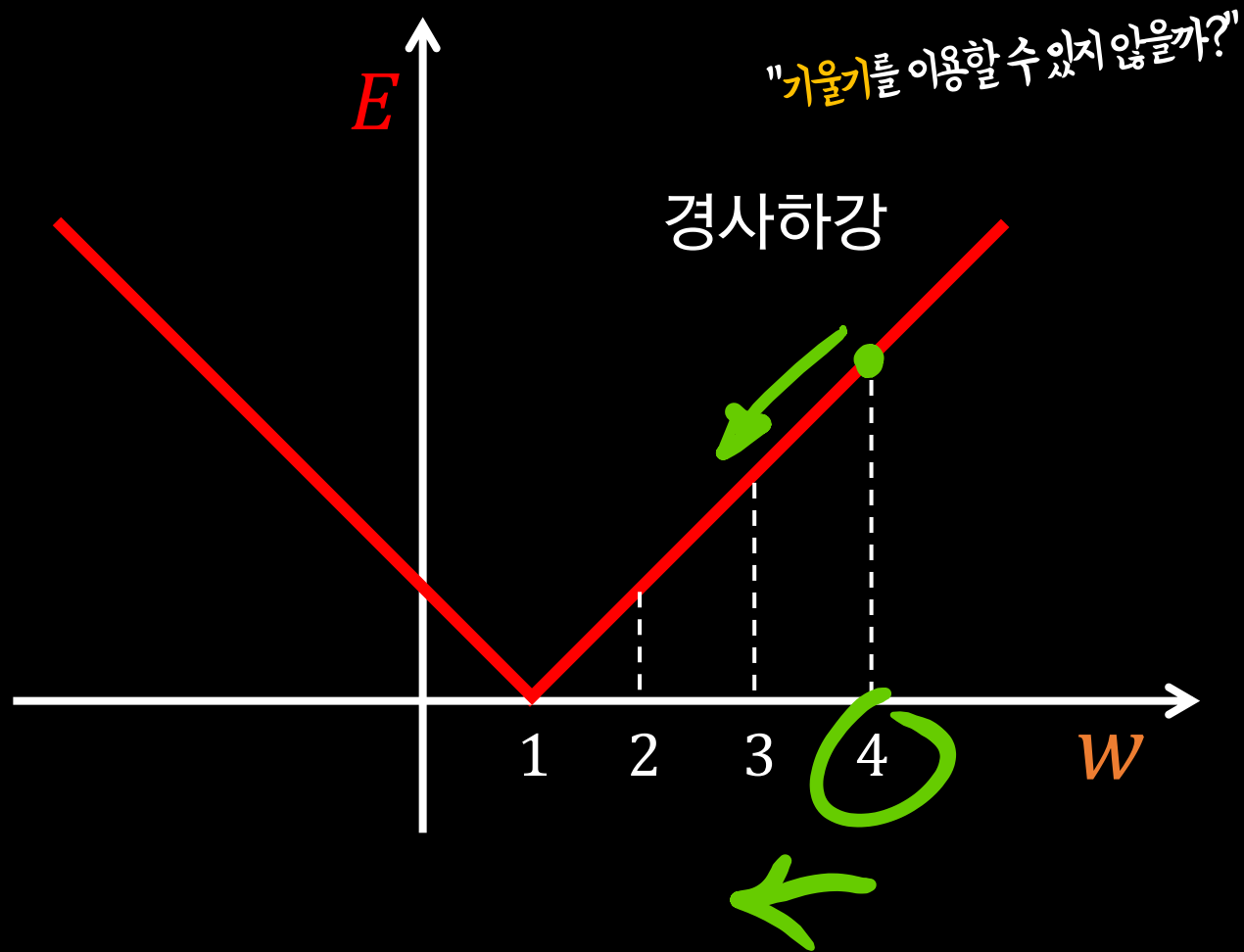


$$E = |w \cdot 1 - 1|$$



오류가 줄어들도록

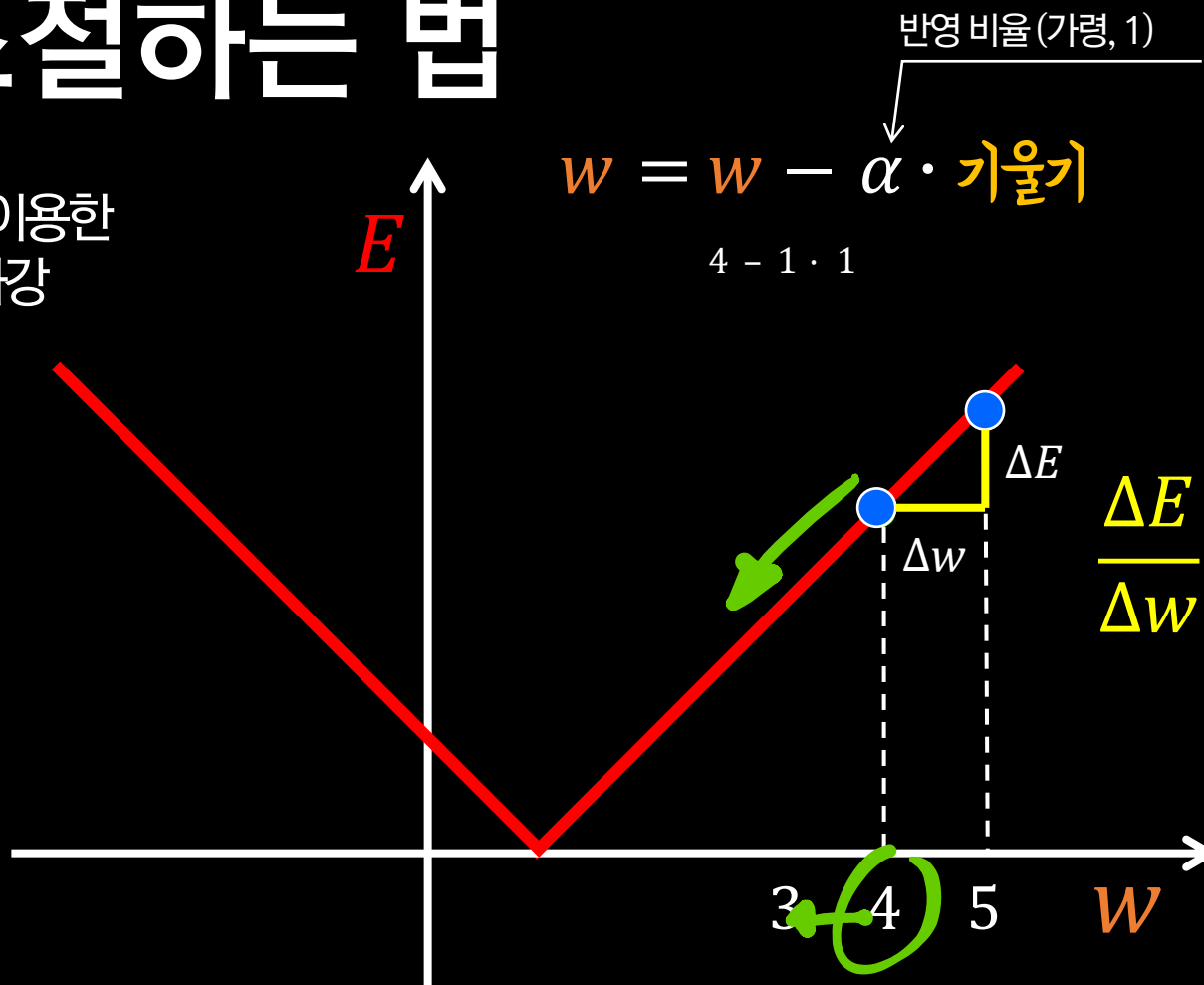
w 조절하는 법



오류가 줄어들도록

w 조절하는 법

기울기를 이용한
경사하강

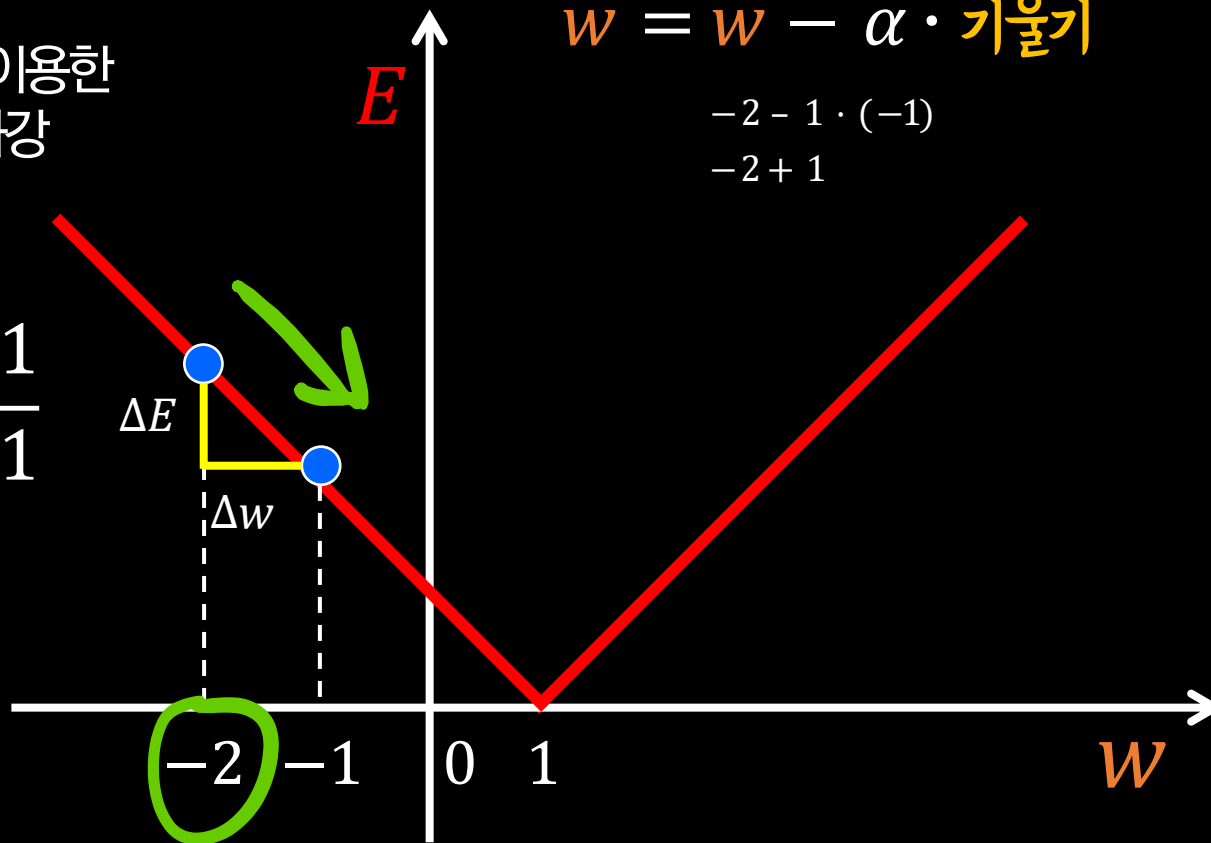


오류가 줄어들도록

w 조절하는 법

기울기를 이용한
경사하강

$$\frac{\Delta E}{\Delta w} = \frac{-1}{+1}$$



어디에서의 기울기?

