

AI and Deep Learning

선형회귀(1)

오류 그래프와 기울기, 그리고 경사하강

제주대학교

변영철

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

공부할 내용

- 회귀의 의미
- 뉴런의 출력과 절대값 오류
- 기울기와 경사하강
- 절대값 오류의 문제점과 평균
제곱오류
- 기울기를 구하는 방법, 미분
- 기울기가 갖는 의미

회귀(Regression)

인류는 고향을 떠나도 나이가
들면 언젠가는 고향으로
회귀하고(돌아가고) 싶어한다.
(인류학)

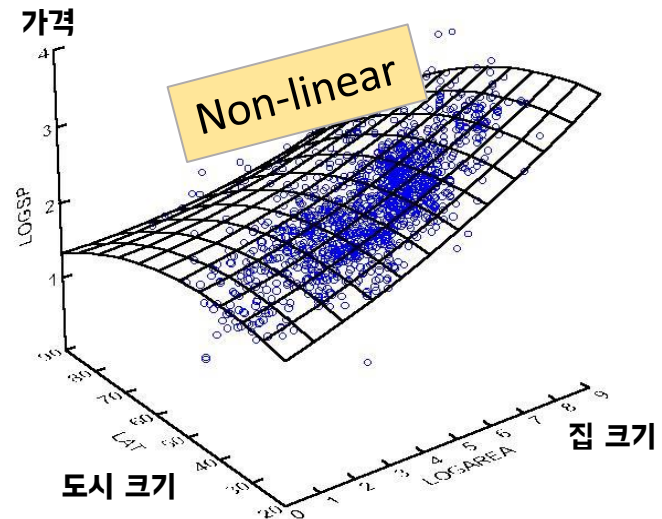
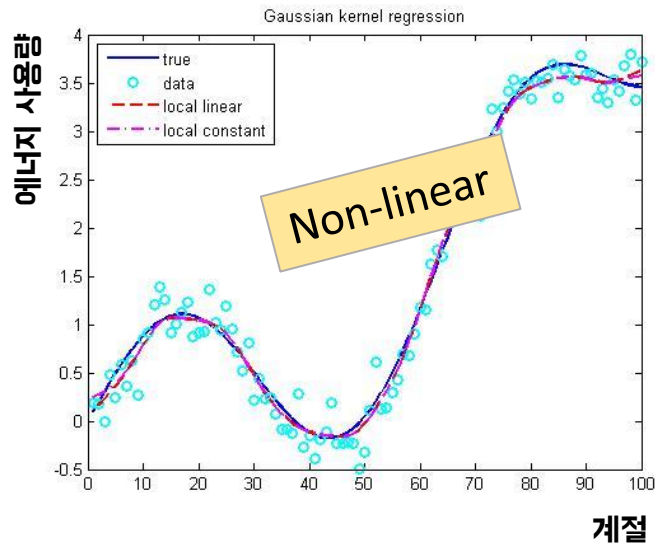
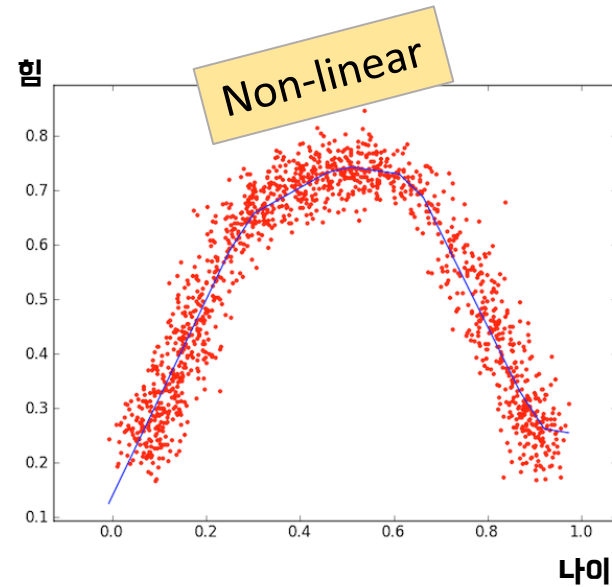
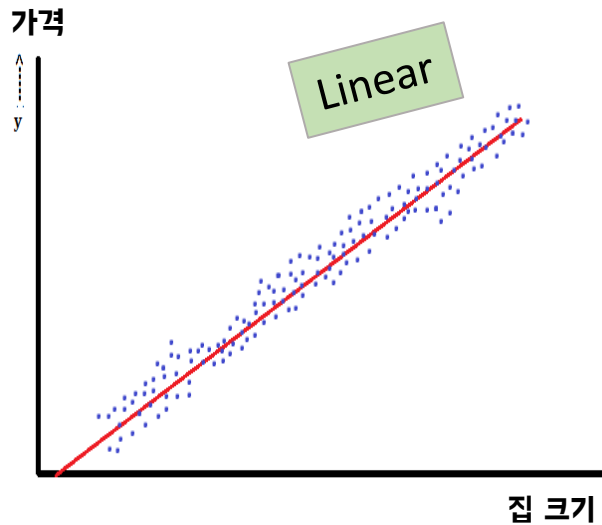


회귀=자연의 법칙, 현상

- 연어는 자라서 태어난 곳으로 돌아온다.
- 집은 클수록 비싼 편이다.
- 젊을 때는 힘이 세지만 나이가 들수록 약해진다.

일종의 규칙, 이러한 회귀는 지식이 되고 인해 예측(prediction)이 가능

회귀는 **그래프**로 표현하면 이해하기 쉽다.



