AI and Deep Learning

선형회귀

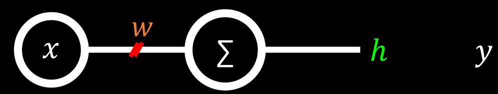
오류 그래프와 기울기, 그리고 경사하강(2/2)

제주대학교
변 영 철

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 w를 ‘**자동으로**’
조절할 것인가?

(학습, Learning)

- 오류 E가 줄어들도록 w를 조절

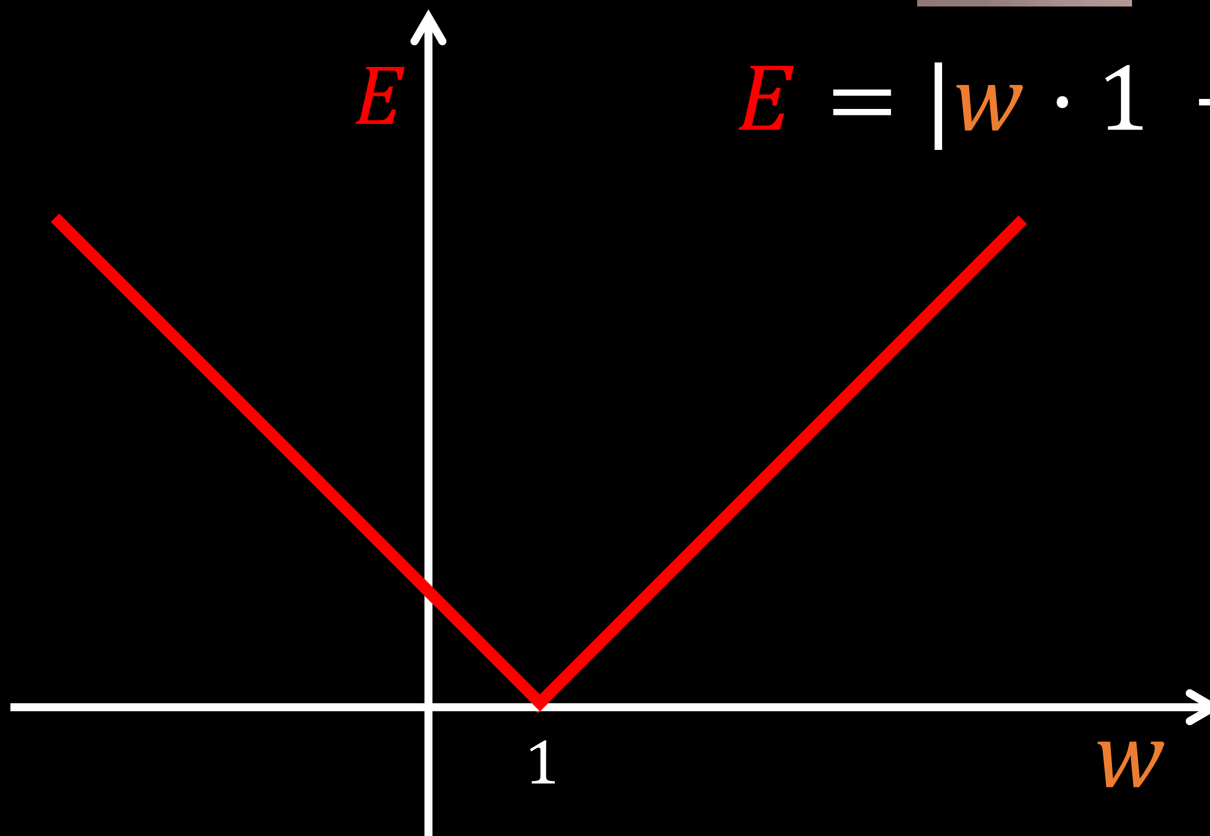
→ 그러면 뉴런 대답이 정답에 가까워 짐.

www.desmos.com

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

절대값 오류 함수 L1

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



오류가 줄어들도록

w 조절하는 법

“기울기를 이용할 수 있지 않을까?”

경사하강 하도록 w 조절

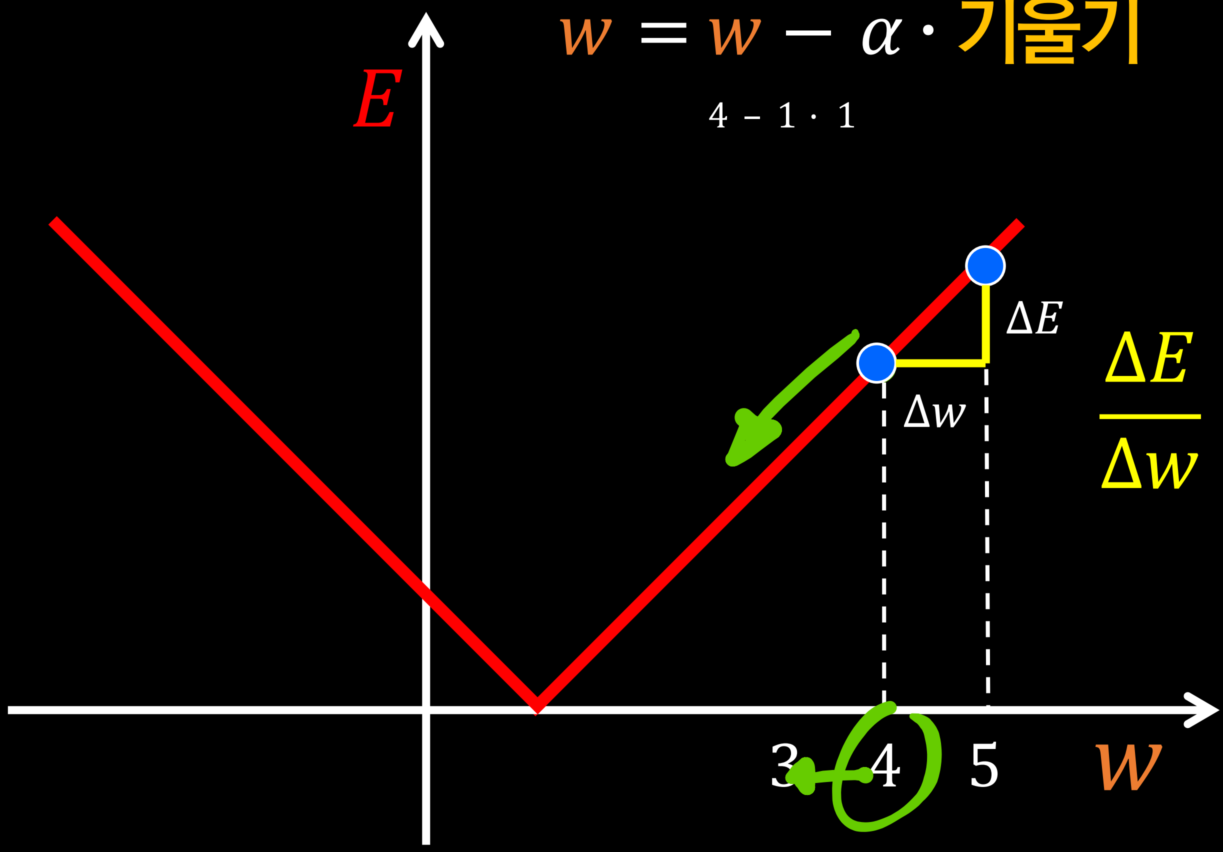
오류가 줄어들도록

w 조절하는 법

반영 비율(가령, 1)

$w = w - \alpha \cdot$ 기울기

$4 - 1 \cdot 1$



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

오류가 줄어들도록

w 조절하는 법

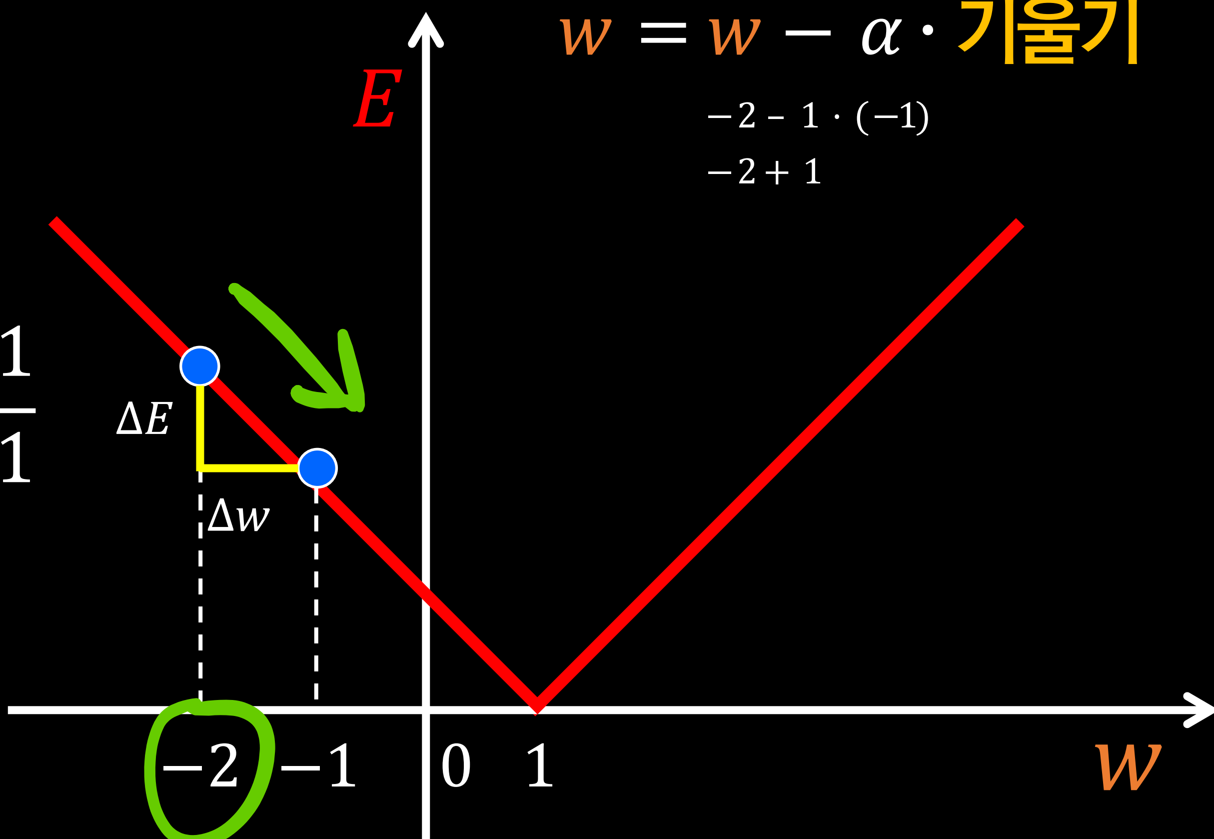
반영 비율(가령, 1)

$$w = w - \alpha \cdot \text{기울기}$$

$-2 - 1 \cdot (-1)$

$-2 + 1$

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



w 를 어떻게 조절할까? = 경사하강 수식

어디에서의 기울기?

w지점에서의 기울기

$$w = w - \alpha \cdot \text{기울기}$$

$$\boxed{4} \quad 1 \quad 1$$

$$w - \alpha \cdot \text{기울기} \rightarrow w$$

$$4 - 1 \cdot 1 \longrightarrow 3$$ error E = 2

$$3 - 1 \cdot 1 \longrightarrow 2$$ error E = 1

$$2 - 1 \cdot 1 \longrightarrow 1$$ error E = 0

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

$$w - \alpha \cdot \text{기울기} \rightarrow w$$

$$-2 - 1 \cdot (-1) \longrightarrow -1$$

error E = 2

$$-1 - 1 \cdot (-1) \longrightarrow 0$$

error E = 1

$$0 - 1 \cdot (-1) \longrightarrow 1$$

error E = 0

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

$$w - \alpha \cdot \text{기울기} \longrightarrow w$$

$$\begin{aligned}
-2 - 2 \cdot (-1) &\longrightarrow 0 \\
0 - 2 \cdot (-1) &\longrightarrow 2 \\
2 - 2 \cdot (1) &\longrightarrow 0 \\
0 - 2 \cdot (-1) &\longrightarrow 2
\end{aligned}$$

error E = 1

error E = 1

error E = 1

error E = 1

절대값 오류, 어떤 문제?