w 지점에서의 기울기


$$w = w - \alpha \cdot \text{기울기}$$

$$\boxed{4}$$

$$1$$

$$1$$

$$W - \alpha \cdot \text{가울기} \rightarrow W$$

$$4 - 1 \cdot 1 \rightarrow$$

$$3$$

$$\text{error } E = 2$$

$$3 - 1 \cdot 1 \rightarrow$$

$$2$$

$$\text{error } E = 1$$

$$2 - 1 \cdot 1 \rightarrow$$

$$1$$

$$\text{error } E = 0$$

$$\boxed{-2} \quad 1 \quad -1$$

$$W - \alpha \cdot \text{가운데} \rightarrow W$$

$$\begin{array}{rcl} -2 - 1 \cdot (-1) & \rightarrow & \boxed{-1} \\ -1 - 1 \cdot (-1) & \rightarrow & \boxed{0} \\ 0 - 1 \cdot (-1) & \rightarrow & \boxed{1} \end{array} \quad \begin{array}{l} \text{error } E = 2 \\ \text{error } E = 1 \\ \text{error } E = 0 \end{array}$$

$$\boxed{-2} \quad \textcircled{2} \quad -1$$

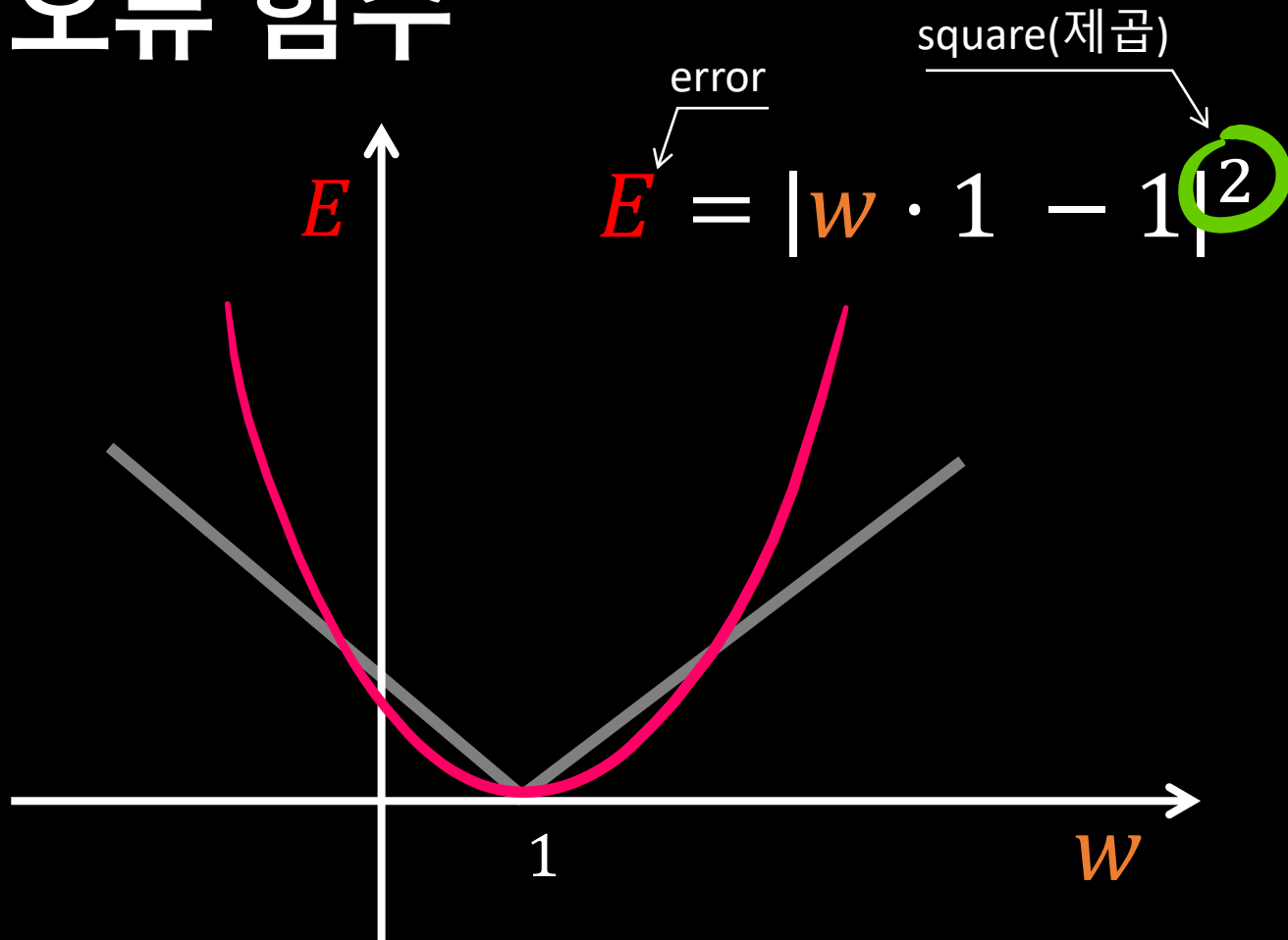
$$W - \alpha \cdot \text{가울기} \rightarrow W$$

	$-2 - 2 \cdot (-1) \rightarrow$	0	error E = 1
!	$0 - 2 \cdot (-1) \rightarrow$	2	error E = 1
	$2 - 2 \cdot (1) \rightarrow$	0	error E = 1
	$0 - 2 \cdot (-1) \rightarrow$	2	error E = 1

절대값 오류, 어떤 문제?

- 왼쪽, 오른쪽, 각각은 항상 같은 기울기, 따라서 항상 같은 경사하강 속도
- w 값, 조절 실패할 수도
- 기울기 값만 보고 w 값을 짐작할 수 없음.
- w 가 1일 때는 기울기 구할 수 없음.