Lab. Linear Regression

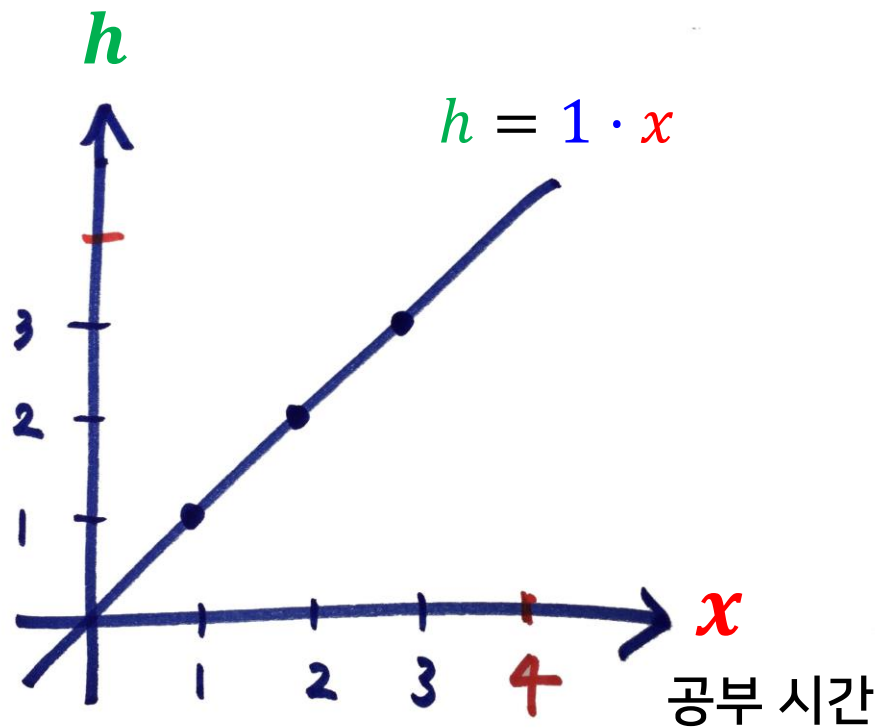
using  desmos

www.desmos.com

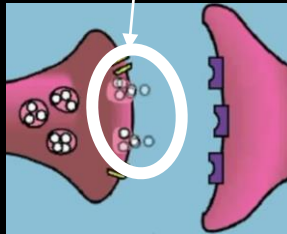
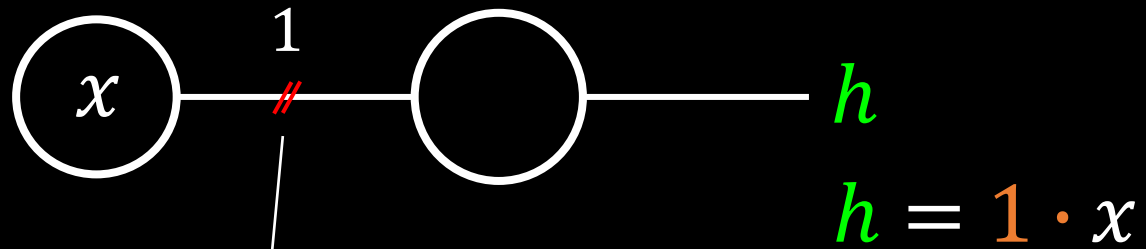
1. 점 $(1, 1)$ 표시
2. 점 $(2, 2), (3, 3)$ 추가
3. $h = x$
4. $h = 2x$
5. $h = wx$ (회전)
6. 모든 점 y 에 1 더함
7. $h = wx + 1$ (이동)
8. $h = wx + b$ (회전과 이동)

www.desmos.com

게임 시간

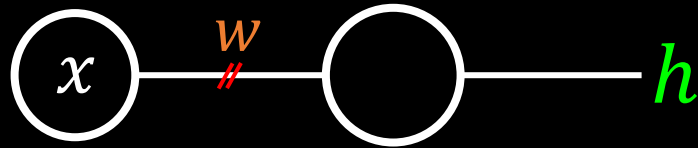


$$h = wx$$



- 신경전달 물질의 양
- 현재 값 = 1
- 많으면? 적으면? 없으면?

뉴런과 회귀



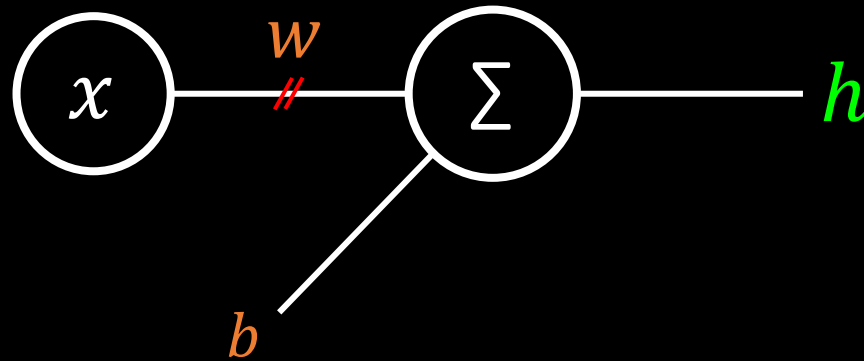
- 하나의 뉴런은 하나의 선형 회귀(Linear Regression)를 표현
- h ypothesis : 뉴런의 대답, a proposed explanation for a phenomenon (a regression).
- 증명되지 않았으나 w 조절을 통하여 회귀(regression)를 표현할 수 있음.



바이어스 b 의미

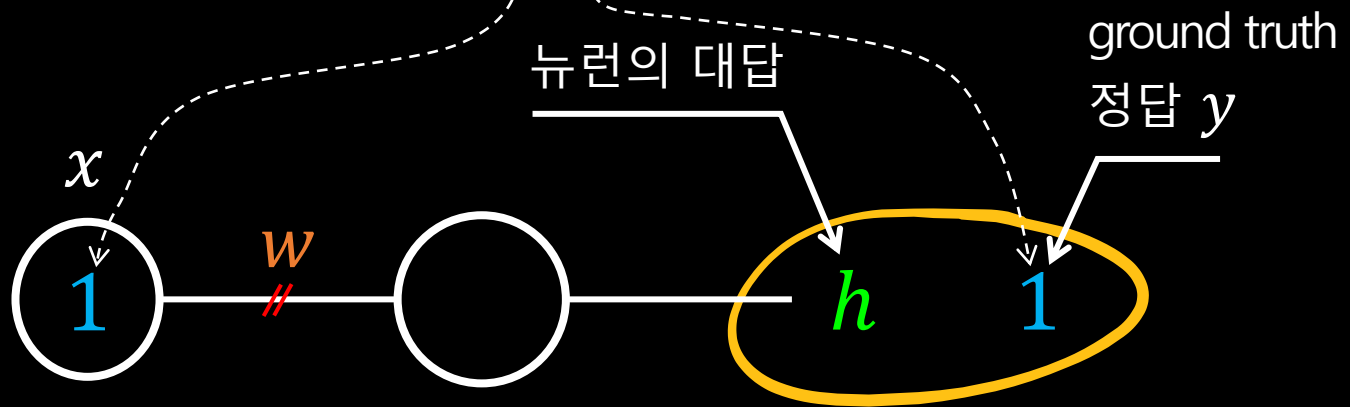
$$h = wx + b$$

바이어스로
다양한 회귀를
표현할 수 있음.



데이터와 학습

- 데이터 (x, y) 가 $(1, 1)$ 일 경우 w 는?



- 따라서 학습이란?