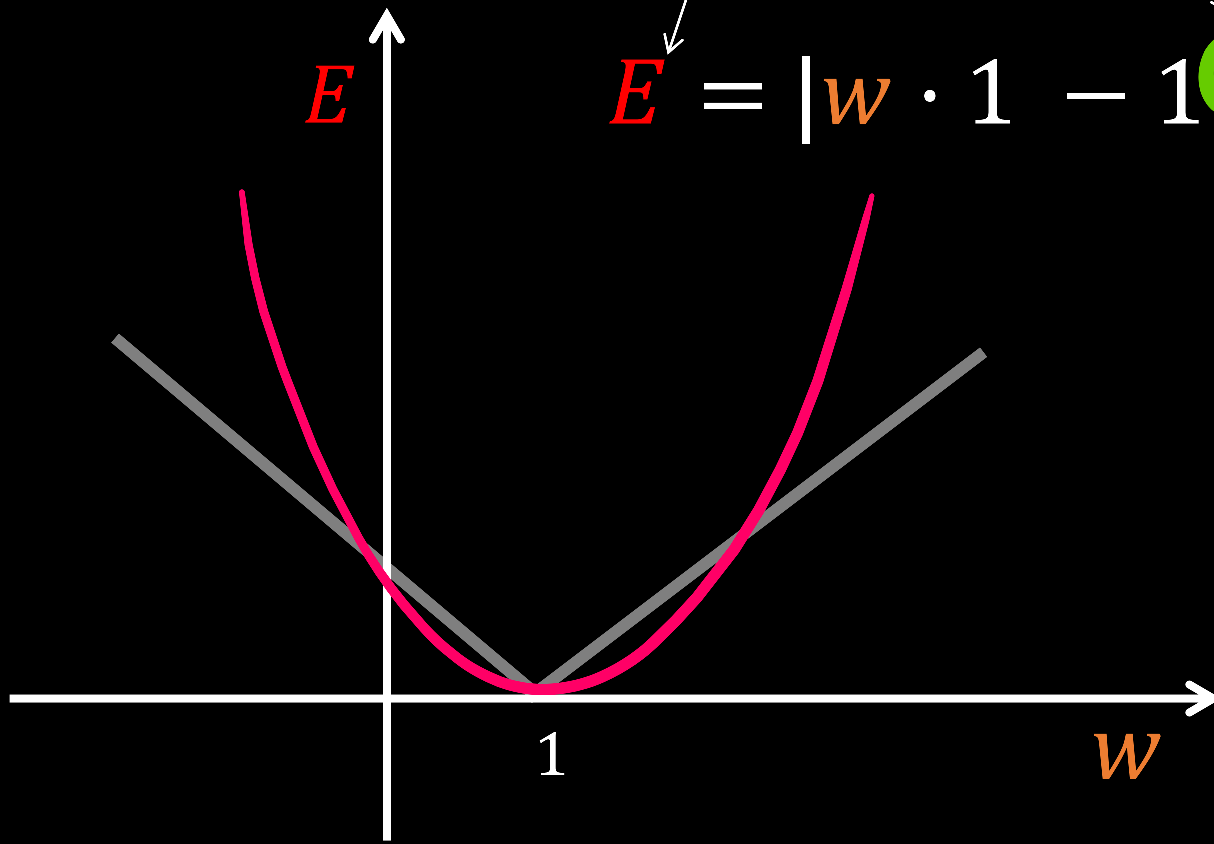
- 왼쪽, 오른쪽, 각각은 항상 같은 기울기, 따라서 항상 같은 경사하강 속도
- w값, 조절 실패할 수도
- 기울기 값만 보고 w값을 짐작할 수 없음.
- w가 1인 지점에서는 기울기 구할 수 없음.

제곱 오류 (L2 Loss)

square(제곱)