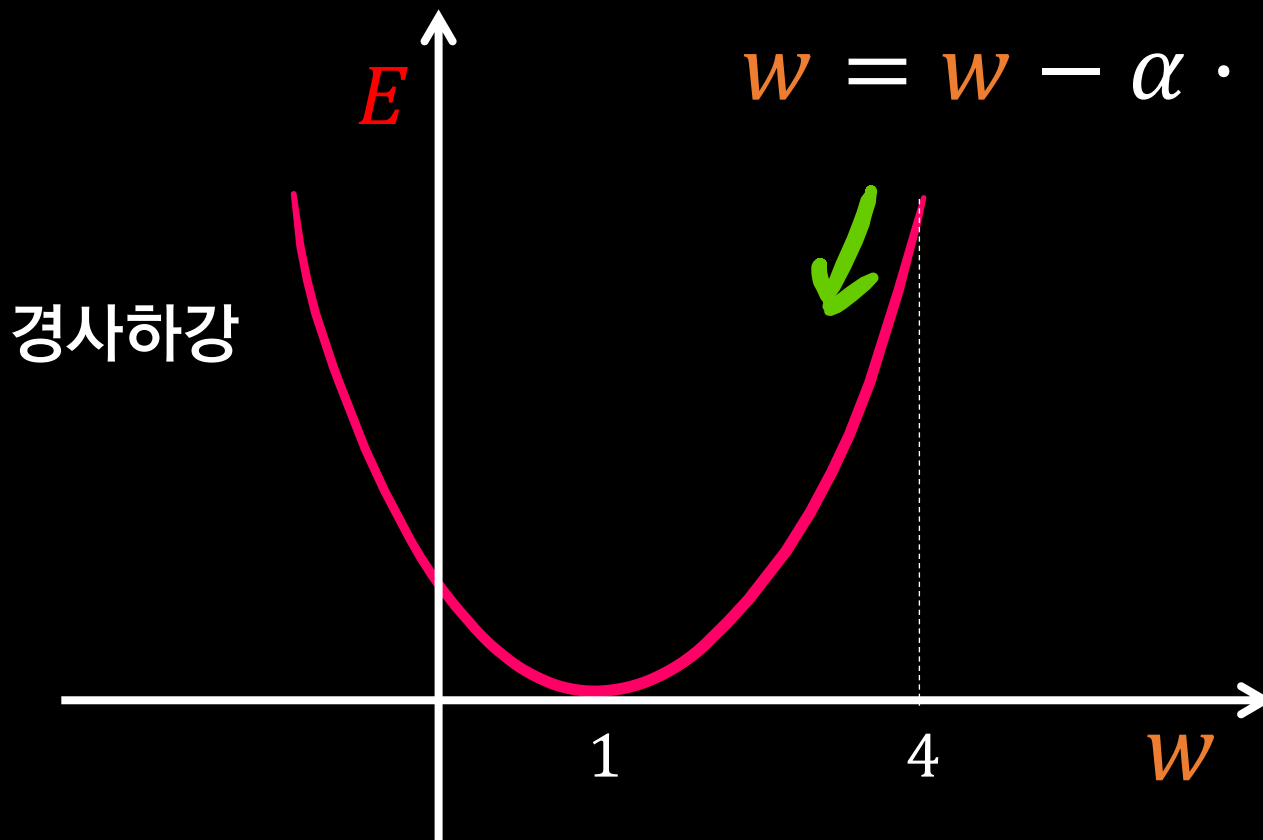
제곱 오류 함수



w 조절하는 법

$$\frac{\Delta E}{\Delta w}$$

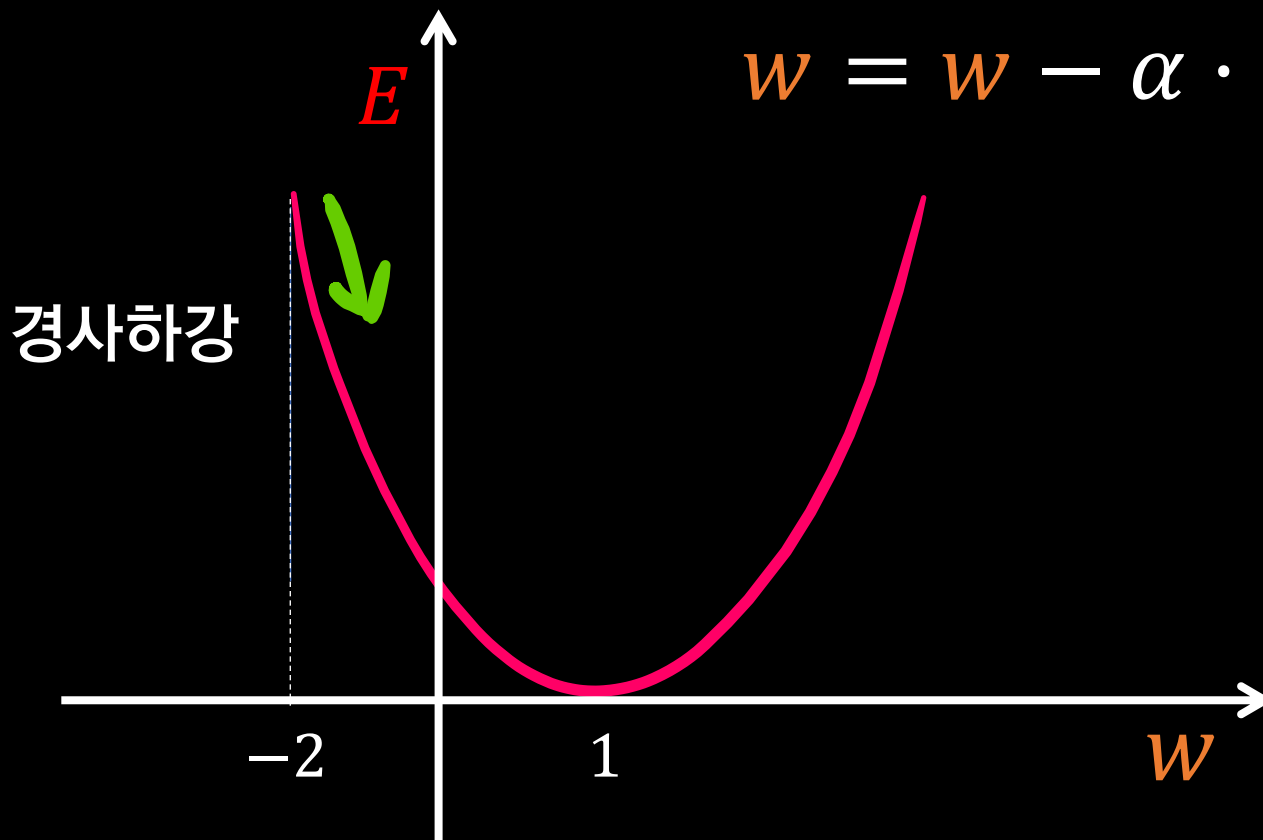
$$w = w - \alpha \cdot \text{기울기}$$



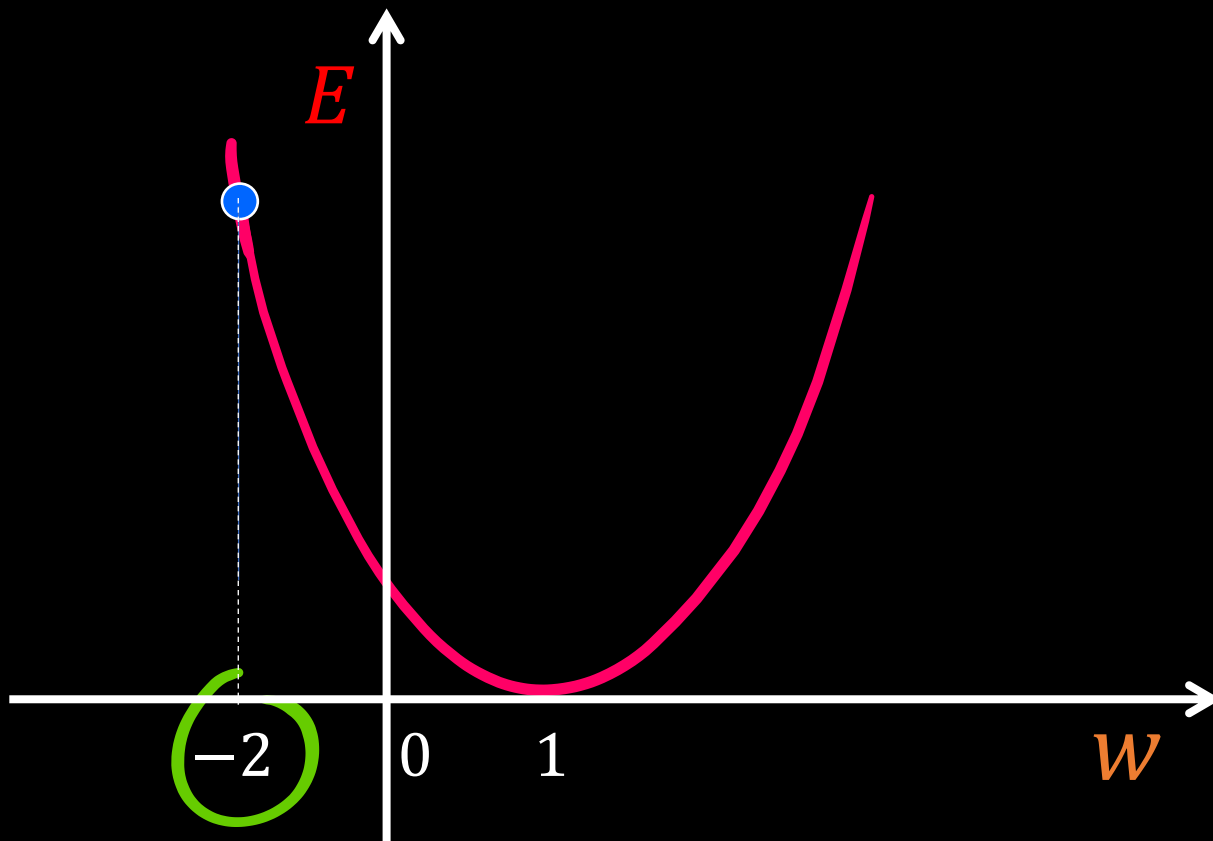
w 조절하는 법

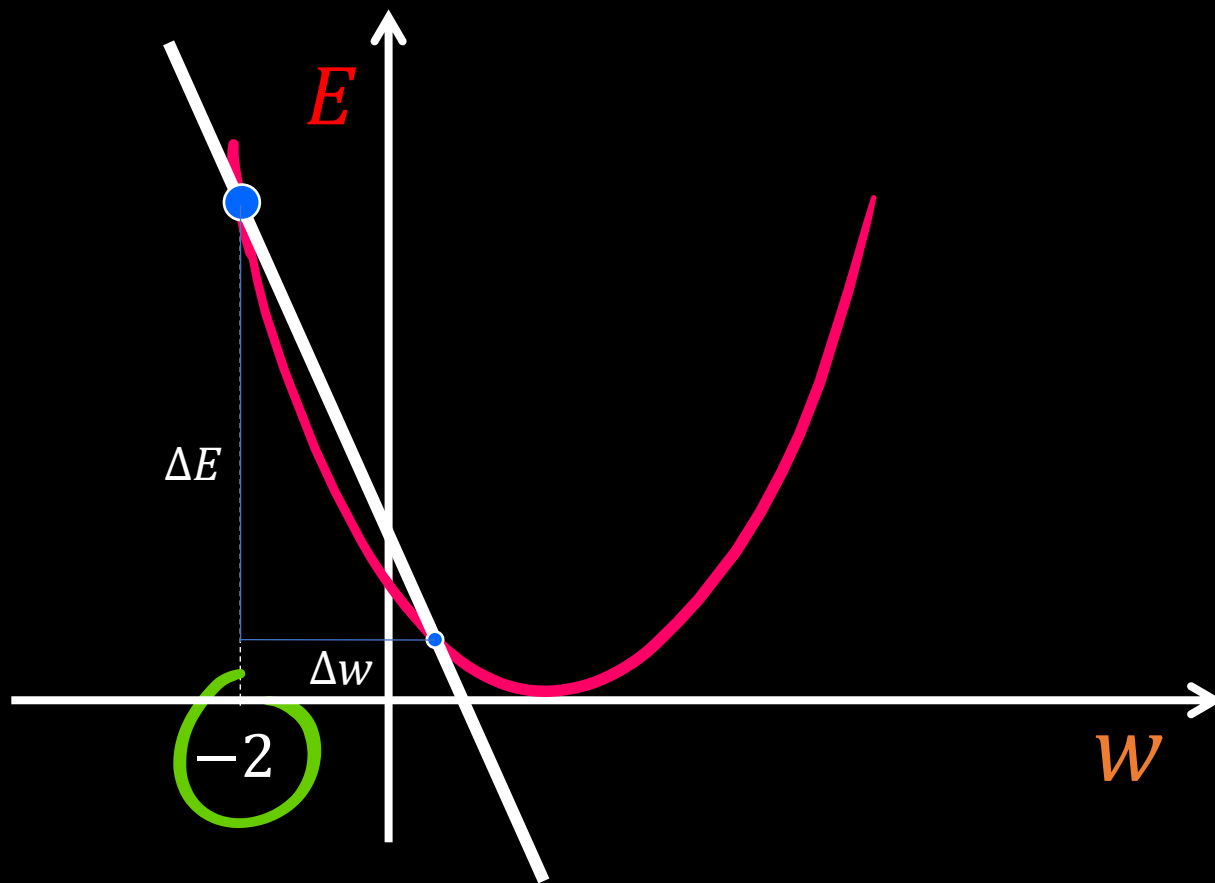
$$\frac{\Delta E}{\Delta w}$$

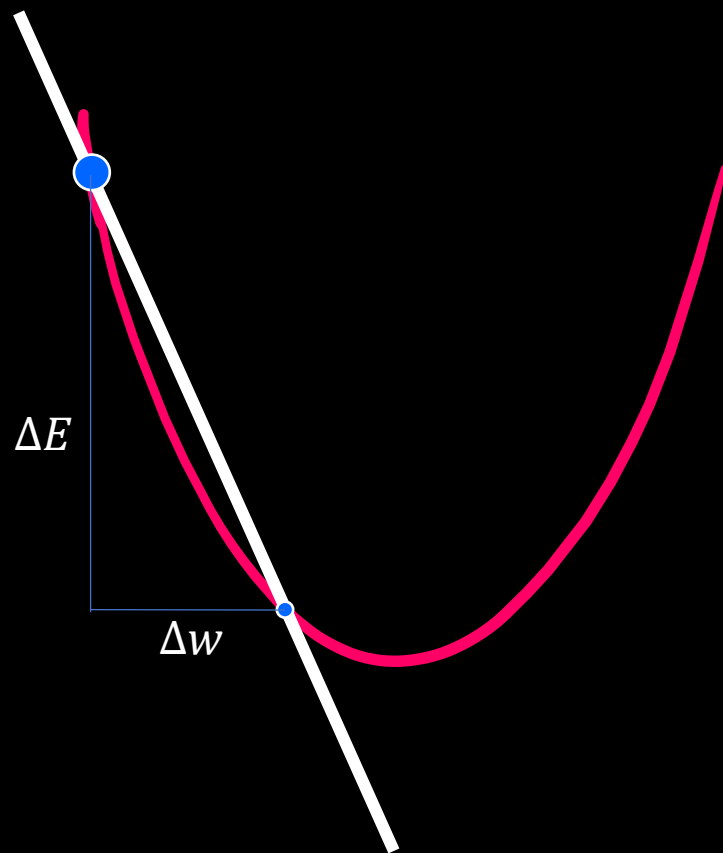
$$w = w - \alpha \cdot \text{기울기}$$

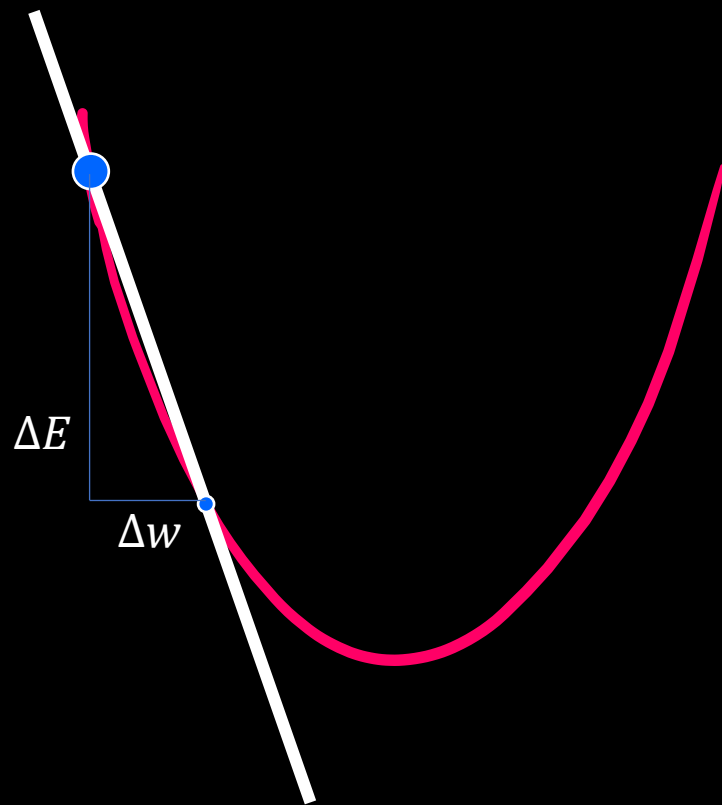


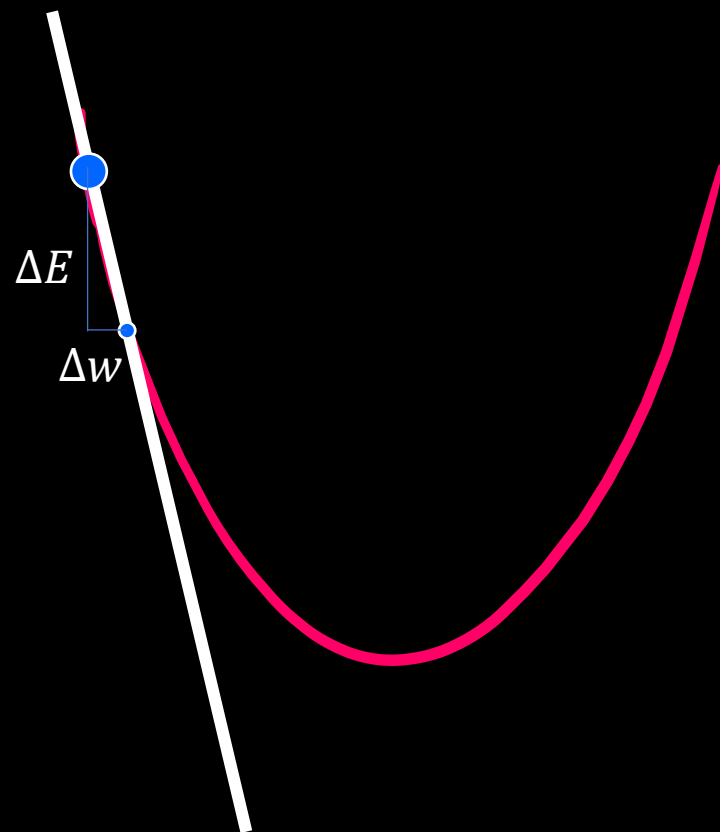
어떻게 기울기를 구할까