지도학습

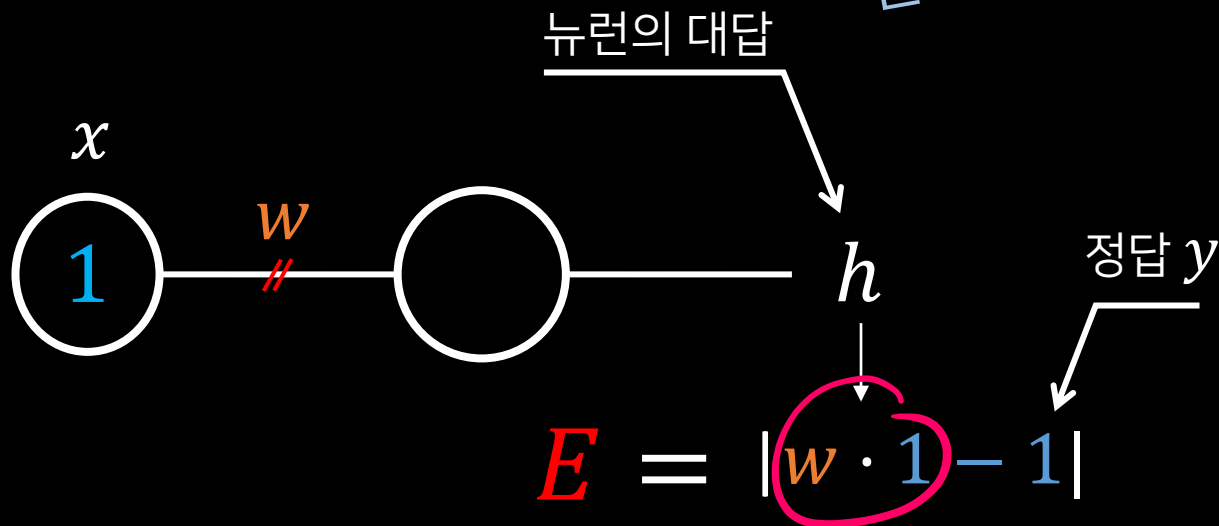
Supervised Learning

w 를 어떻게 조정할 것인가?

- 뉴런의 대답과 정답의 차이(오류, 에러 E , 비용, loss)를
- 차이가 많으면 야단침.
- 그러면 오류가 줄어들도록 가중치 w 가 자동으로 조정
- 정답에 가까울 때까지 w 를 조정

오류(차이) 함수

뉴런의 대답이 정답과
얼마나 차이가 있나?



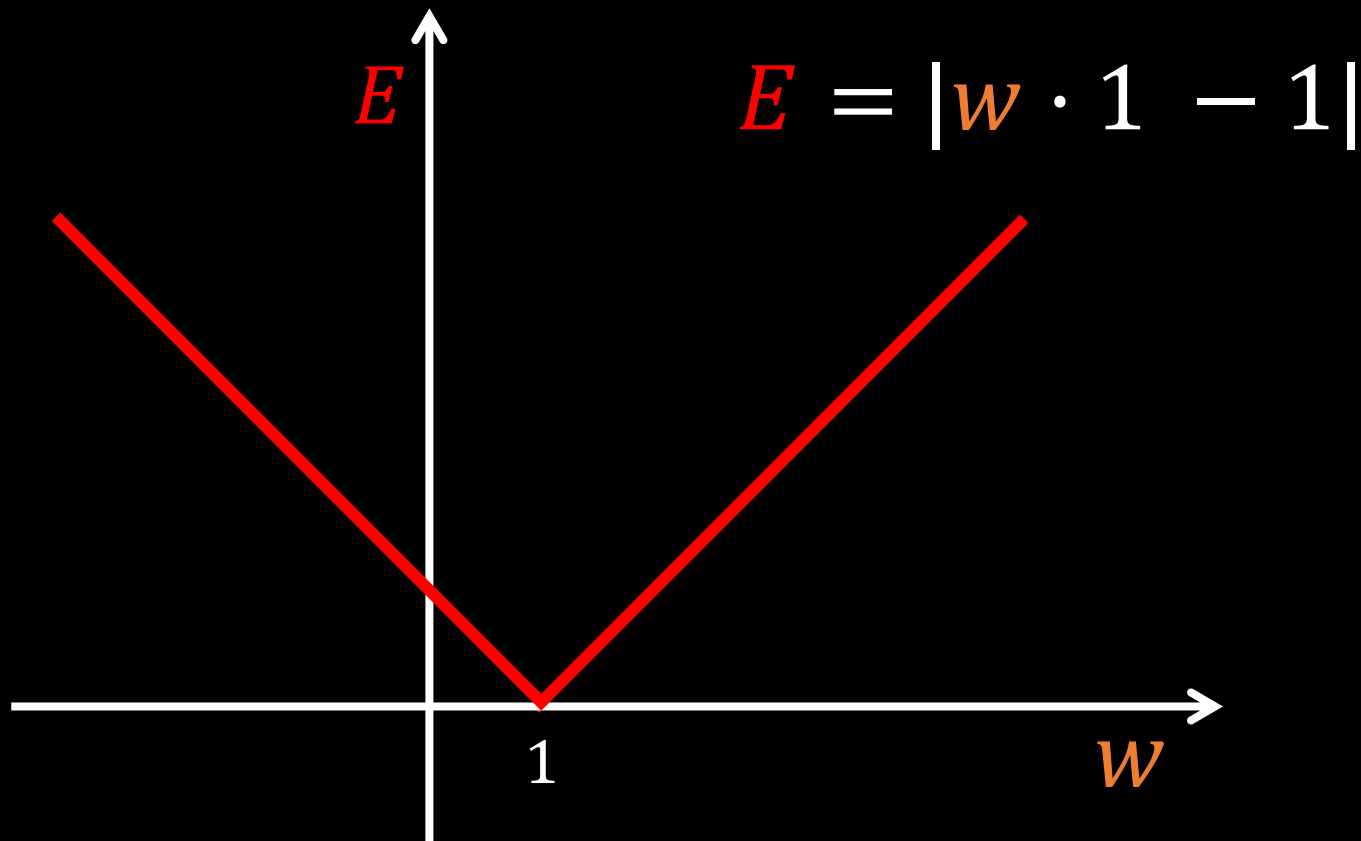
왜 절대값인가?

www.desmos.com

1. 점 $(1, 1)$ 표시
2. $h = w \cdot x$
3. $E = w \cdot 1 - 1$
4. $E = |w \cdot 1 - 1|$