error

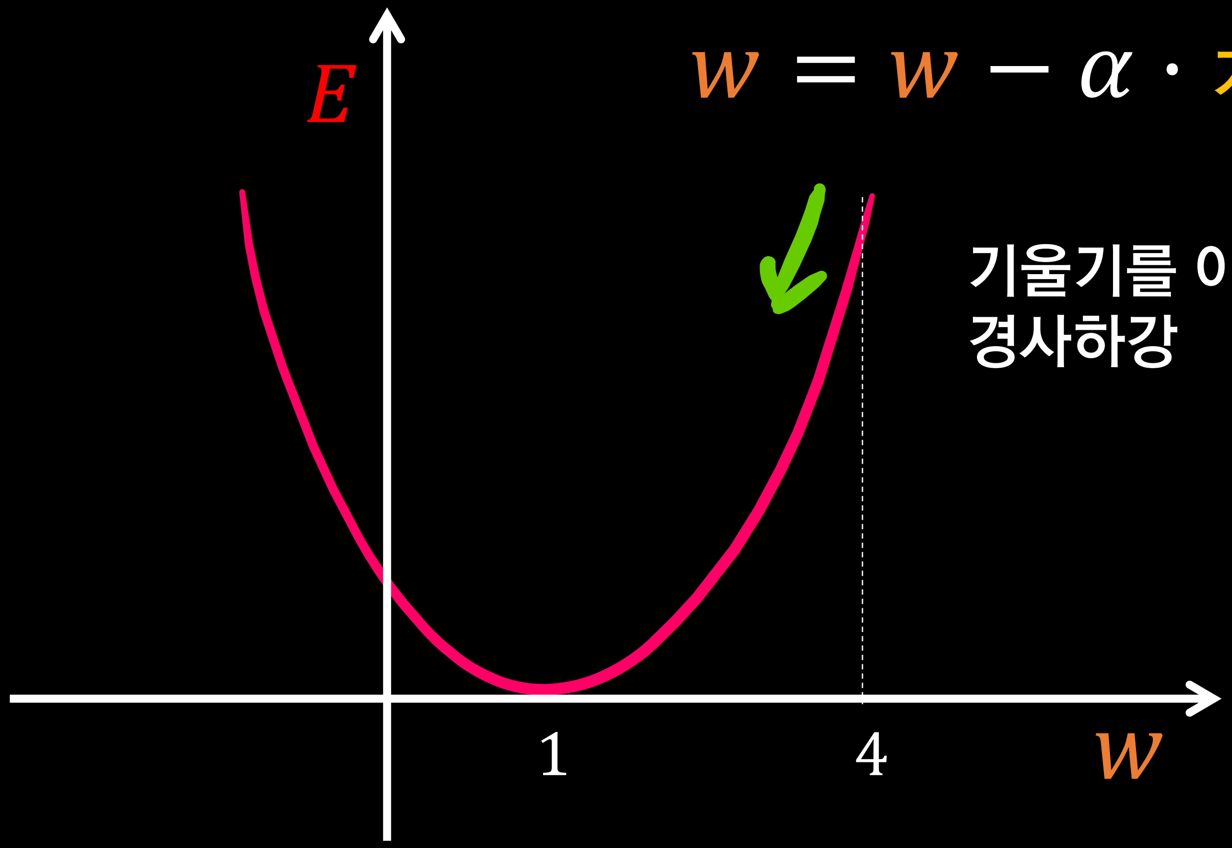
$$E = |w \cdot 1 - 1|^2$$



w 조절하는 법

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

$$w = w - \alpha \cdot \text{기울기}$$



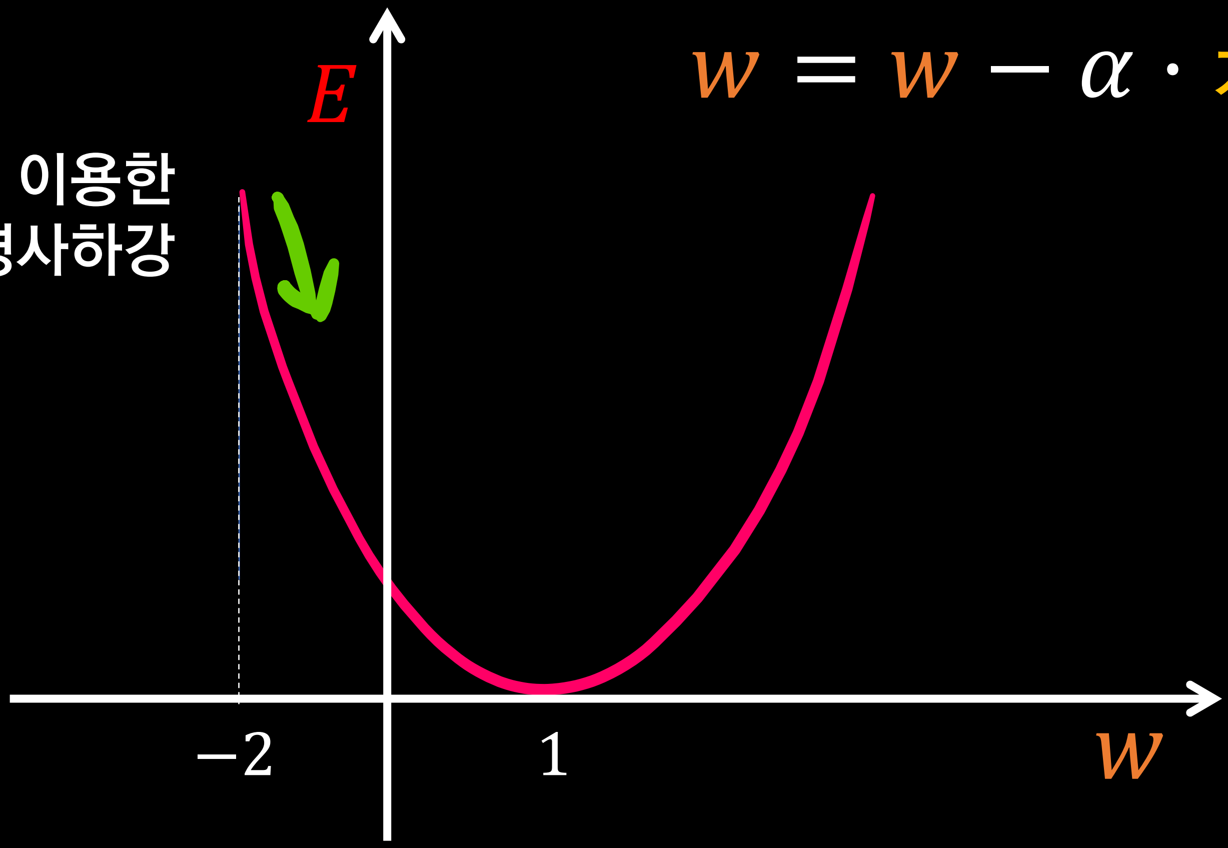
기울기를 이용한
경사하강

w 조절하는 법

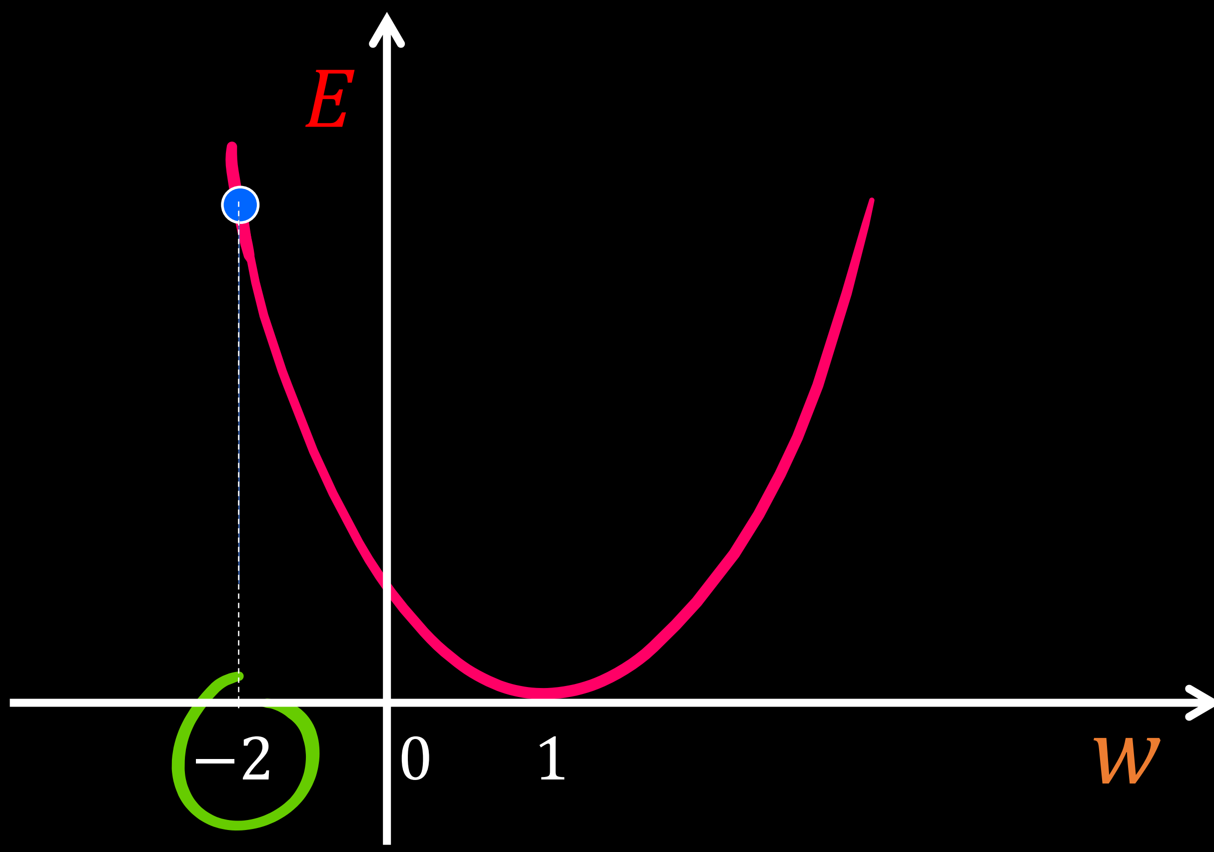
$$\frac{\Delta E}{\Delta w}$$

$$w = w - \alpha \cdot \text{기울기}$$

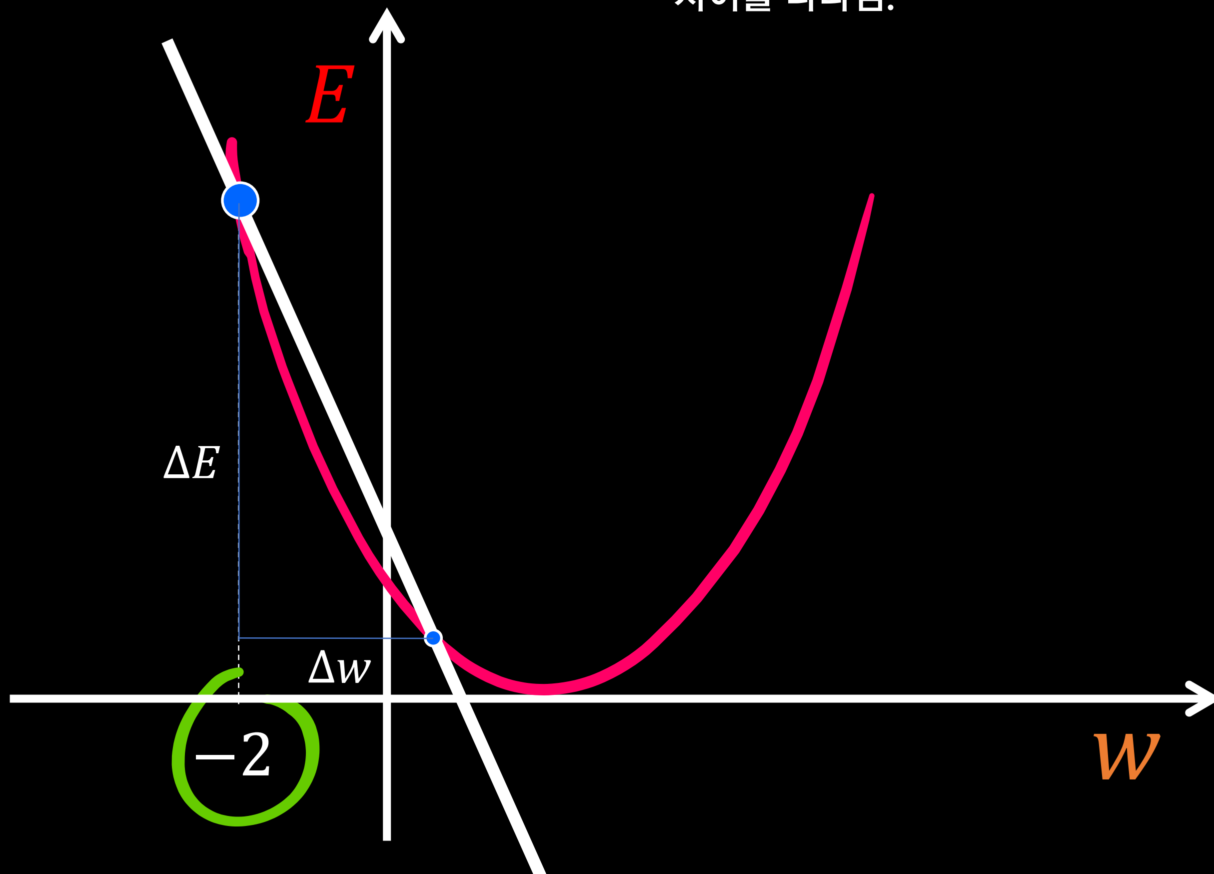
기울기를 이용한
경사하강

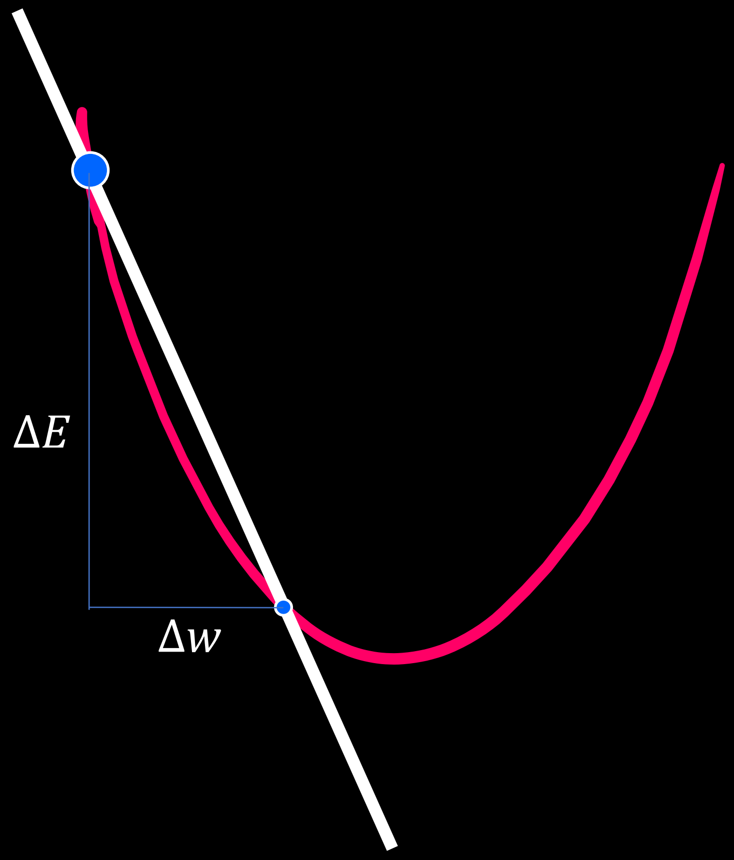


기울기는 어떻게 구할까?

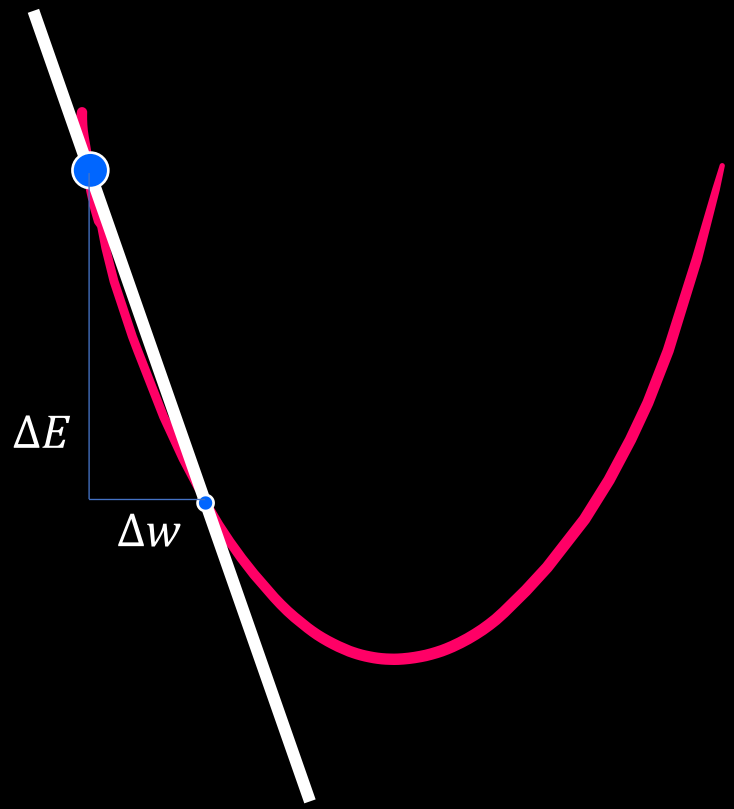


- 대문자 Δ, 소문자 δ 는 그리스 문자 중 네 번째 글자로, 로마자 D, d가 여기에서 비롯됨.
- 차이나 변화를 나타내는 기호로 많이 사용됨. 예를 들어, 델타 w (Δw)는 w의 작은 변화나 차이를 나타냄.