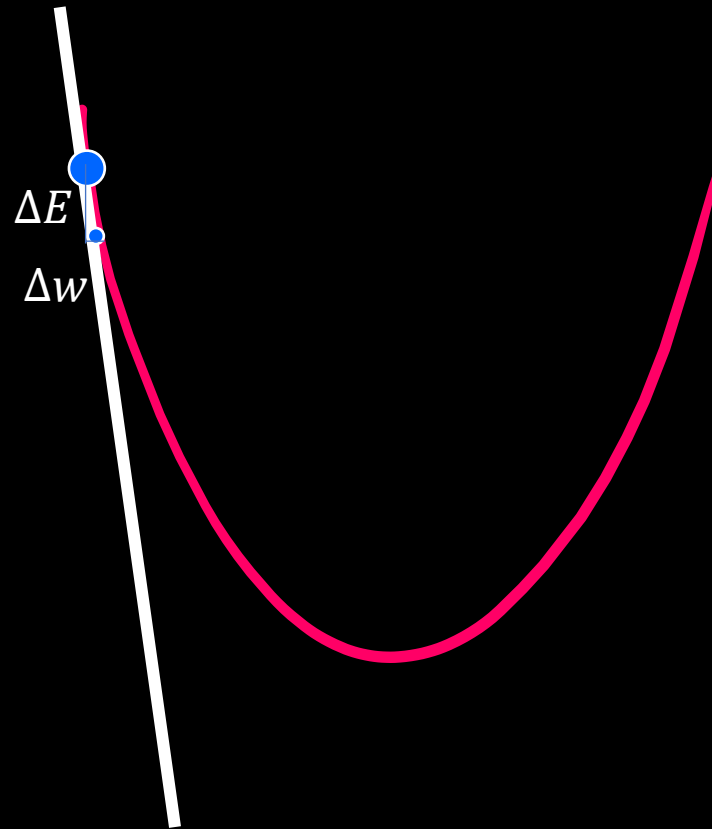


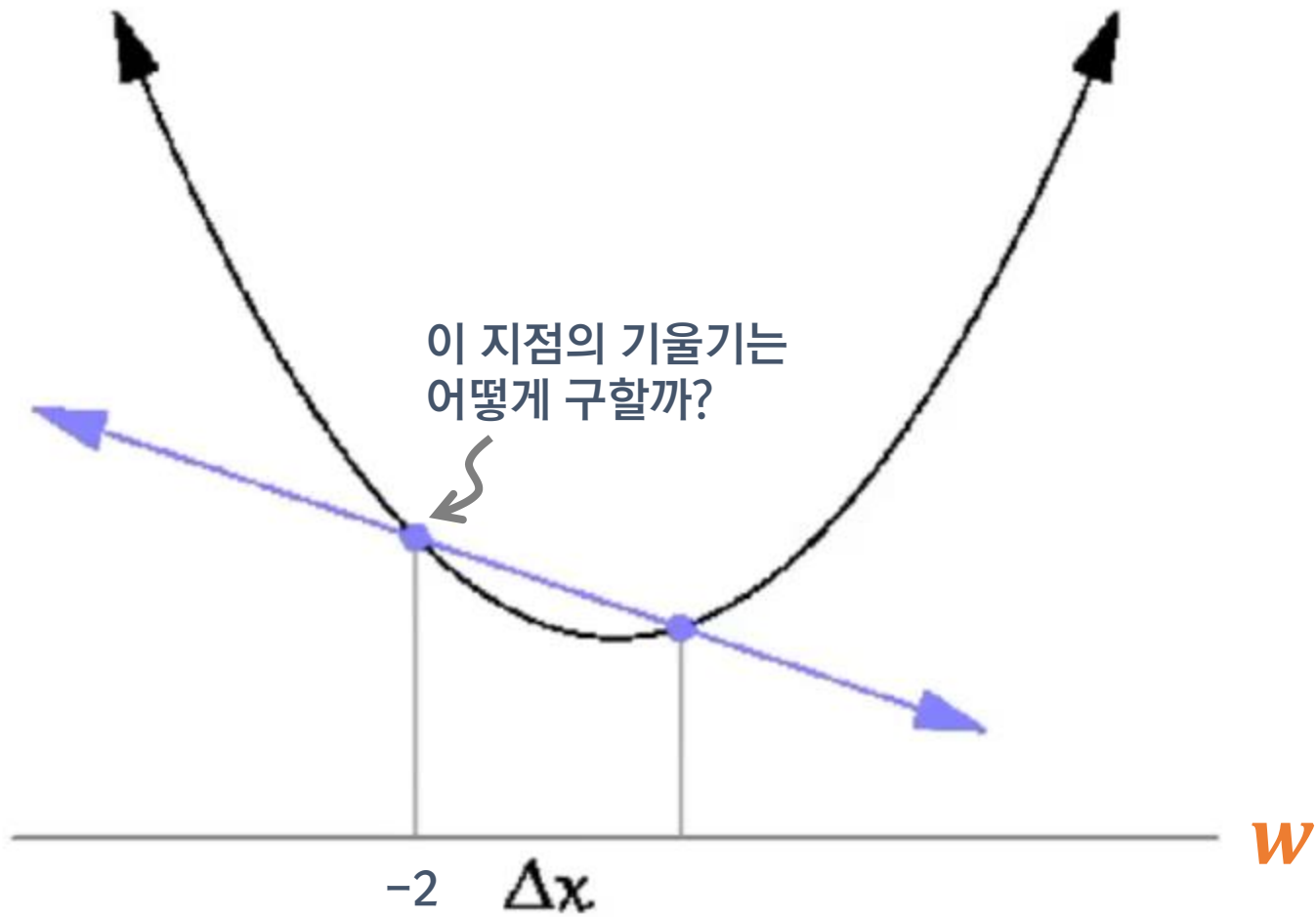




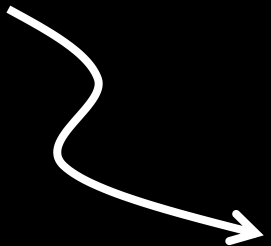


접선·Tangent line





접선으로 정확한
기울기를 구하기 위함.


$$\lim_{\Delta w \rightarrow 0} \frac{\Delta E}{\Delta w}$$

$$= \frac{\partial E}{\partial w}$$

w 를 아주 조금만 늘렸을 때 (∂w)
 E 는 얼마나 늘어날까? (∂E)
두 값으로 구한 값, 기울기

아래 표현은 무슨 의미인지?

$$\lim_{\Delta w \rightarrow 0} \frac{\Delta E}{\Delta w}$$

$$= \frac{\partial E}{\partial w}$$

아주 잘게 자름, 미분(微分)