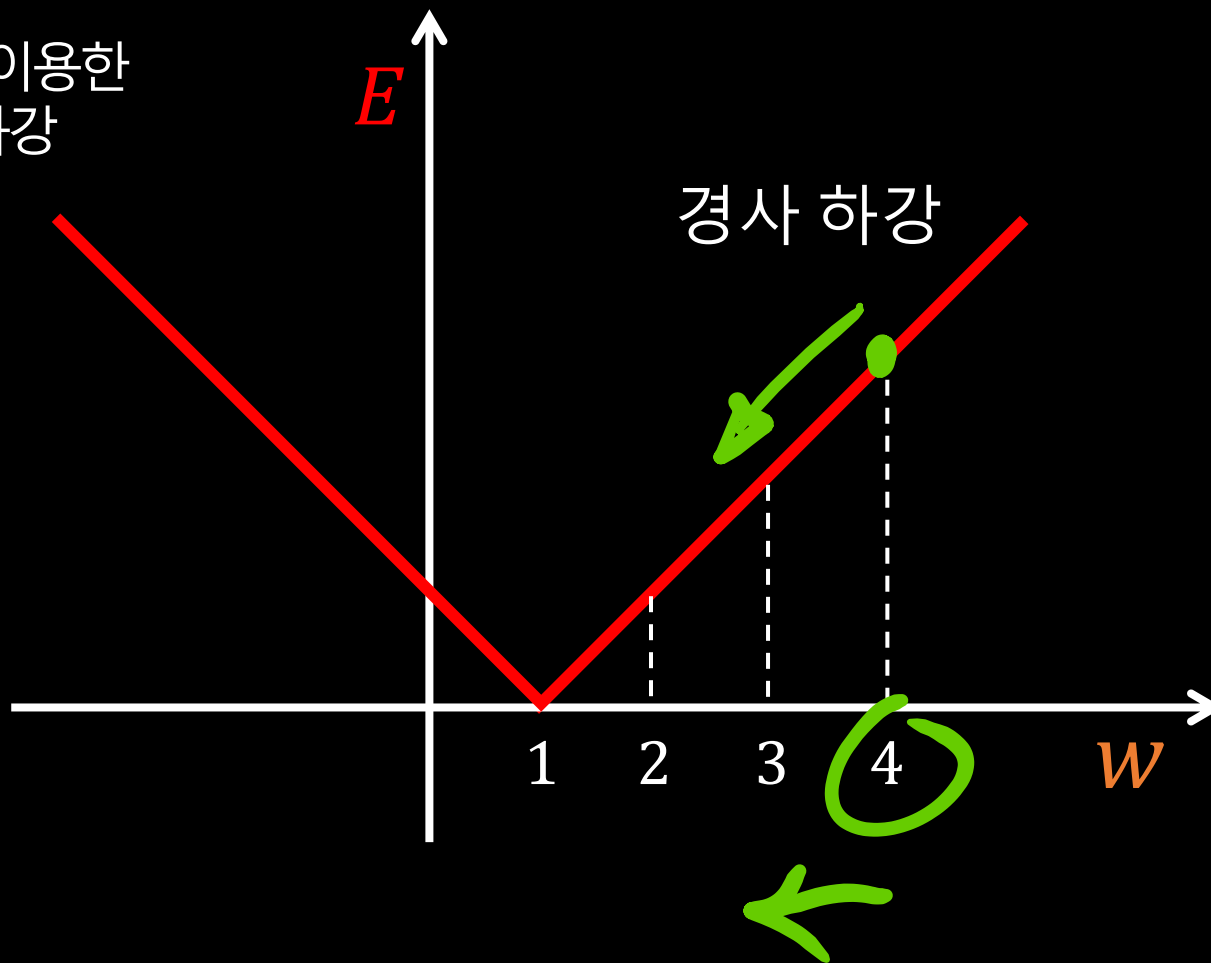
w 에 대한 오류 함수



오류가 줄어들도록

w 업데이트 하는 법

기울기를 이용한
경사하강



오류가 줄어들도록

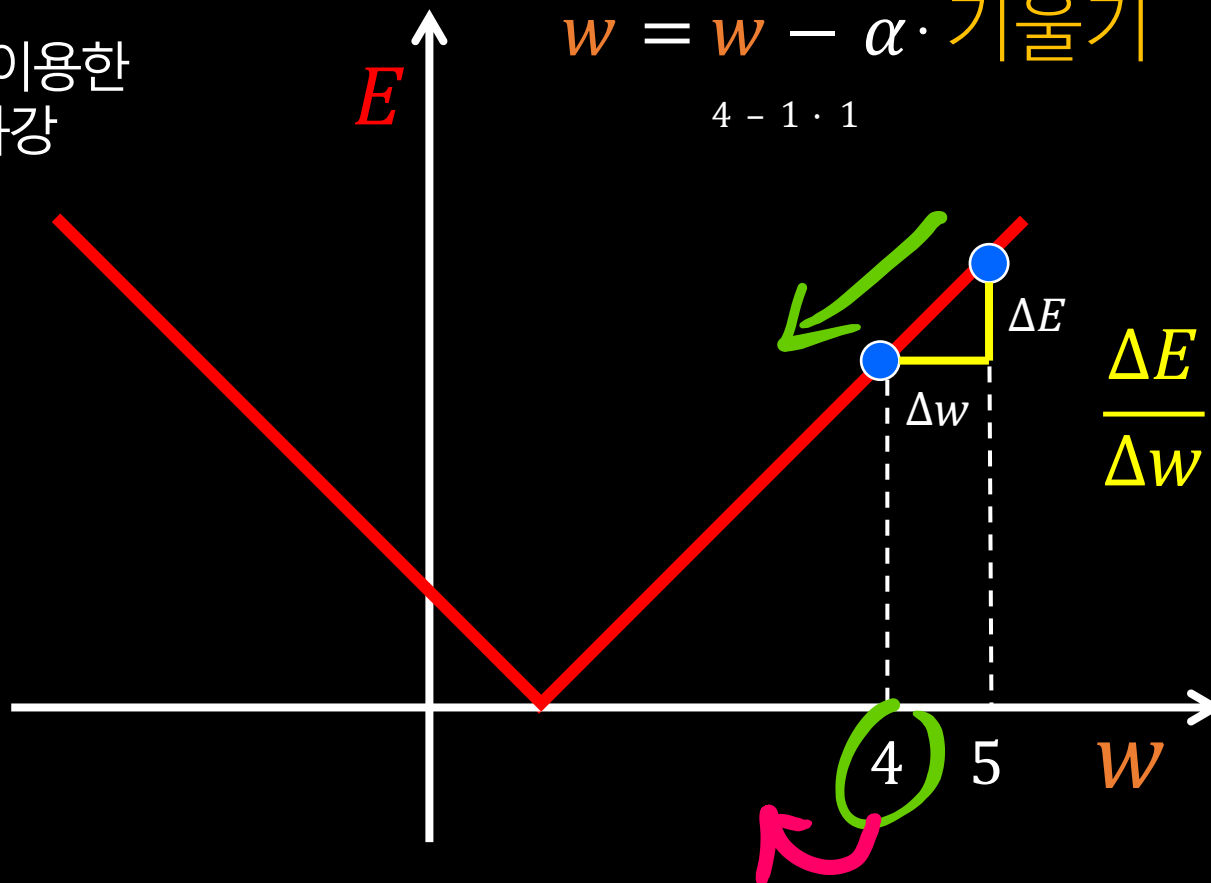
w 업데이트 하는 법

기울기를 이용한
경사하강

반영 비율 (가령, 1)