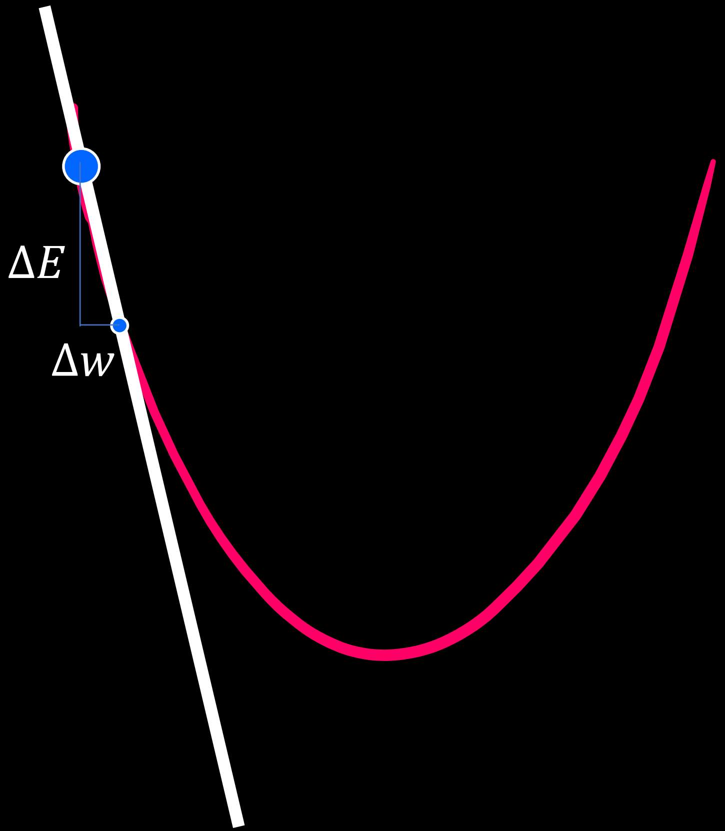




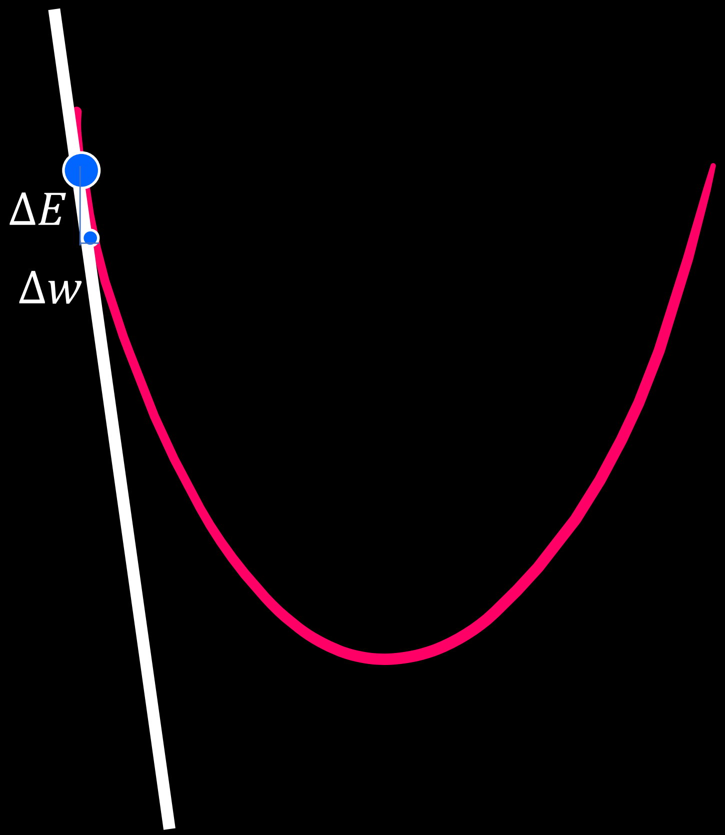
ΔE
Δw

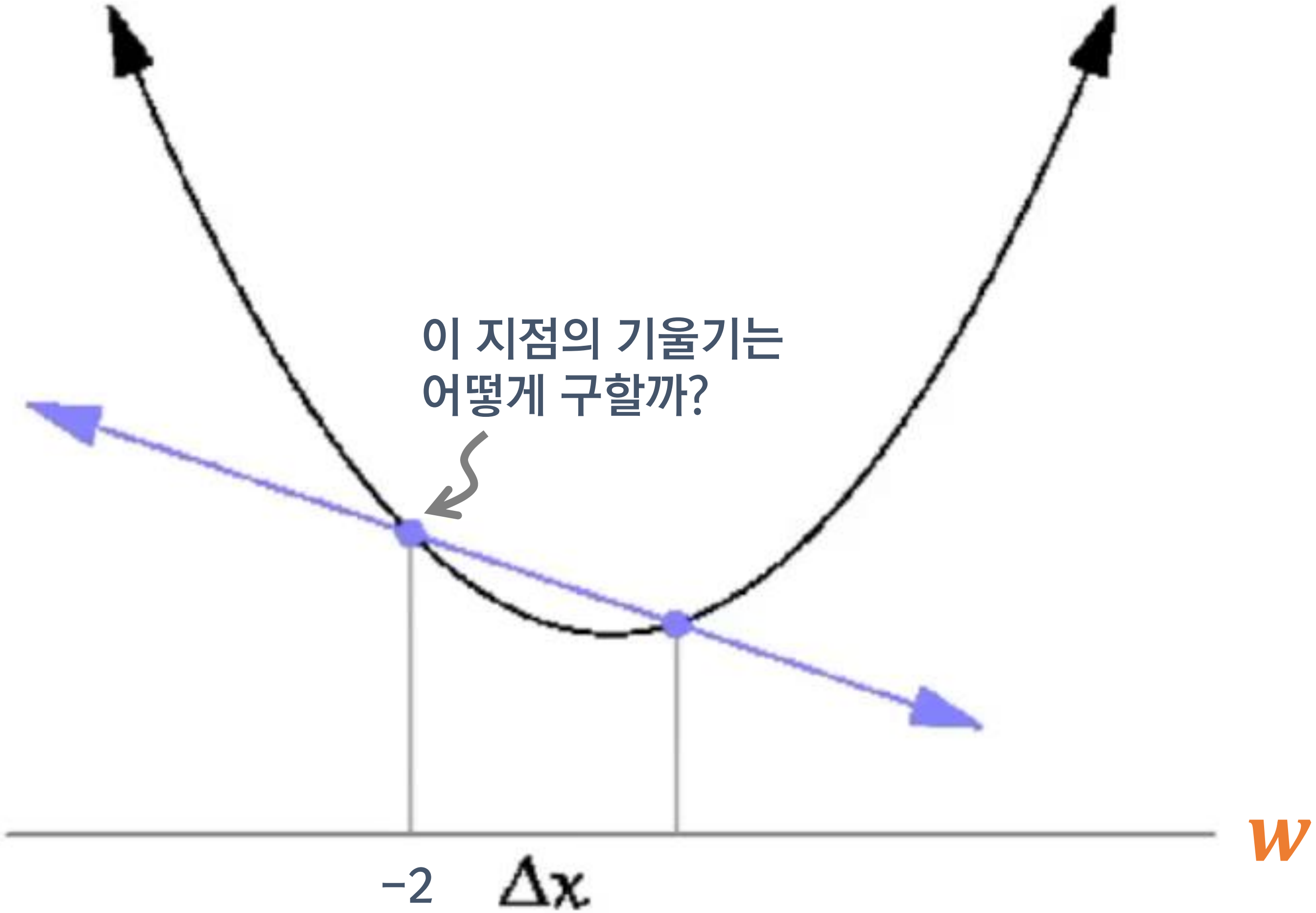


ΔE
Δw


ΔE
Δw

접선·Tangent line





$\Delta w(\Delta x) \rightarrow 0$

접선으로 정확한
기울기를 구하기 위함.

lim : limit, 극한,
무한히 접근한다.

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

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

w를 아주 조금만 늘렸을 때 (∂w)
E는 얼마나 늘어나는지 (∂E)
두 값으로 구한 값, 기울기

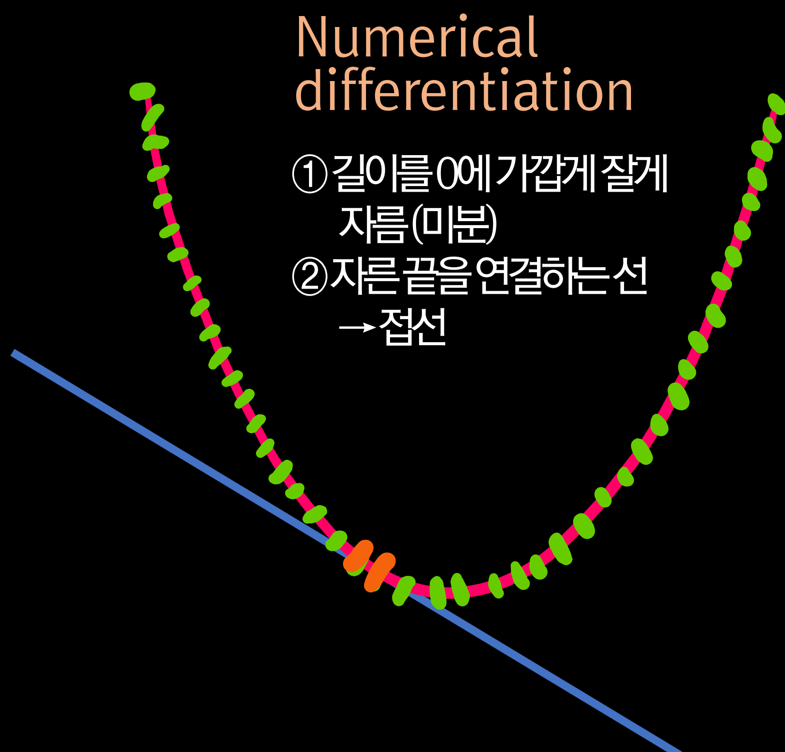
아래 표현은 무슨 의미인지?