Numerical
differentiation

- ① 길이를 0에 가깝게 잘게
자름(미분)
- ② 자른 끝을 연결하는 선
→ 접선

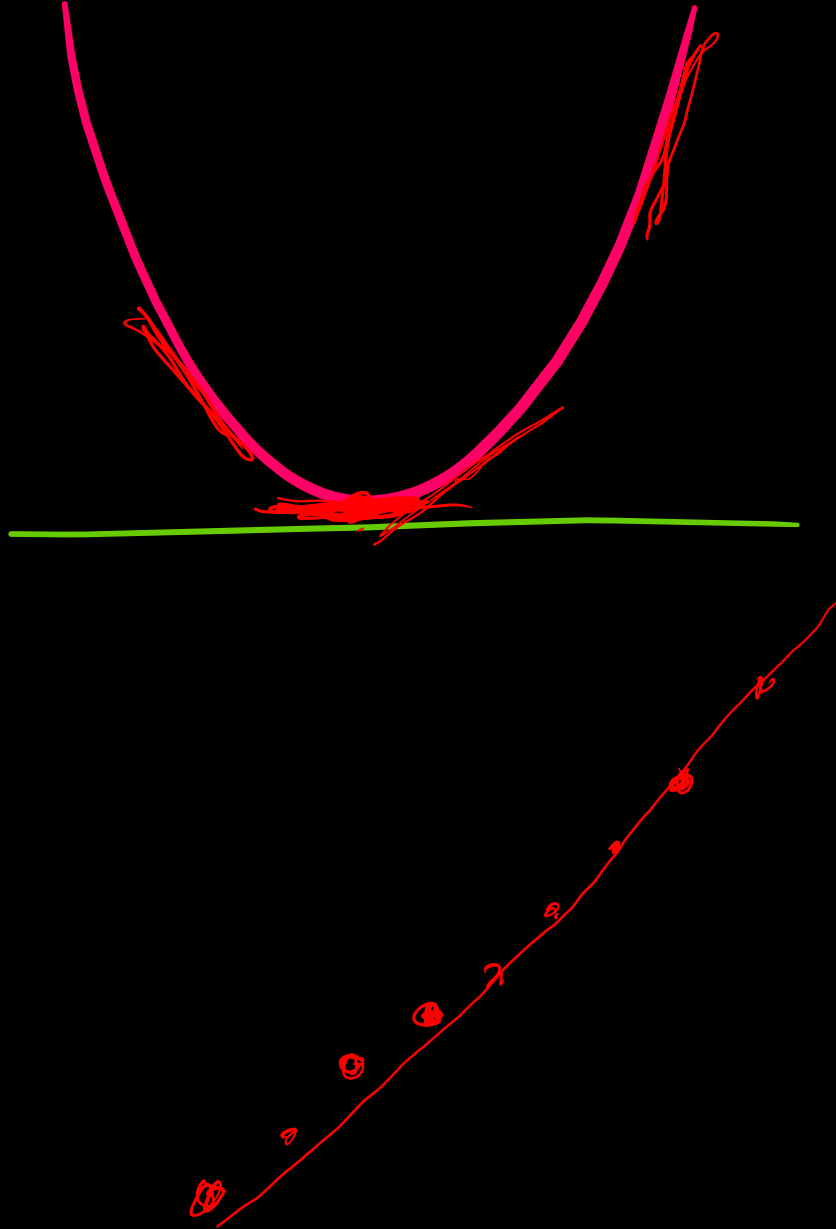


$$\lim_{\Delta w \rightarrow 0} \frac{\Delta E}{\Delta w}$$

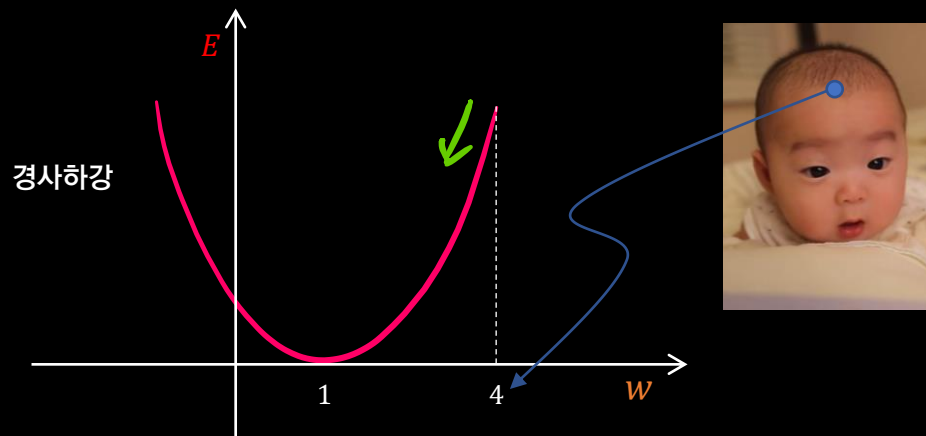
$$= \frac{\partial E}{\partial w}$$

= 결과적으로
선이 아주 작게
나뉘(미분)

아래 그래프의 모든 곳에서의 기울기를 구하시오(미분).



아래 그래프에 대해 기울기를 구하시오(미분).



$$w = w - \alpha \cdot \text{기울기}$$

$$\frac{\partial E}{\partial w}$$

w 를 아주 조금만 늘렸을 때 (∂w)
 E 는 얼마나 늘어날까? (∂E)
 두 값으로 구한 값, 기울기

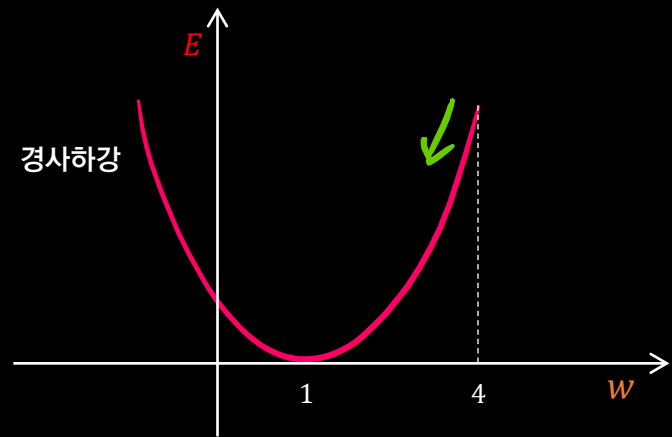
경사하강 수식을 작성하시오.

정답은?

경사하강 수식을 작성하시오.

$$w = w - \alpha \cdot \text{기울기}$$

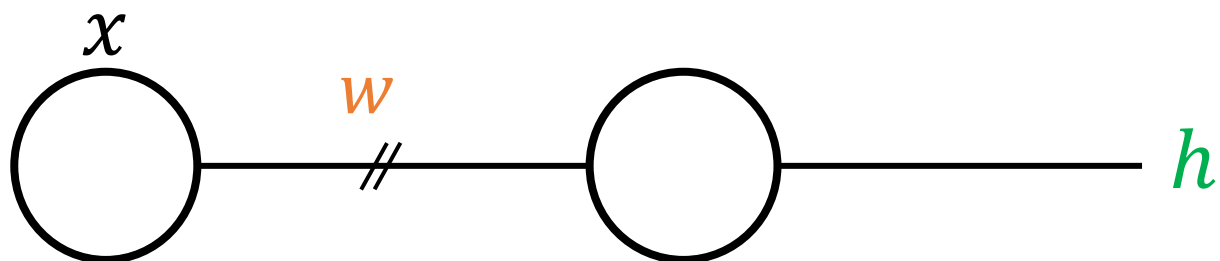
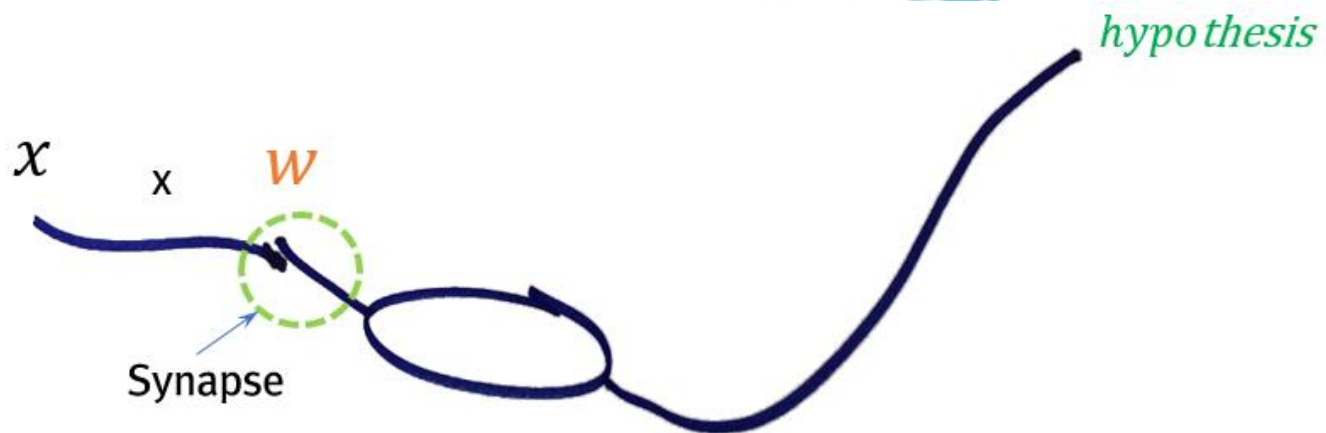
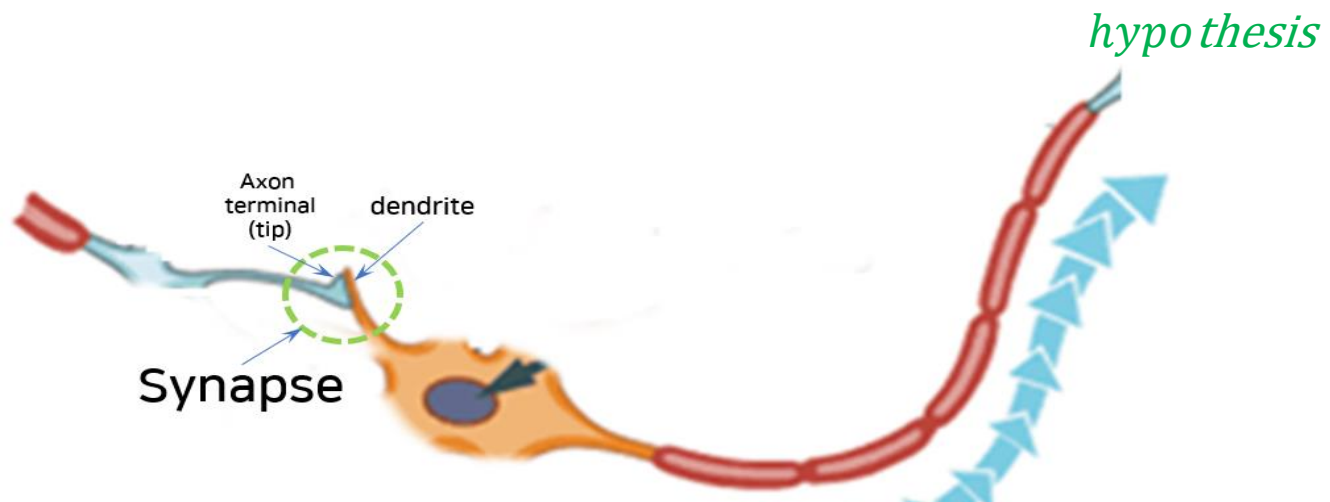
제곱 에러 장점



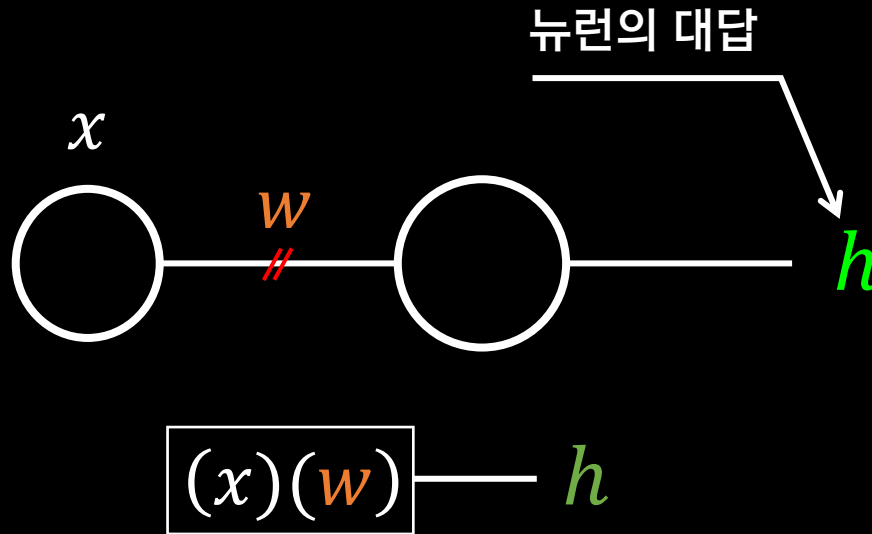
- 양쪽으로 멀리 떨어질 수록 **급경사**
- 가운데로 갈 수록 **완만**
- 따라서, 기울기 값의 크기에 따라 w 가 어디에 있는지 알 수 있음.
- 모든 곳에서 기울기 계산 가능(미분가능)

절대값 오류에서는?