$$w = w - \alpha \cdot \text{기울기}$$

$4 - 1 \cdot 1$



오류가 줄어들도록

w 업데이트 하는 법

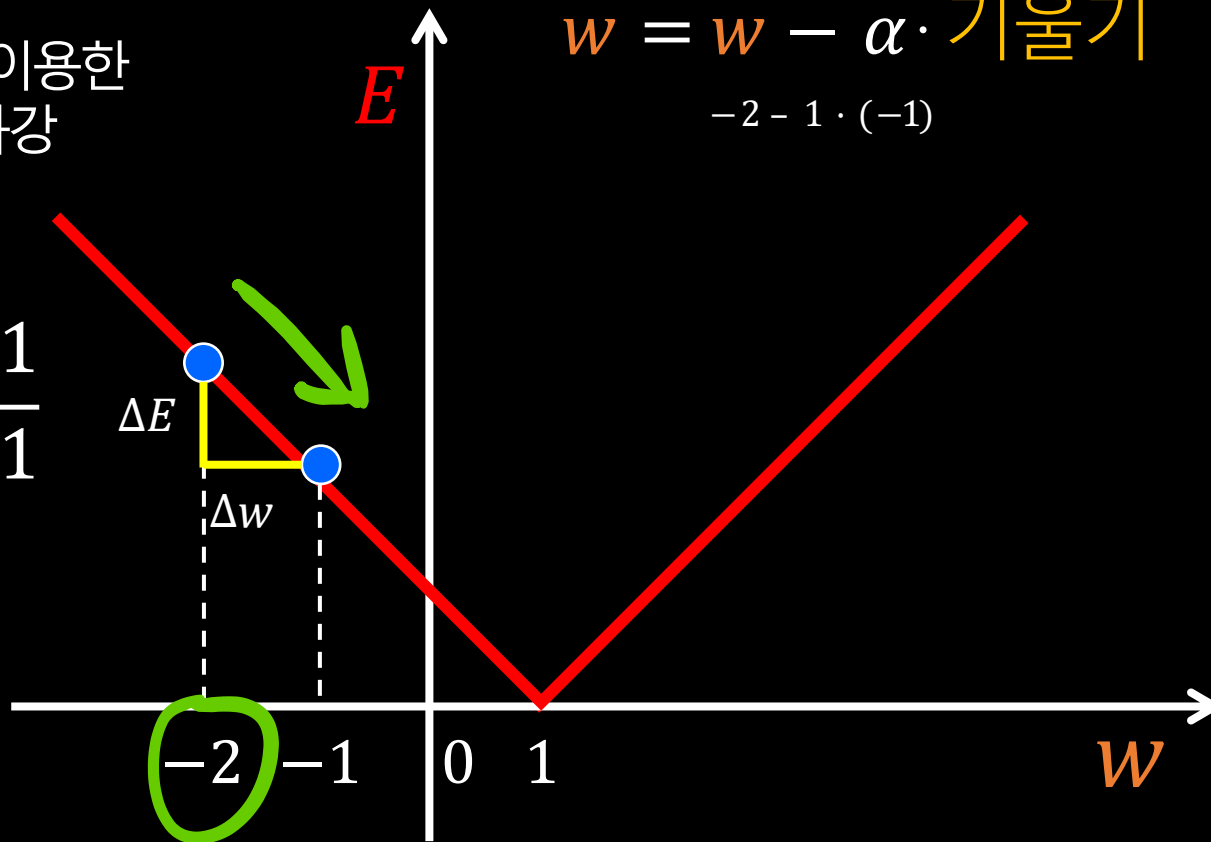
반영 비율 (가령, 1)

기울기를 이용한
경사하강

$$w = w - \alpha \cdot \text{기울기}$$

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

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



4

1

1

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

$$\begin{array}{l} 4 - 1 \cdot 1 \rightarrow 3 \\ 3 - 1 \cdot 1 \rightarrow 2 \\ 2 - 1 \cdot 1 \rightarrow 1 \end{array}$$

error E = 2

error E = 1

error E = 0

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

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

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

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

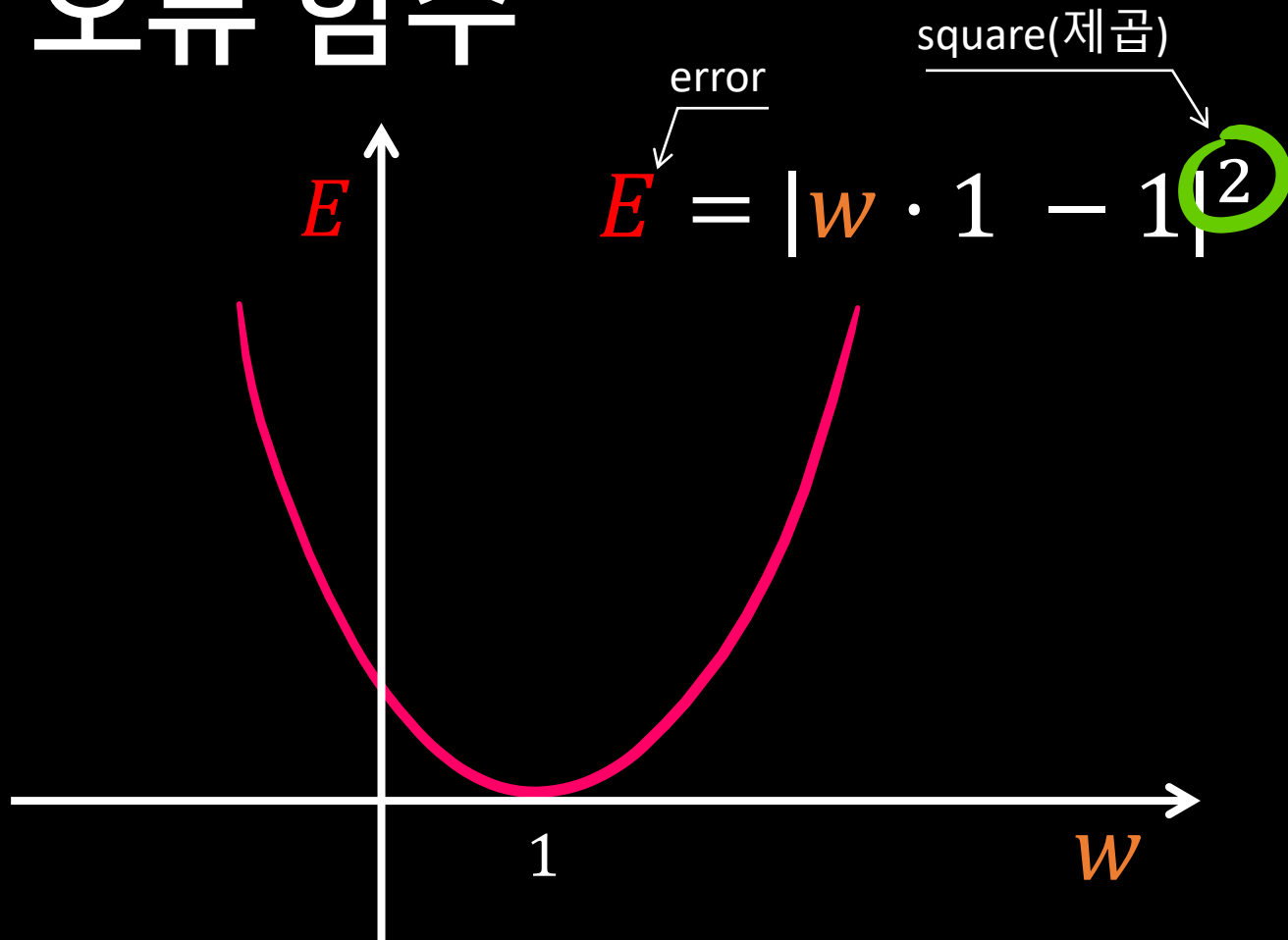
$$w - \alpha \cdot \text{기울기} \rightarrow w$$