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

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

w를 *아주 조금만* 늘렸을 때 (∂w)
E는 얼마나 늘어날까? (∂E)

아주 잘게 자름



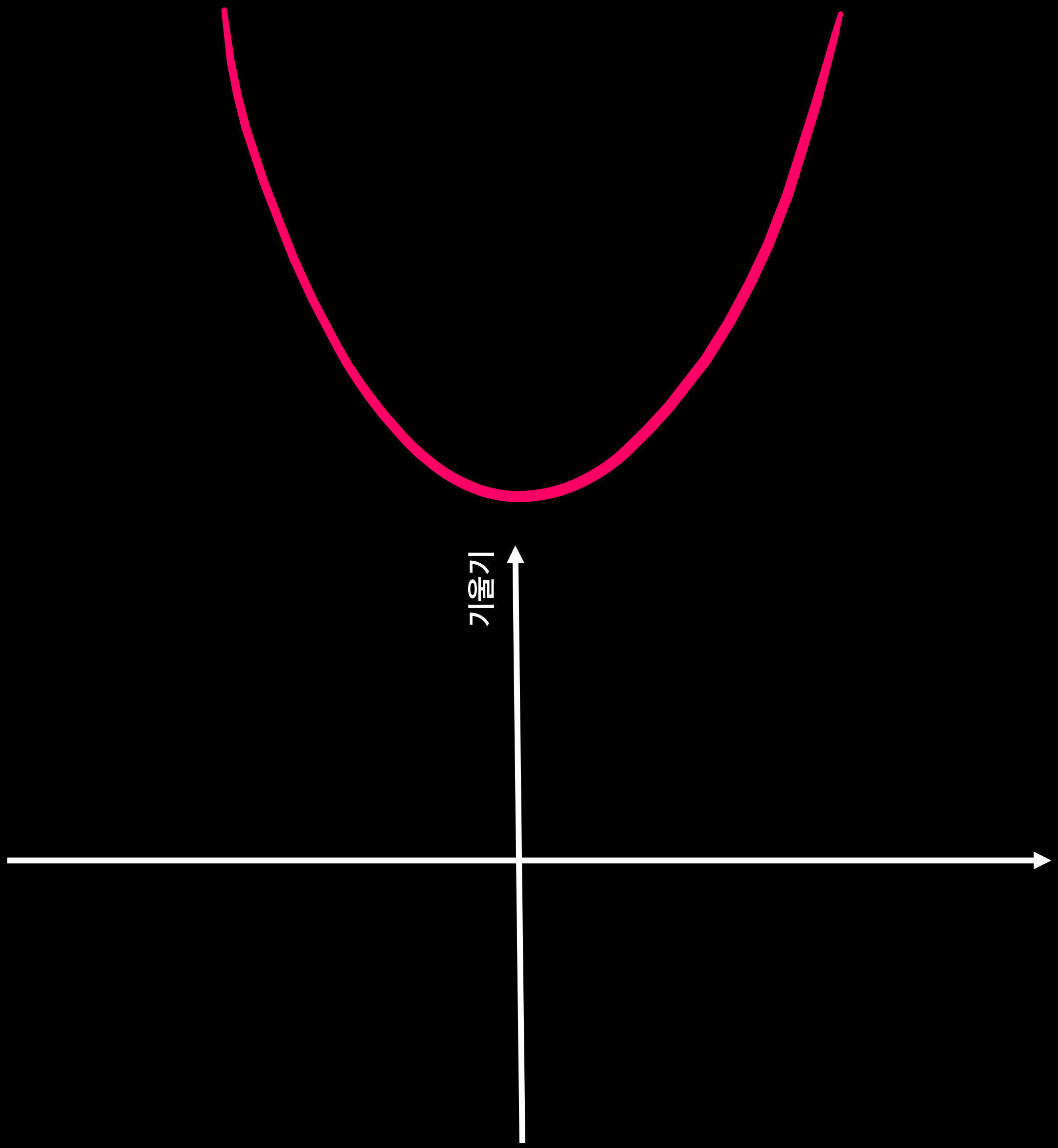
$$\lim_{\Delta w \to 0} \frac{\Delta E}{\Delta w}$$

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

= 결과적으로 선이 아주 작게 나뉨

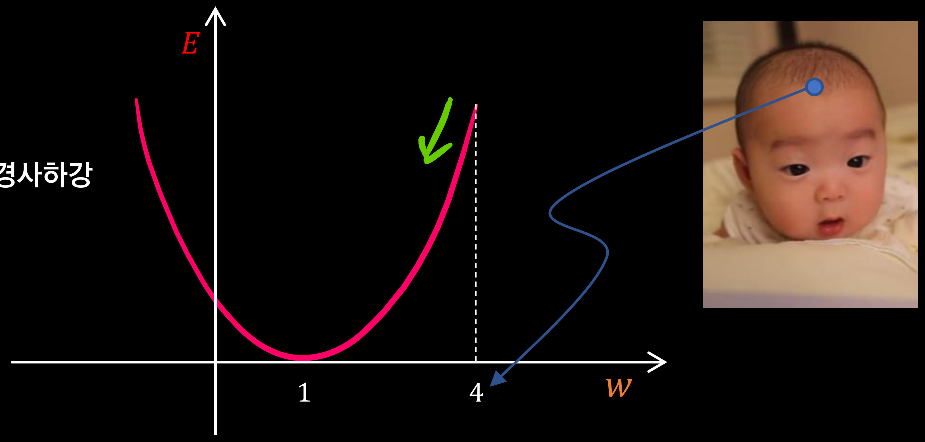
아주 잘게 자름 → 미분(微分)

아래 그래프의 모든 곳에서의 기울기를 구하시오(미분).



아래 그래프에 대해 기울기를 구하시오(미분).

경사하강



$$w = w - \alpha \cdot \text{기울기}$$

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

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

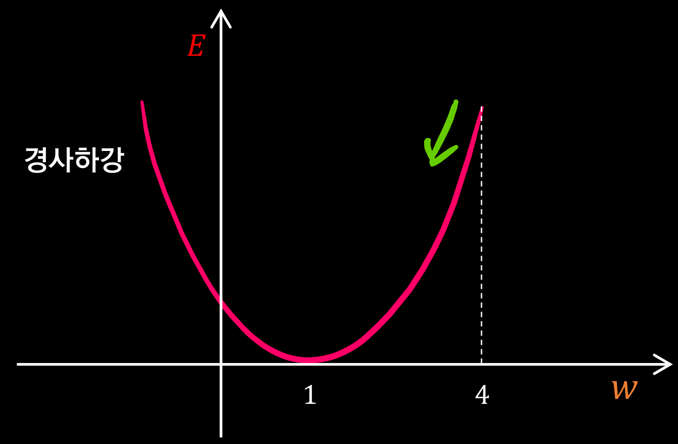
학습이란 무엇을 하는 것인가?
학습을 위한 경사하강 수식을 작성하시오.

정답은?

학습이란 무엇을 하는 것인가?
학습을 위한 경사하강 수식을 작성하시오.

$$w = w - \alpha \cdot \text{기울기}$$

제곱 오류(L2) 장점



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

Error Function

- 모델의 Loss를 구하는 방법으로 L1, L2 loss가 사용될 때 아래의 식을 따른다.

L1 loss

- L1 loss를 보면, 식처럼 실제 값 y_i와 예측값 $f(x_i)$ 사이의 차이값에 절댓값을 취해 그 오차 합을 최소화하는 방향으로 loss를 구한다.
- Least Absolute Deviations, LAD라고도 한다.

$$L = \sum_{i=1}^{n} |y_i - f(x_i)|$$

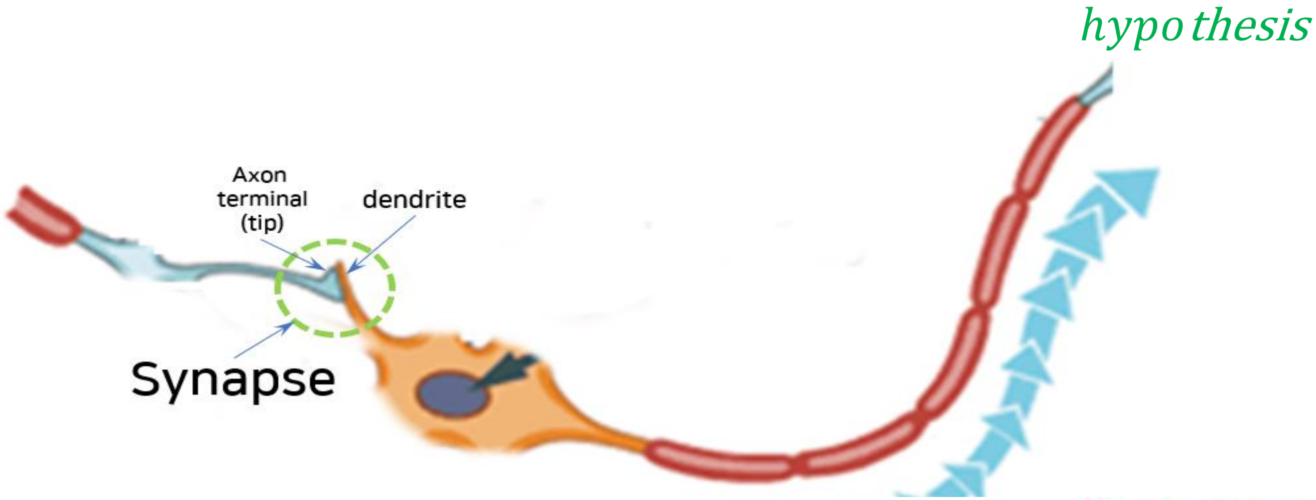
L2 loss

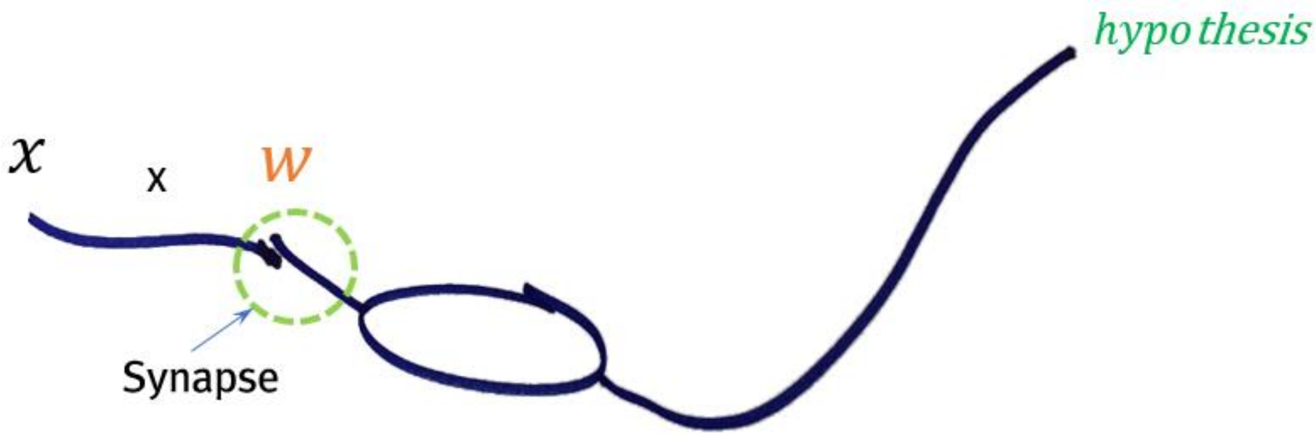
- L2 loss는 MSE(Mean Square Error)를 안다면 아주 익숙한 개념으로 target value인 실제값 y_i와 예측값 $f(x_i)$ 사이의 오차를 제곱한 값들을 모두 합하여 loss로 계산한다.
- Least square error, LSE라고도 한다.