- 왼쪽, 오른쪽, 각각은 항상 같은 기울기
- 따라서 항상 같은 경사하강 속도
- w 값, 조절 실패할 수도
- 기울기 값만 보고 w 값을 짐작할 수 **없음**.
- w 가 1일 때는 기울기 구할 수 **없음**.



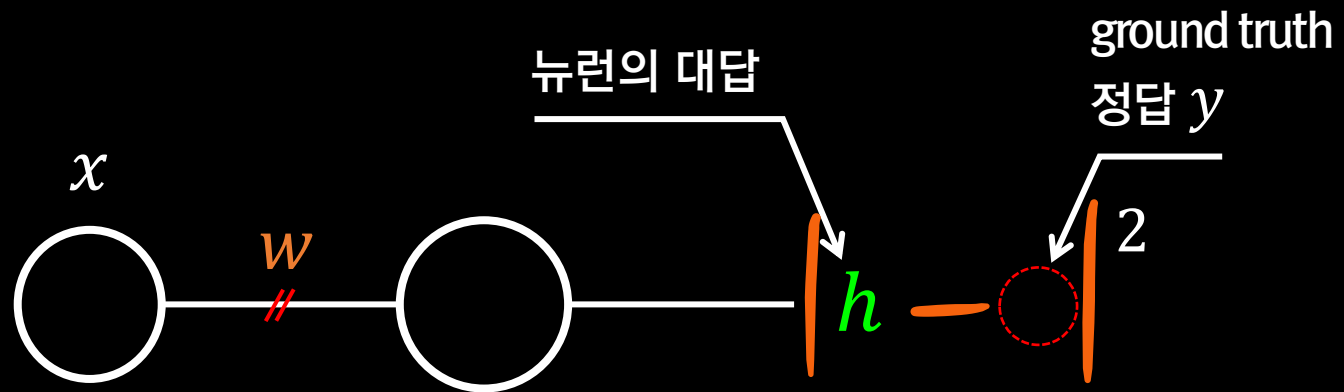
다음과 같은 뉴런(신경세포)이 있다고 하자.
이 뉴런이 표현하는 회귀(대답)를 그려보자.



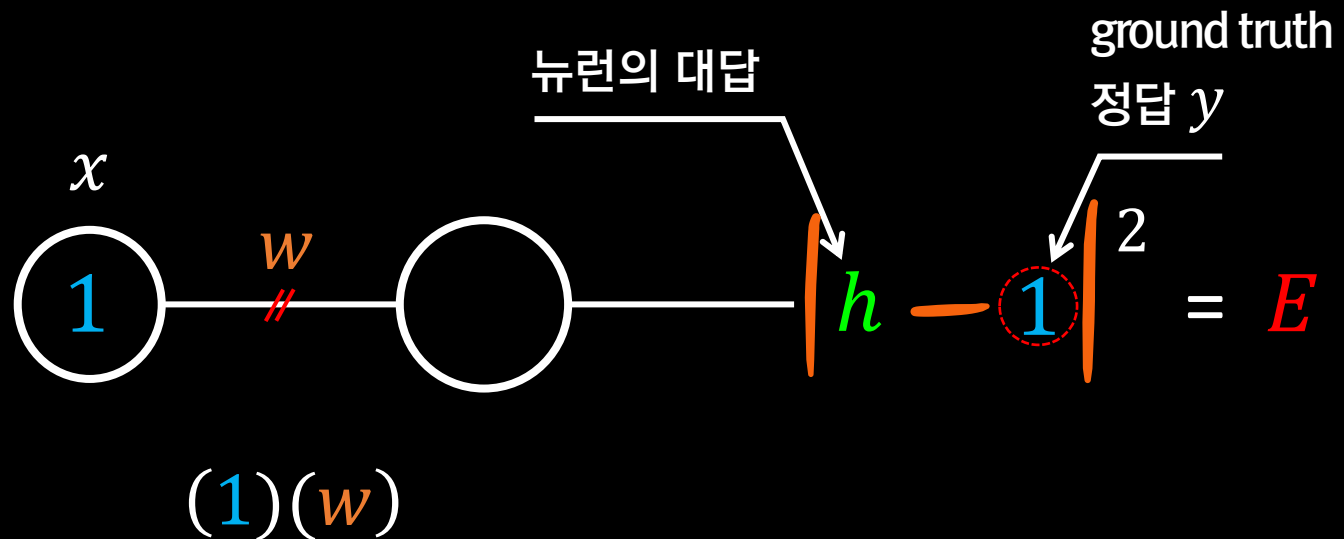
데이터 (1, 1) 표시

$$h = w \cdot x$$

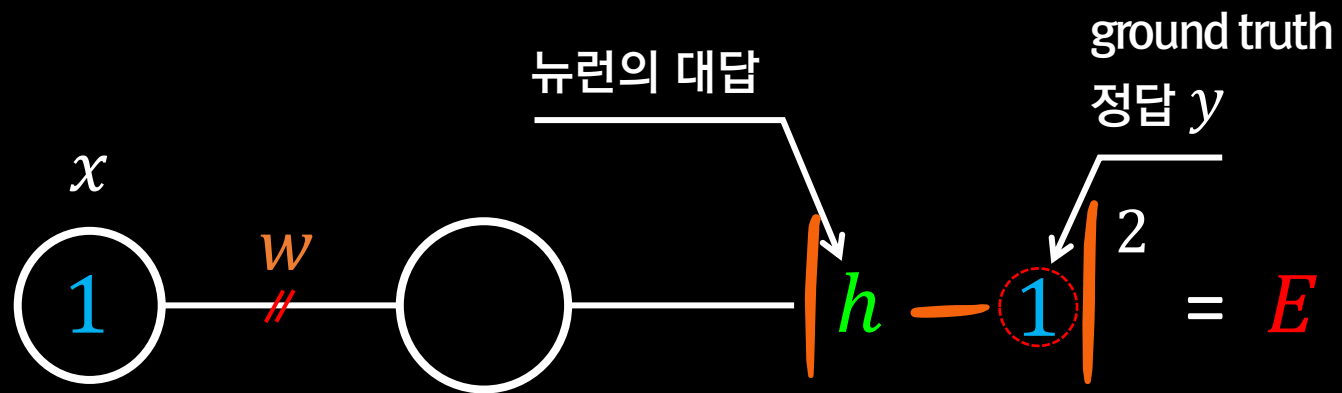
데이터가 (1, 1)일 때 w 에 따른 제곱 오류 함수 E 는?



데이터가 (1, 1)일 때 w 에 따른 제곱 오류 함수 E 는?



데이터가 (1, 1)일 때 w 에 따른 제곱 오류 함수 E 는?



$$(1)(w)$$

$$E = (w \cdot 1 - 1)^2$$

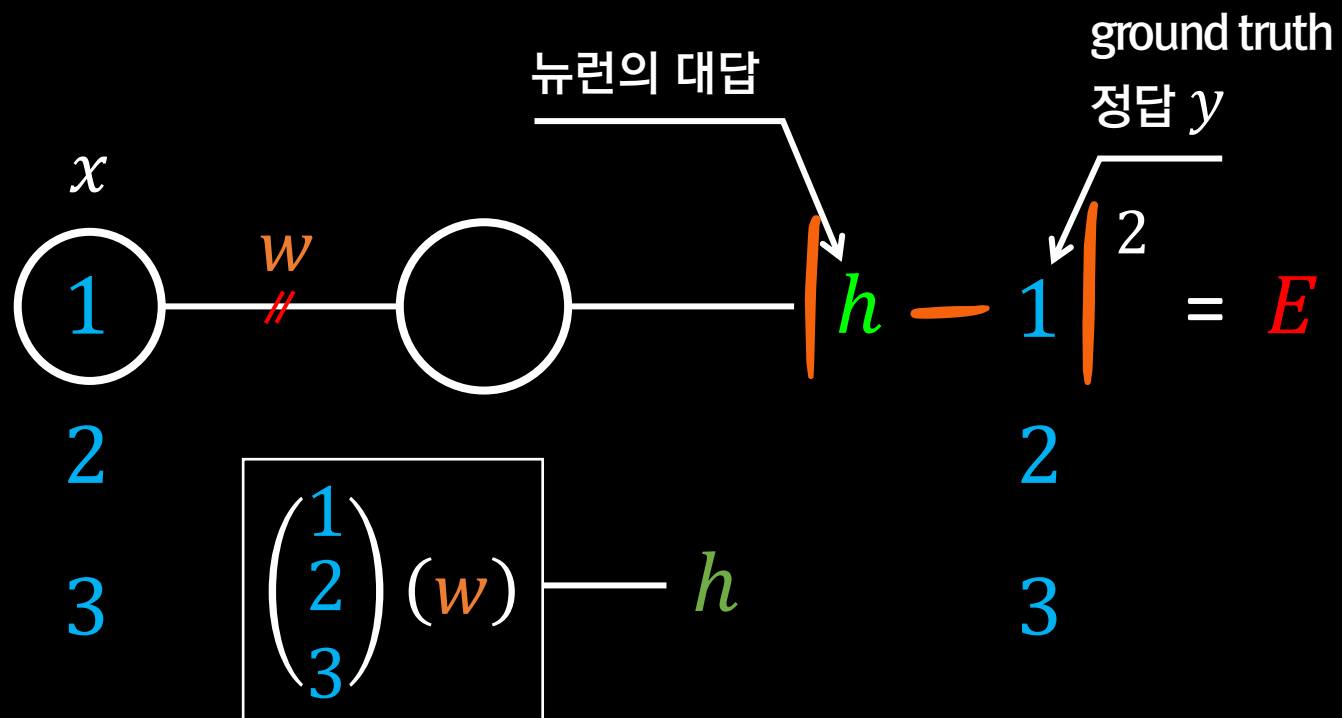


$$E = (w \cdot 1 - 1)^2$$

(w, E)