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

- 현재 w 값에 관계 없이 항상 같은 기울기
- 따라서 항상 같은 경사하강 속도
- w 값이 수렴하기 어려울 수도 있음.
- ~~기울기 값만 보고 w 가 어디에 있는지 알 수 없음.~~
- ~~w 가 1일 때는 기울기 구할 수 없음.~~

제공 오류 함수



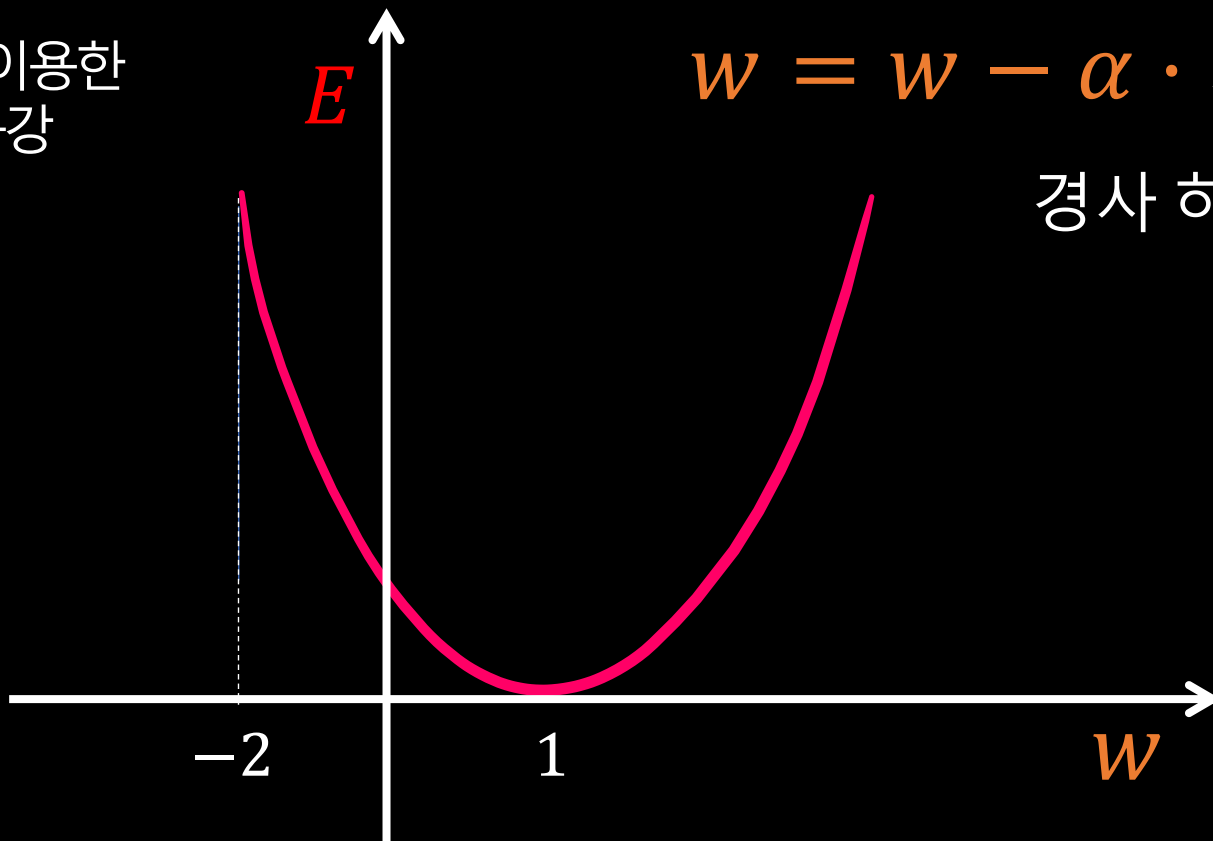
w 업데이트 하는 법

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

기울기를 이용한
경사하강

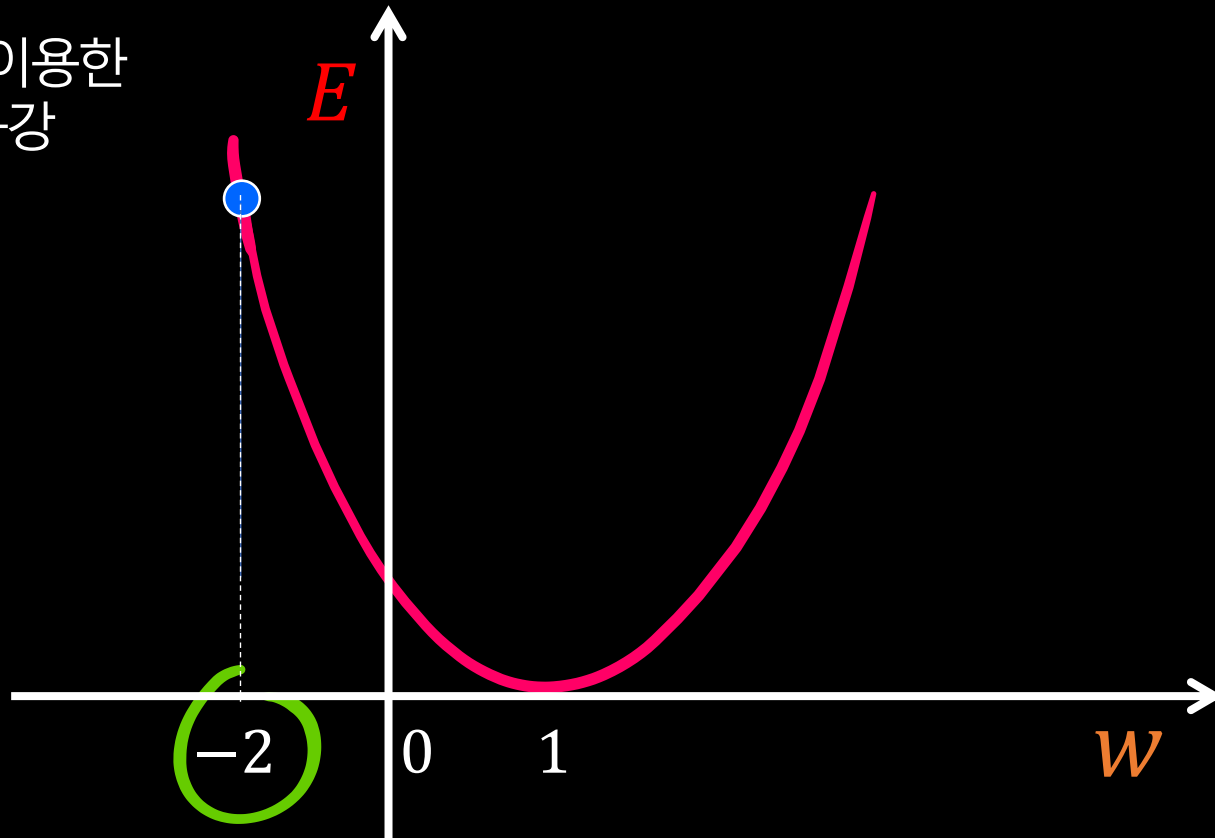
$$w = w - \alpha \cdot \text{기울기}$$

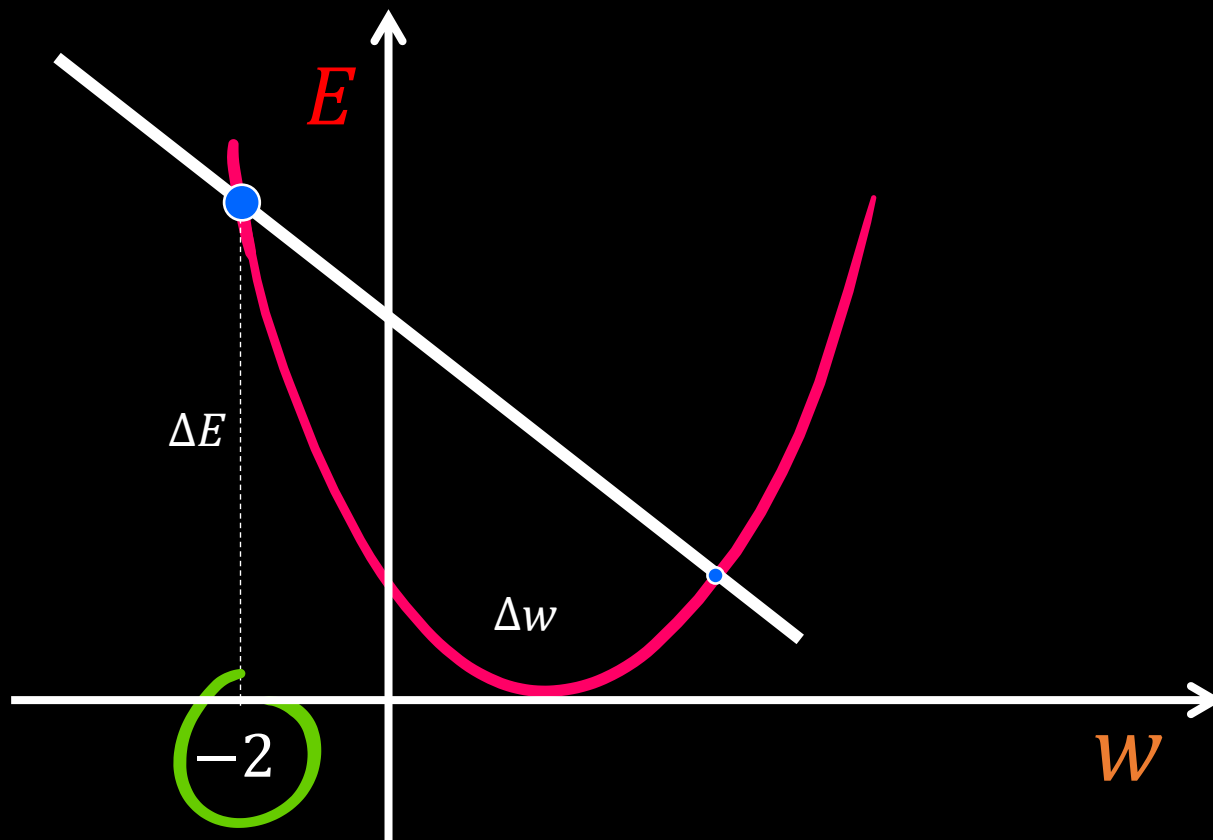