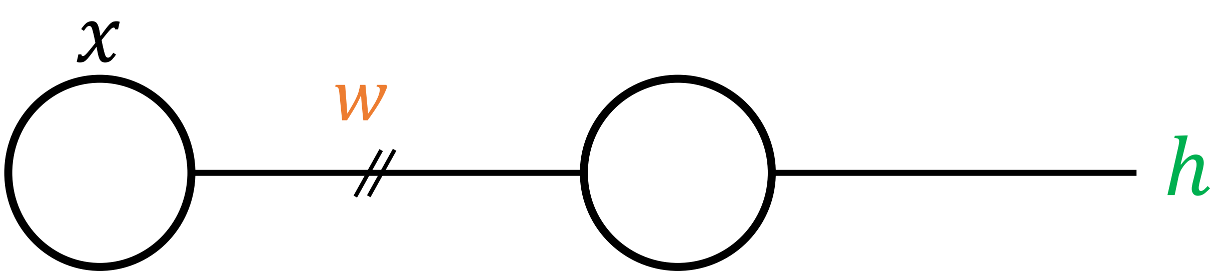
$$L = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

절대값 오류(L1)에서는?

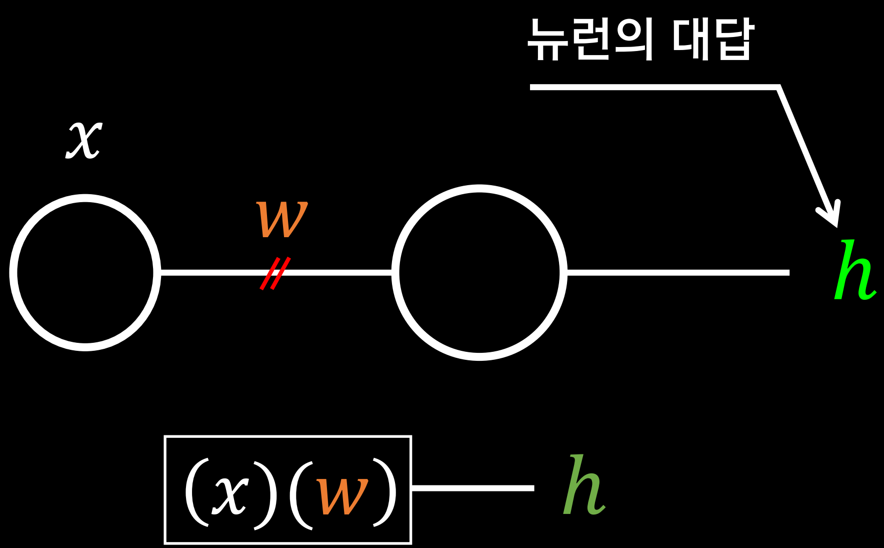
- 왼쪽, 오른쪽, 각각은 항상 같은 기울기
- 따라서 항상 같은 경사하강 속도
- w 값, 조절 실패할 수도
- 기울기 값만 보고 w값을 짐작할 수 **없음**.
- w 가 1일 때는 기울기 구할 수 **없음**.


hypo thesis
Axon
terminal
(tip)
dendrite
Synapse


hypo thesis
x
x
w
Synapse


x
w
h

다음과 같은 뉴런(신경세포)이 있다고 하자.
이 뉴런이 표현(대답)하는 회귀를 그려보자.

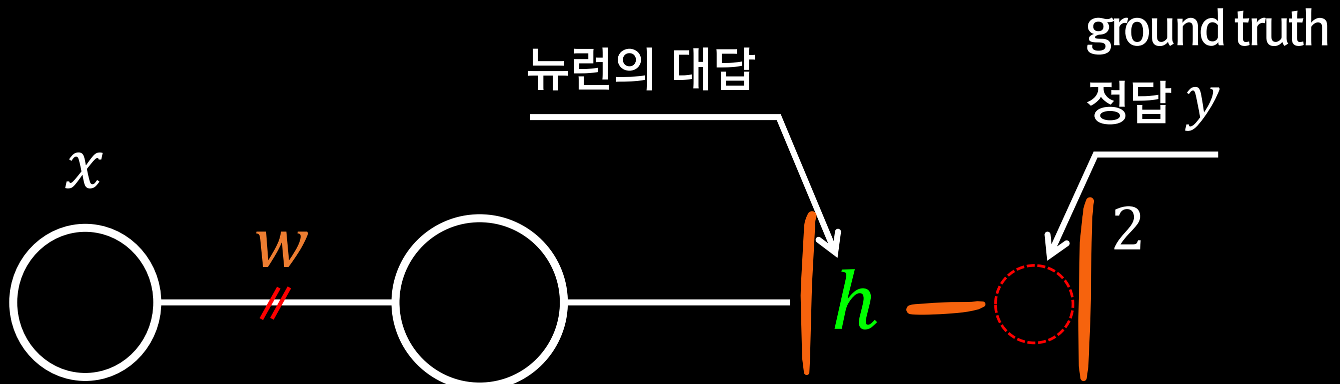


desmos

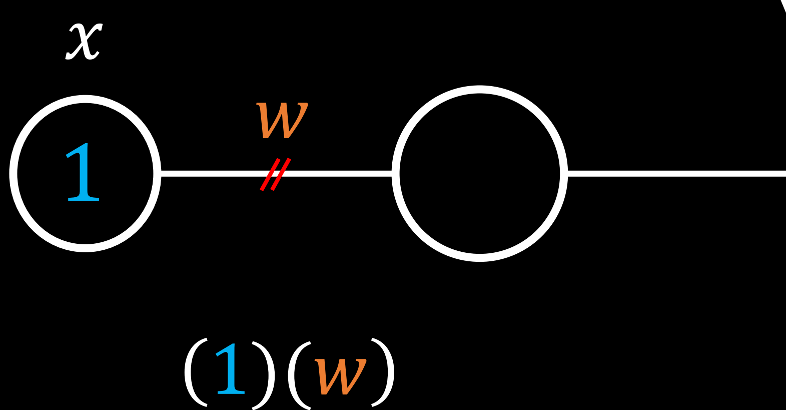
데이터 (1, 1) 표시

$h = w \cdot x$

데이터가 $(1, 1)$일 때 w에 따른 제곱 오류 함수 E는?



데이터가 $(1, 1)$일 때 w에 따른 제곱 오류 함수 E는?



뉴런의 대답

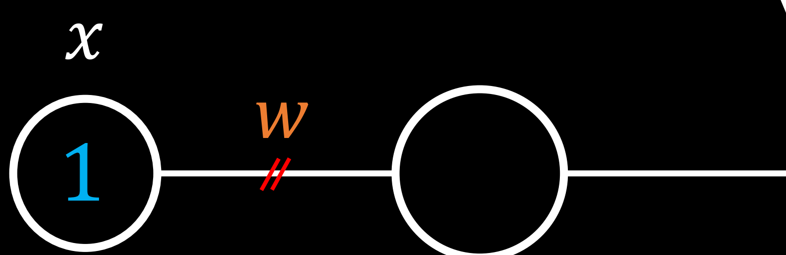
ground truth
정답 y

$$|h - 1|^2 = E$$

데이터가 $(1, 1)$일 때 w에 따른 L2 E는?

csv 파일

x_i	y_i
1	1



뉴런의 대답

ground truth
정답 y

$$\left| h - 1 \right|^2 = E$$

$$(1)(w)$$

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

desmos

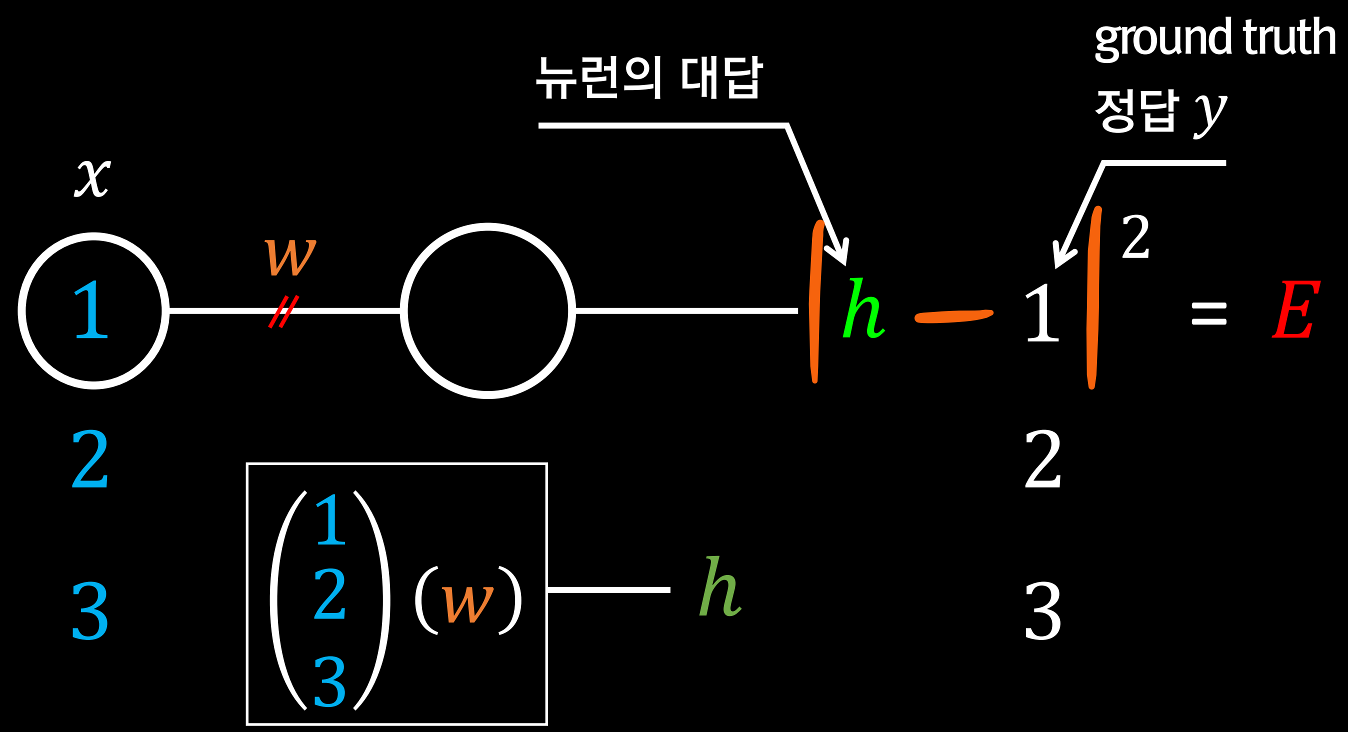
$$E = (w \cdot 1 - 1)^2$$

$$(w, E)$$

데이터가 $(1, 1), (2, 2), (3, 3)$이면?