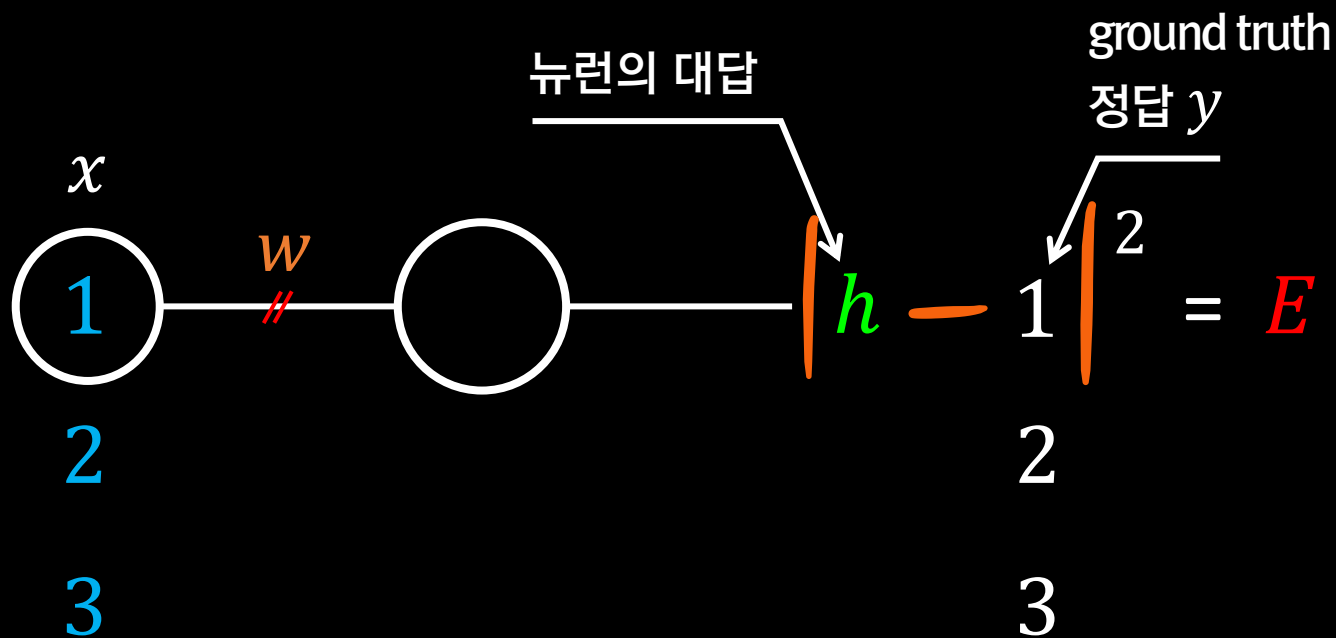
데이터가 (1, 1), (2, 2), (3, 3)이면?

x_i	y_i
1	1
2	2
3	3



데이터가 (1, 1), (2, 2), (3, 3)이면?

x_i	y_i
1	1
2	2
3	3



$$E = (w \cdot 1 - 1)^2 + (w \cdot 2 - 2)^2 + (w \cdot 3 - 3)^2$$

(입력,정답) 데이터가
(1, 1), (2, 2), (3, 3)이면?

x_i	y_i
1	1
2	2
3	3

$$E = \frac{1}{3} ((w \cdot 1 - 1)^2 + (w \cdot 2 - 2)^2 + (w \cdot 3 - 3)^2)$$

Mean Square Error
평균 제곱 오류 함수



$$E = \frac{1}{3} ((w \cdot 1 - 1)^2 + (w \cdot 2 - 2)^2 + (w \cdot 3 - 3)^2)$$

(w, E)

$(w, \frac{E}{3})$

N개의 데이터

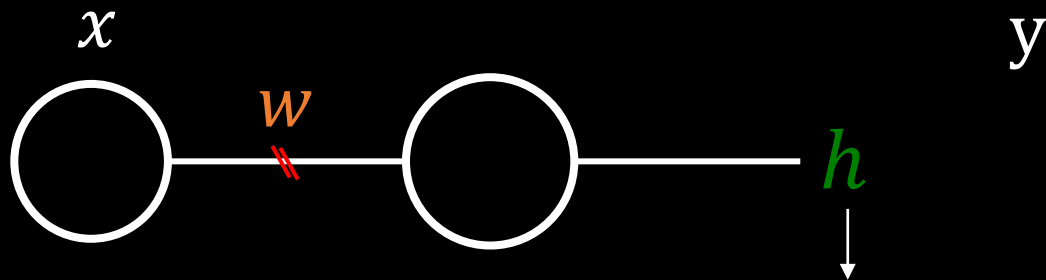
x_i	y_i
1	1
2	2
3	3

Mean Square Error
평균 제곱 오류 함수

뉴런이 대답(예측)한 값 (가설)

$$E = \frac{1}{N} \sum_{i=1}^N (wx_i - y_i)^2$$

정답

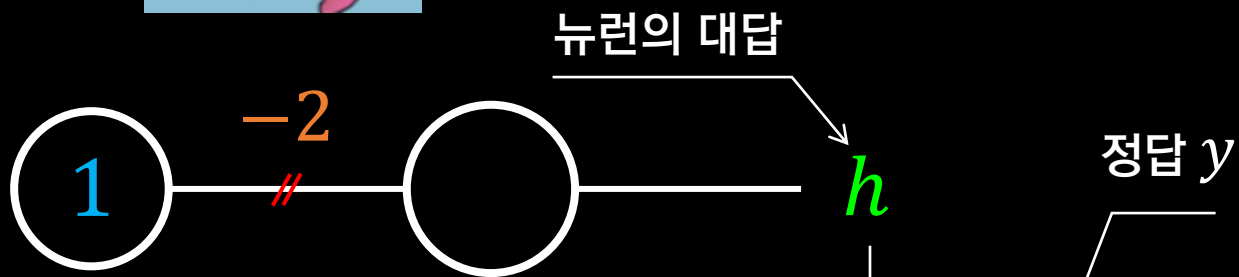


$$E = (h - y)^2$$

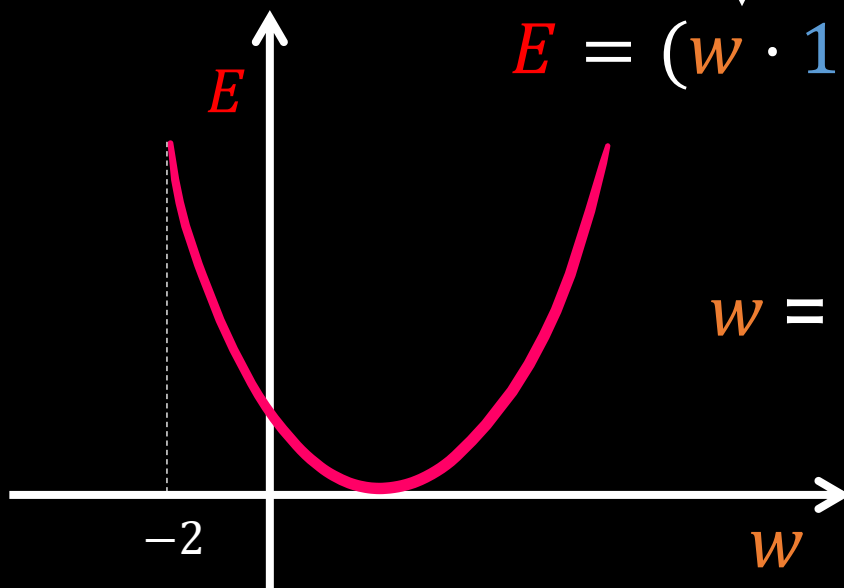
$$E = (w \cdot x - y)^2$$

x	y
1	1

$$E = (w \cdot 1 - 1)^2$$



$$E = (w \cdot 1 - 1)^2$$



$w = w - \text{기울기}$

기울기 의미 (기미)

$$\frac{\Delta E}{\Delta w}$$

- 기울기가 크다는 의미는? w 를 조금만 조절해도 E 가 많이 달라진다는 의미 $\rightarrow w$ 변화가 E 에 미치는 영향이 크다는 뜻
- 기울기가 작다는 의미는? w 를 바꿔도 E 는 별로 변하지 않는다는 의미 \rightarrow 미치는 영향이 작다는 뜻

$\frac{8}{1}$

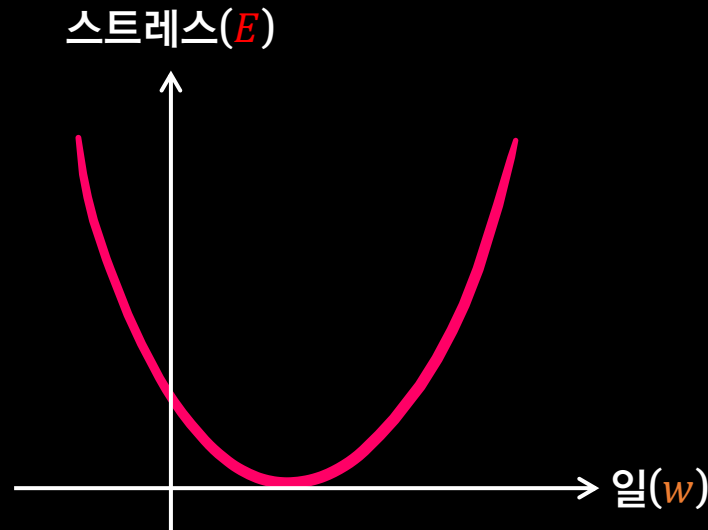
$\frac{1}{10}$

오류 그래프 E 의 w 지점에서의 기울기
 w 변화가 E 에 미치는 영향

이를 수학 공식으로 표현해보라.

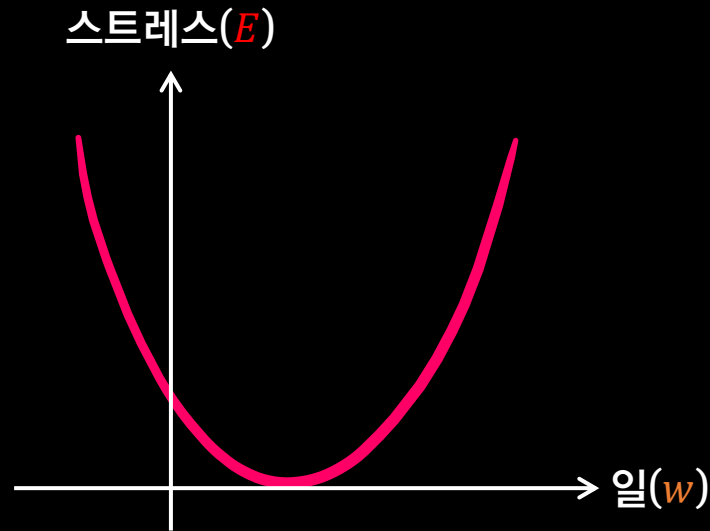
$$\lim_{\Delta w \rightarrow 0} \frac{\Delta E}{\Delta w}$$

$$= \frac{\partial E}{\partial w}$$

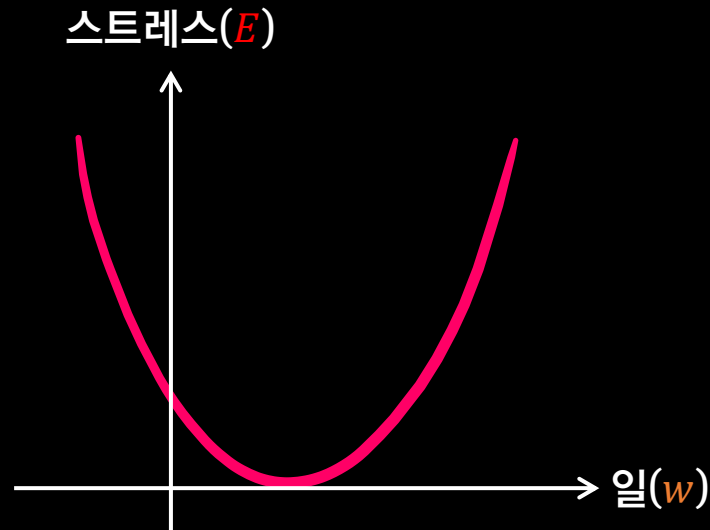


일(w)의 양의 변화가
내 스트레스(E)에 미치는 영향

w 변화가 E 에 미치는 영향



스트레스 E 가 최소가 되도록
일의 양 w 를 조절



현재 일의 양이 4일때
일의 변화가 스트레스에 미치는
영향(기울기)을 구하시오.

$$\lim_{\Delta \text{일} \rightarrow 0} \frac{\Delta \text{스트레스}}{\Delta \text{일}} \\ = \frac{\partial \text{스트레스}}{\partial \text{일}}$$

(Q) 미치는 영향 구하기

$$E = (wx - y)^2$$

데이터 (x, y) 가 $(1, 1)$ 로 주어졌을 때

$w = 3$ 인 지점에서 w 변화가 오류 E 에 미치는 영향을 구하라.