csv 파일

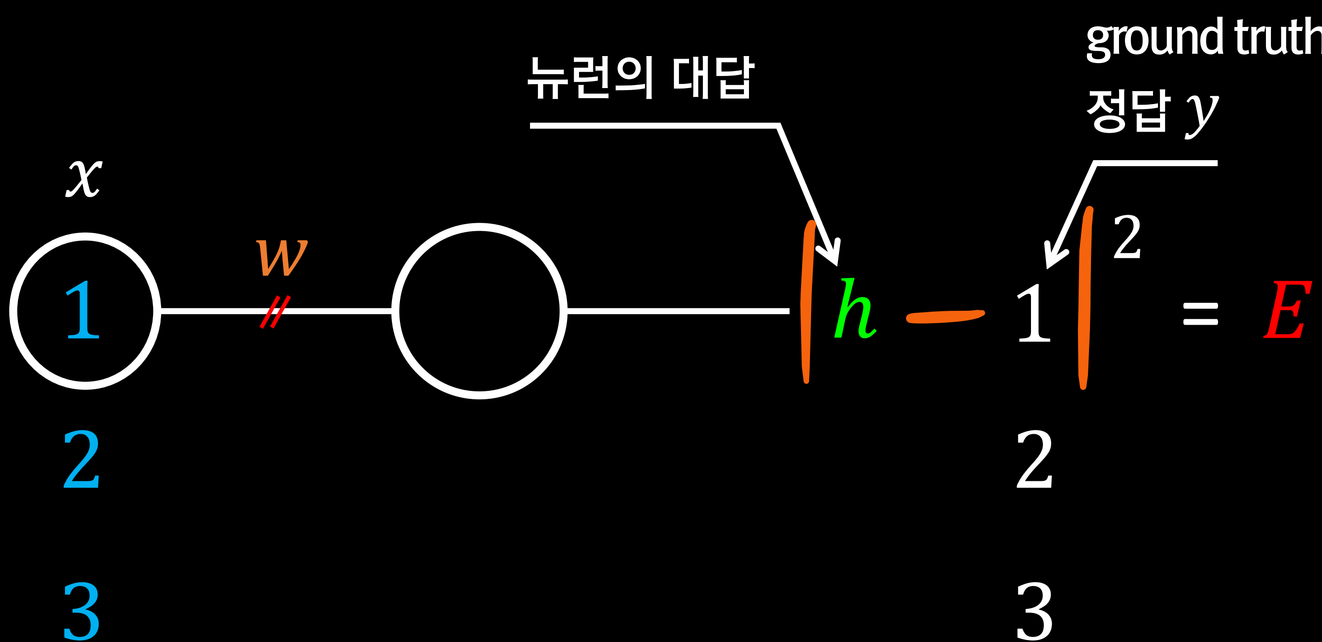
x_i	y_i
1	1
2	2
3	3



데이터가 너무 많으면?
데이터를 나눠서.. batch_size

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

desmos

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

$$(w, E)$$

$$(w, \frac{E}{5})$$

N개의 데이터

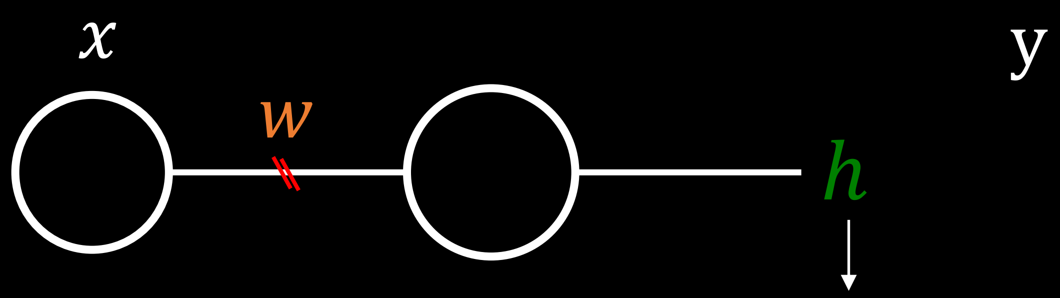
x_i	y_i
1	1
2	2
3	3

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



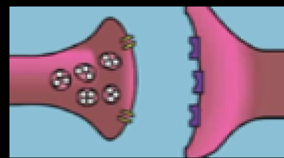
$$E = (h - y)^2$$

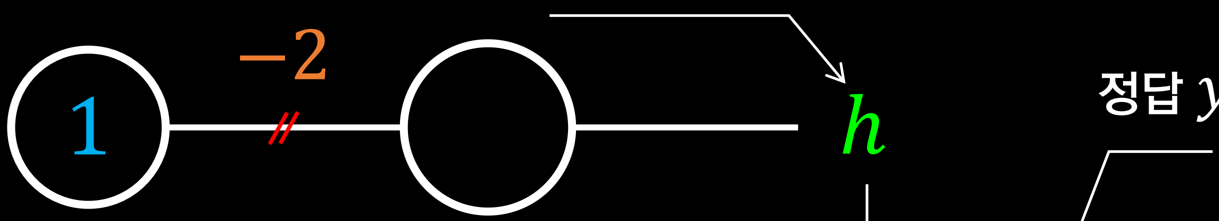
$$E = (w \cdot x - y)^2$$

if

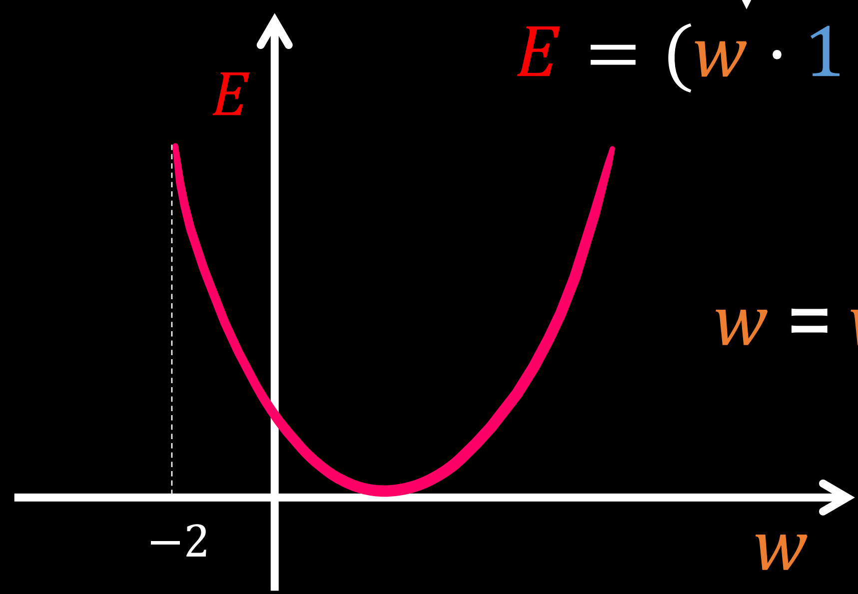
x	y
1	1

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



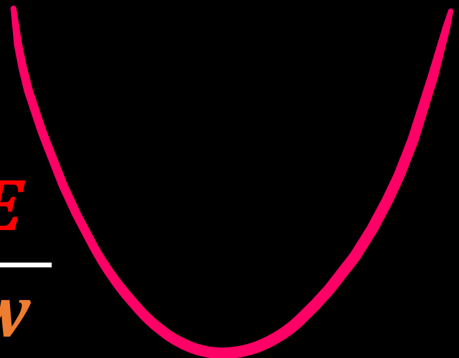


$$E = (w \cdot 1 - 1)^2$$



$$w = w - \text{기울기}$$

기울기 의미 (기울기는 미치는 영향) $\frac{\Delta E}{\Delta w}$



$\frac{8}{1}$

$\frac{1}{10}$

- 기울기가 크다는 의미는?

  w를 조금만 조절해도 E가 많이 달라진다는 의미 → w 변화가 E에 미치는 영향이 크다는 뜻

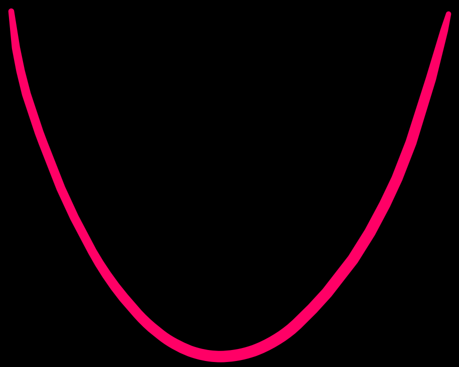
- 기울기가 작다는 의미는?

  w를 바꿔도 E는 별로 변하지 않는다는 의미 → 미치는 영향이 작다는 뜻

오류 그래프 E 의 w 지점에서의 기울기

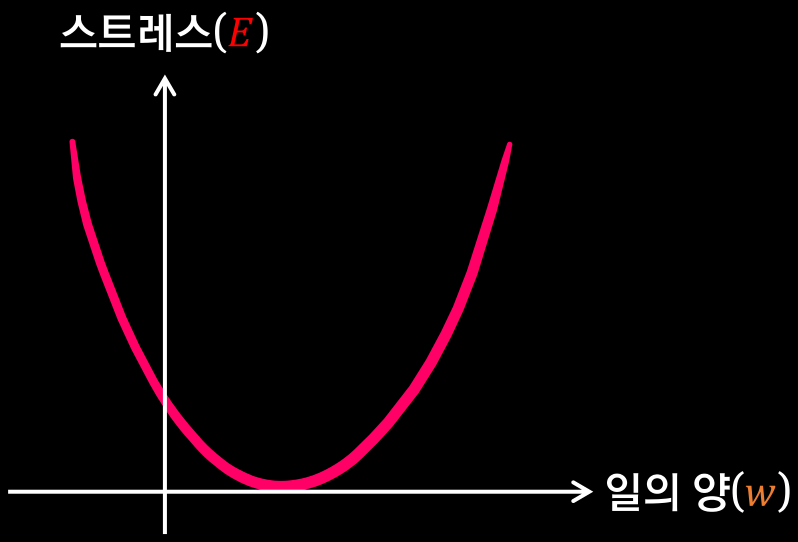
→ w 변화가 E 에 미치는 영향

이를 수학 공식으로 표현해보라.