Compute the influence of w change on E when w is equal to 3.

(A1) 값을 대입하여 구하는 방법

$$E = (w \cdot 1 - 1)^2$$

$$w: 3 \rightarrow E: 4$$

$$w: 3.00001 \rightarrow E: 4.00004$$

w 를 아주 조금(0.00001) 증가 (변화량 $\Delta w = 0.00001$)

E 는 0.00004 증가 (변화량 $\Delta E = 0.00004$)

$$\frac{\Delta E}{\Delta w} = \frac{0.00004}{0.00001} = 4$$

따라서 (w 가 3인 지점에서의) 기울기 = 미치는 영향 = 4

(A2) 어떤 신기한 방법

$$E = (w \cdot 1 - 1)^2$$

$$\rightarrow 2(w \cdot 1 - 1)^1$$

따라서 $w = 3$ 이면

$$\text{그러면, } 2(3 - 1) = 4$$

(A2) 어떤 신기한 방법

$$E = (w \cdot 1 - 1)^2$$

$$\lim_{\Delta w \rightarrow 0} \frac{\Delta E}{\Delta w}$$

$$= 2(w \cdot 1 - 1)^1$$

따라서 $w = 3$ 이면

그러면, $2(3 - 1) = 4$

(A2) 어떤 신기한 방법

$$E = (w \cdot 1 - 1)^2$$

$$\lim_{\Delta w \rightarrow 0} \frac{\Delta E}{\Delta w} = \frac{\partial E}{\partial w}$$

$$= 2(w \cdot 1 - 1)^1$$

따라서 $w = 3$ 이면

그러면, $2(3 - 1) = 4$

(A2) 미분 공식을 이용한 방법

$$E = (w \cdot 1 - 1)^2$$

$$\begin{aligned}\lim_{\Delta w \rightarrow 0} \frac{\Delta E}{\Delta w} &= \frac{\partial E}{\partial w} = \frac{\partial}{\partial w} E = \frac{\partial}{\partial w} (w \cdot 1 - 1)^2 \\ &= 2(w \cdot 1 - 1)^1\end{aligned}$$

따라서 $w = 3$ 이면

그러면, $2(3 - 1) = 4$

오류 함수가 다음과 같을 때
 $w = 3$ 인 곳에서의 기울기는?

$$E = w^3$$

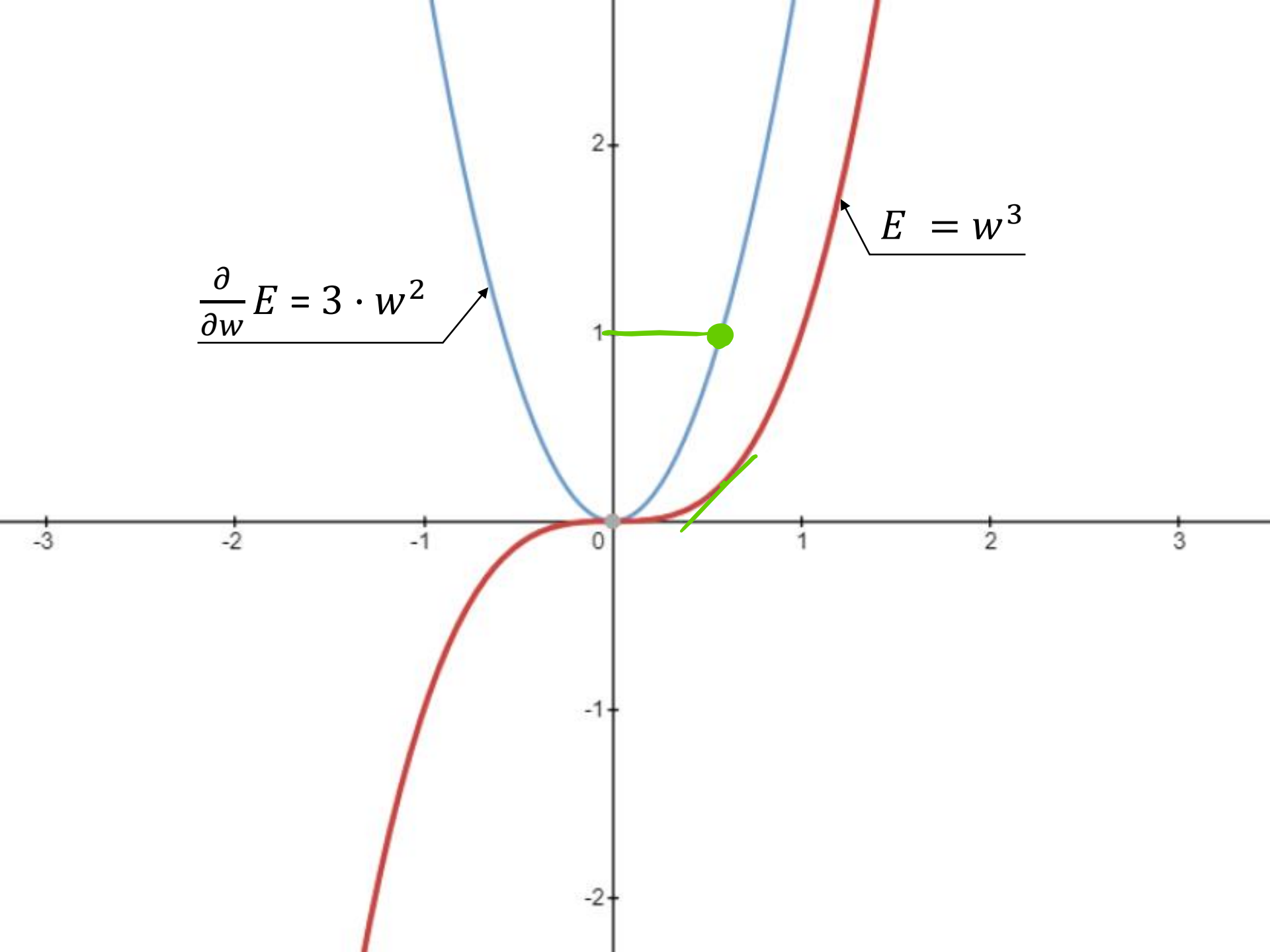
$$\frac{\partial}{\partial w} E = ?$$

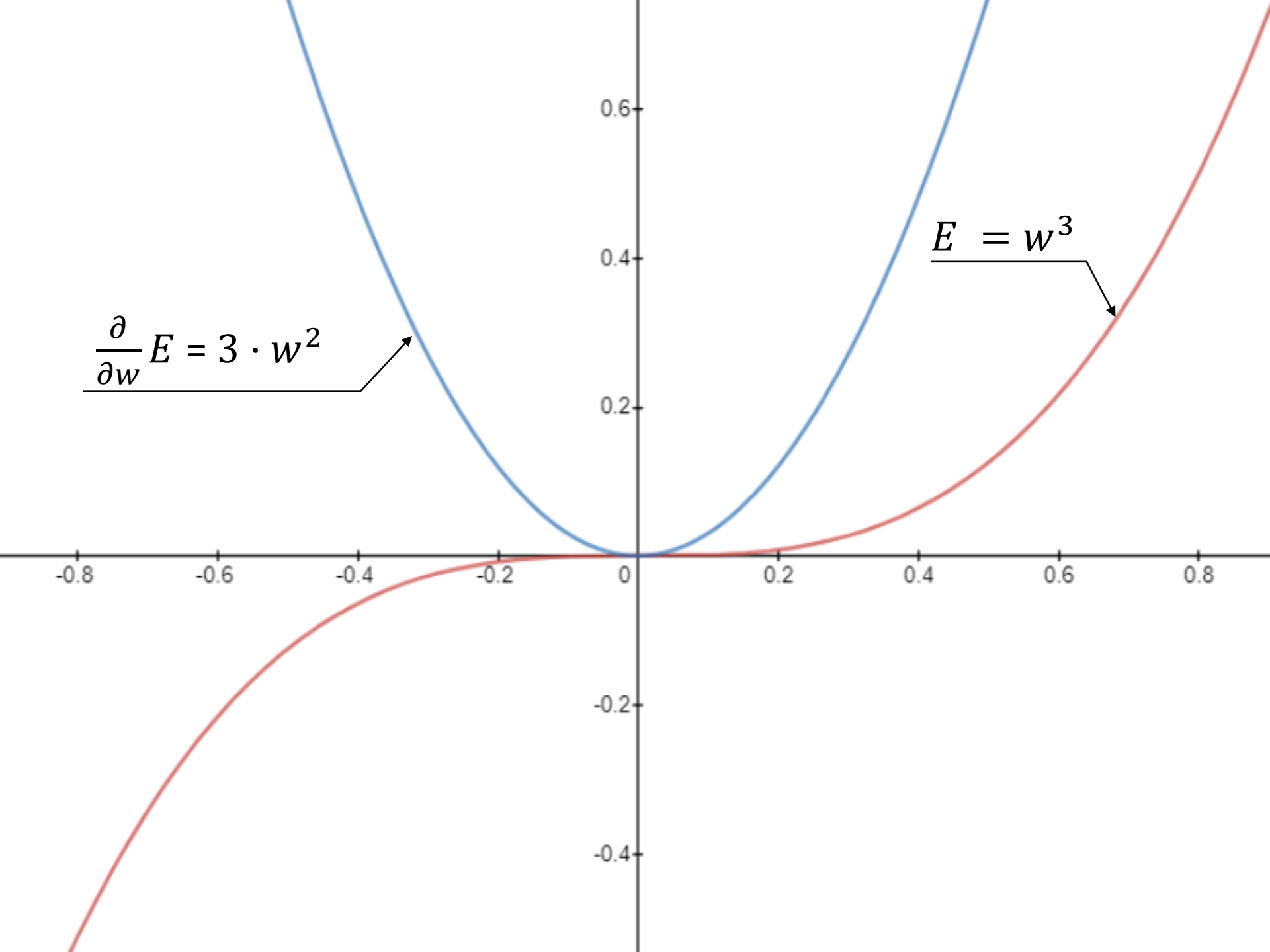
오류 함수가 다음과 같을 때
 $w = 3$ 인 곳에서의 기울기는?

$$E = w^3$$

$$\begin{aligned}\frac{\partial E}{\partial w} &= \frac{\partial}{\partial w} E = 3 \cdot w^2 \\ &= 27\end{aligned}$$









뉴턴(1642~1727)과 라이프니츠(1646~1716),
영국과 독일의 수학자이면서 과학자. 이들의 공통점은?
미분법을 발명했다는 것!

학습(Learning)은

- 오류 E 가 줄어들도록 w 조절
- 경사하강, $w = w - a \cdot 기울기$
- 파라미터 w 튜닝

이번 학습에서는

- 경사하강 방법을 알 수 있다.
- 오류 그래프가 갖는 문제점을 파악할 수 있다.
- 기울기 의미를 이해할 수 있다.