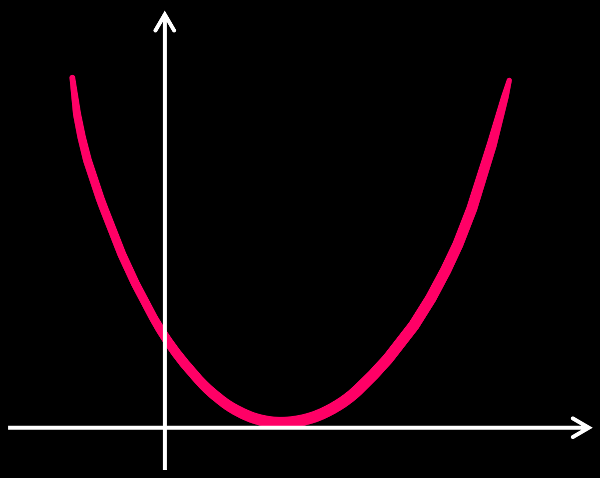
$$\lim_{\Delta w \to 0} \frac{\Delta E}{\Delta w}$$

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

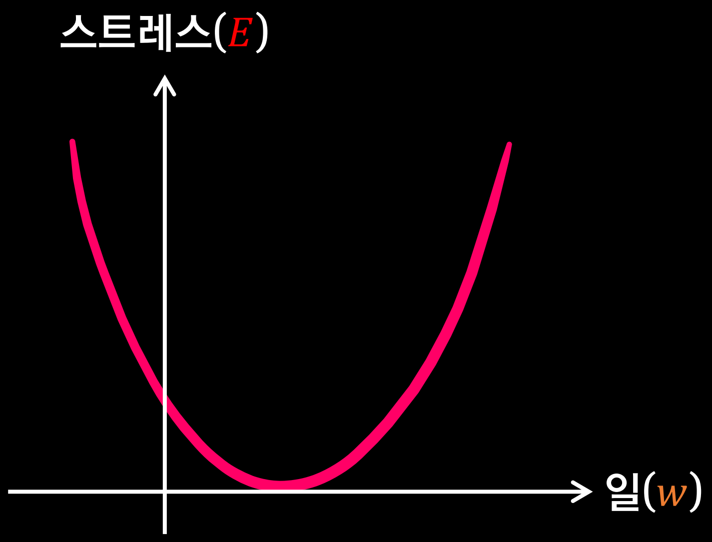


일(w)의 양의 변화가 내 스트레스(E)에 미치는 영향

w 변화가 E에 미치는 영향



스트레스 E가 최소가 되도록
일의 양 w를 조절



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

$$\lim_{\Delta 일 \to 0} \frac{\Delta 스트레스}{\Delta 일}$$

$$= \frac{\partial 스트레스}{\partial 일}$$

(Q) 미치는 영향 구하기

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

데이터 (x, y)가 $(1, 1)$로 주어졌을 때
$w = 3$인 지점에서 w변화가 오류 E에 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 → E: 4
w: 3.00001 → E: 4.00004
w를 아주 조금(0.00001) 증가 (변화량 Δw=0.00001)
E는 0.00004 증가 (변화량 ΔE=0.00004)

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

따라서 (w가 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

그러면, $2(3 - 1) = 4$

(A2) 어떤 신기한 방법

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

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

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

따라서 $w = 3$이면

그러면, $2(3 - 1) = 4$

(A2) 어떤 신기한 방법

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

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

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

따라서 w = 3이면

그러면, $2(3 - 1) = 4$

(A2) 미분 공식을 이용한 방법

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

$$\lim_{\Delta w \to 0} \frac{\Delta E}{\Delta w} = \frac{\partial E}{\partial w} = \frac{\partial}{\partial w} E = \frac{\partial}{\partial w} (w \cdot 1 - 1)^2$$

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

따라서 w = 3이면

그러면, $2(3 - 1) = 4$

오류 함수가 다음과 같을 때
w = 2인 곳에서의 기울기는?

$$E = w^3$$

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

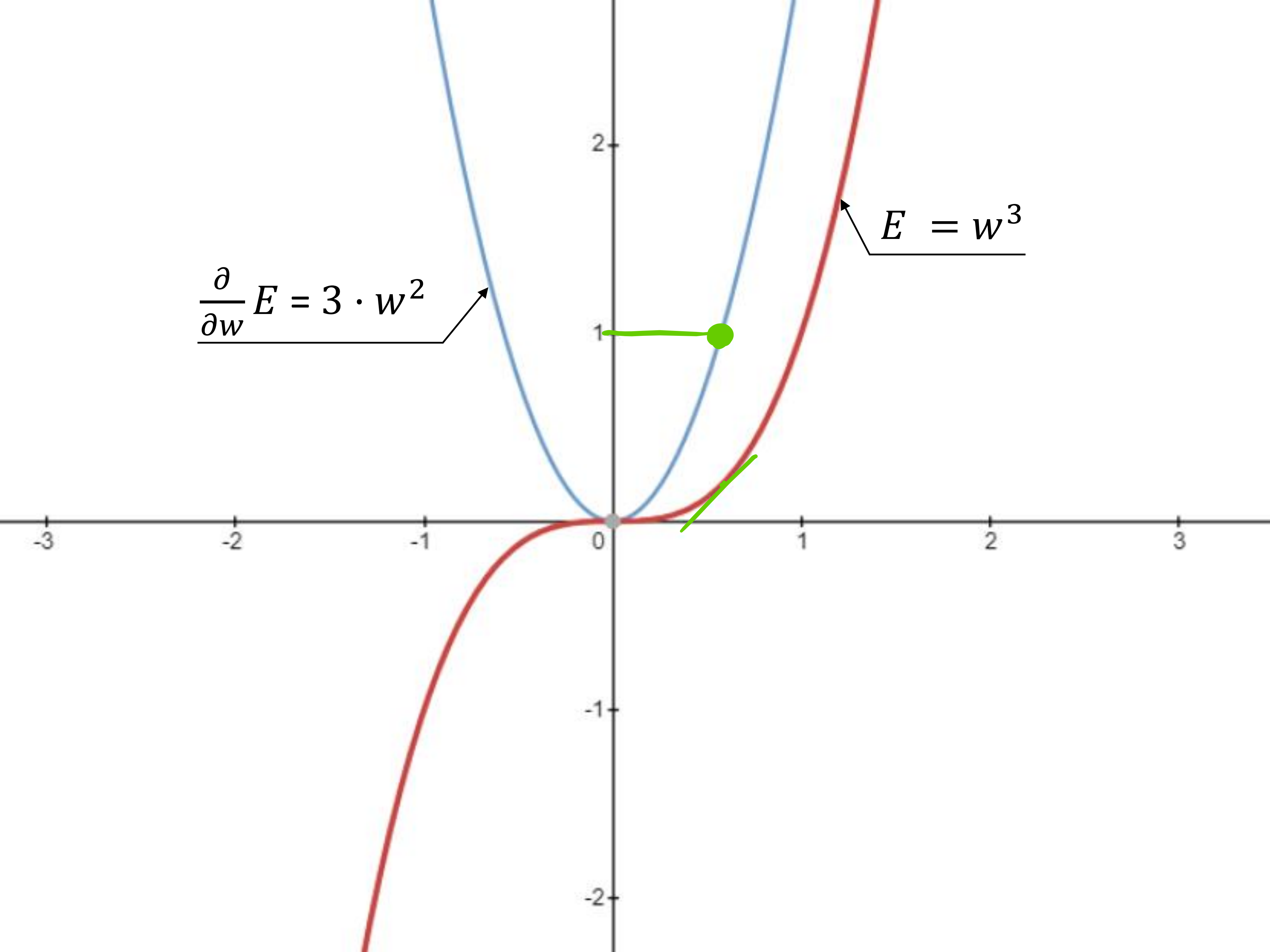
오류 함수가 다음과 같을 때
w = 2인 곳에서의 기울기는?

$$E = w^3$$

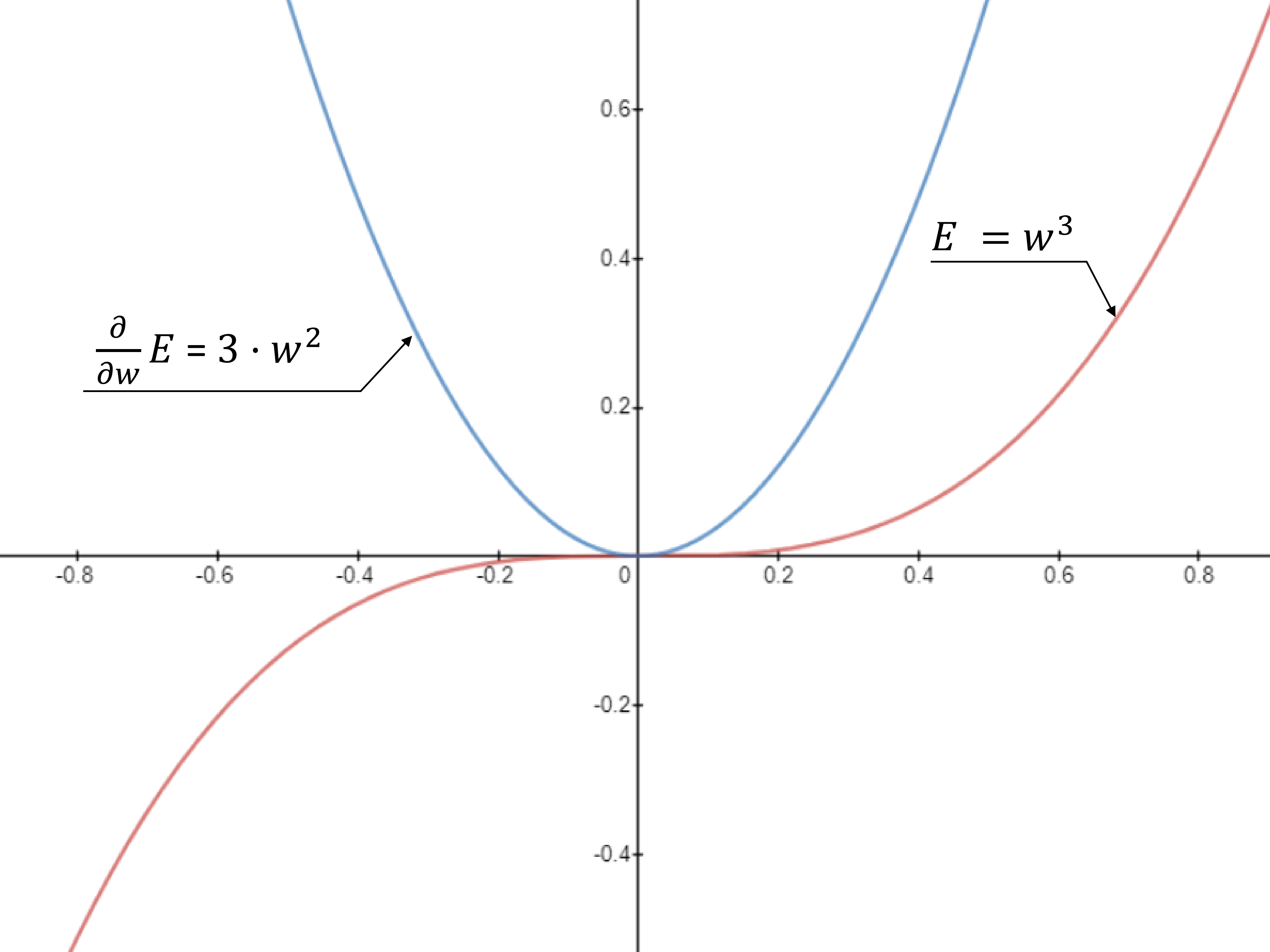
$$\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} E = 3 \cdot w^2$$

$$= 12$$

desmos



$\frac{\partial}{\partial w}E = 3 \cdot w^2$
$E = w^3$
-3
-2
-1
0
1
2
3
2
1
-1
-2



$\frac{\partial}{\partial w} E = 3 \cdot w^2$
$E = w^3$
0.6
0.4
0.2
-0.2
-0.4
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8

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

결국, 학습(Learning)이란?

- 오류 E가 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.