경사 하강

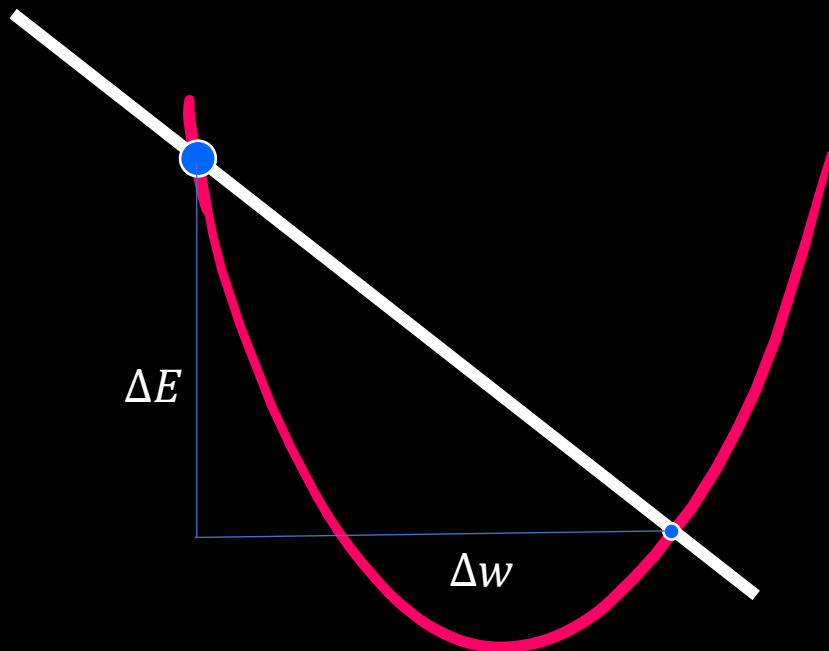


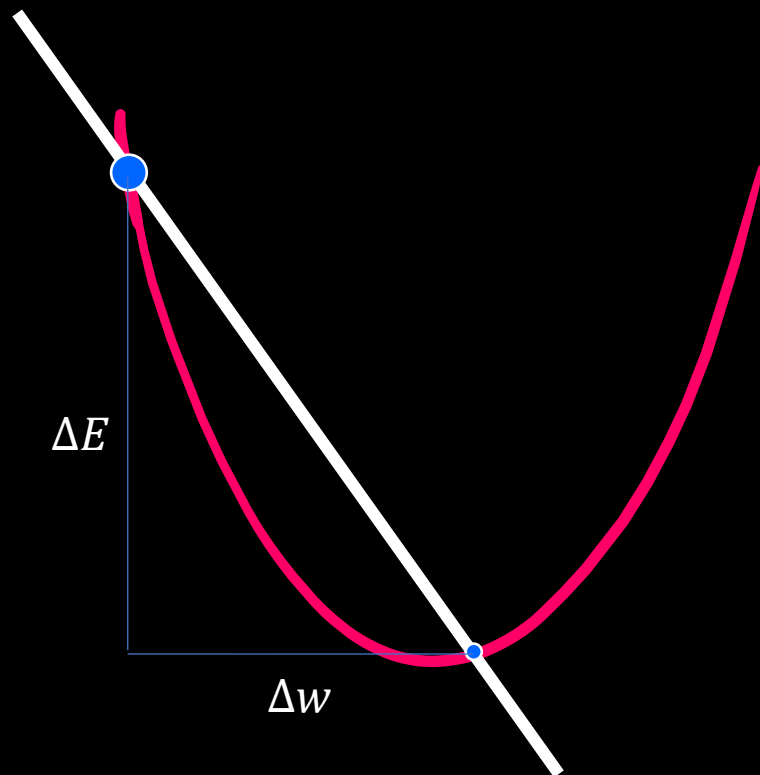
어떻게 기울기를 구할까

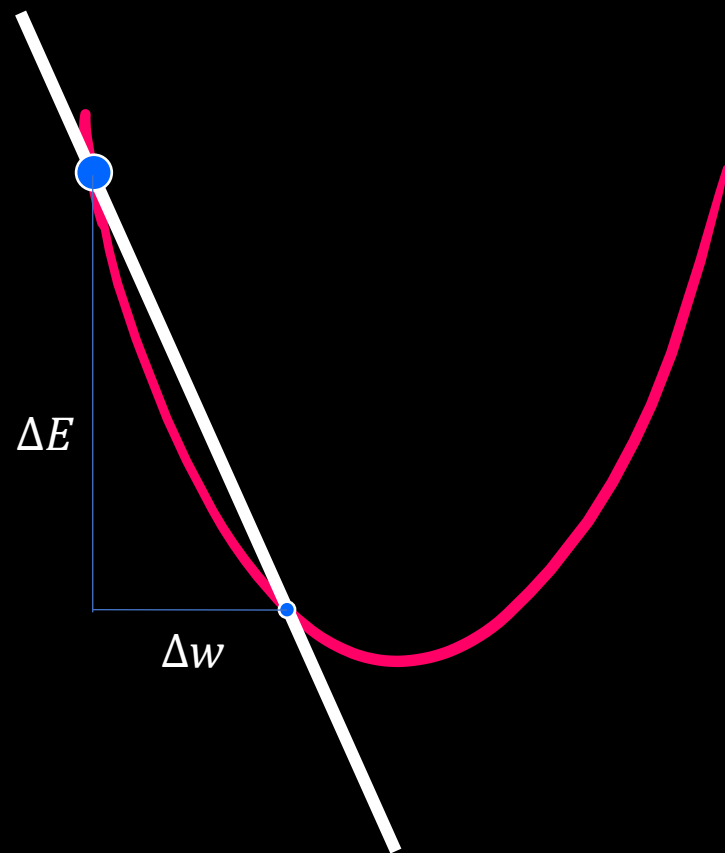
기울기를 이용한
경사하강

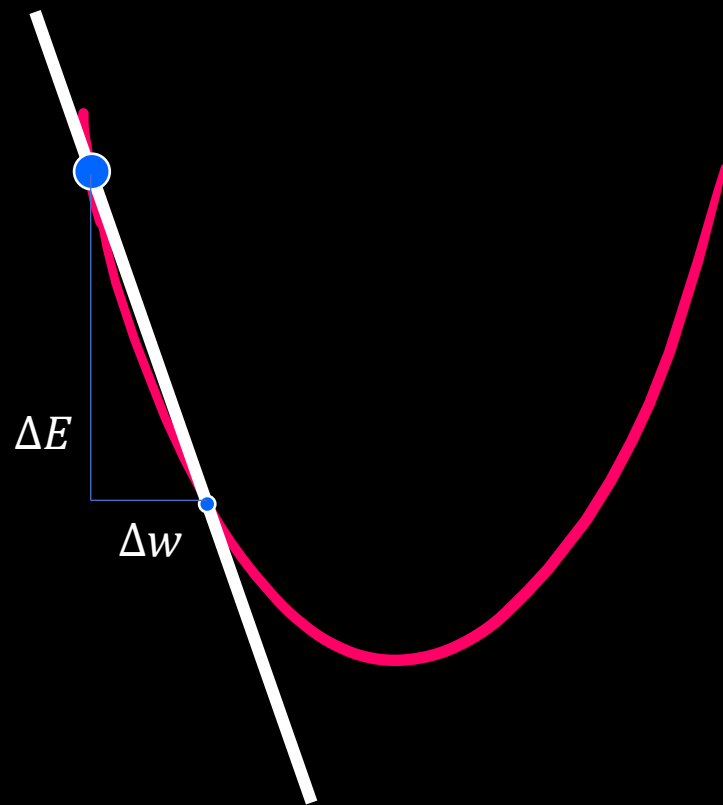


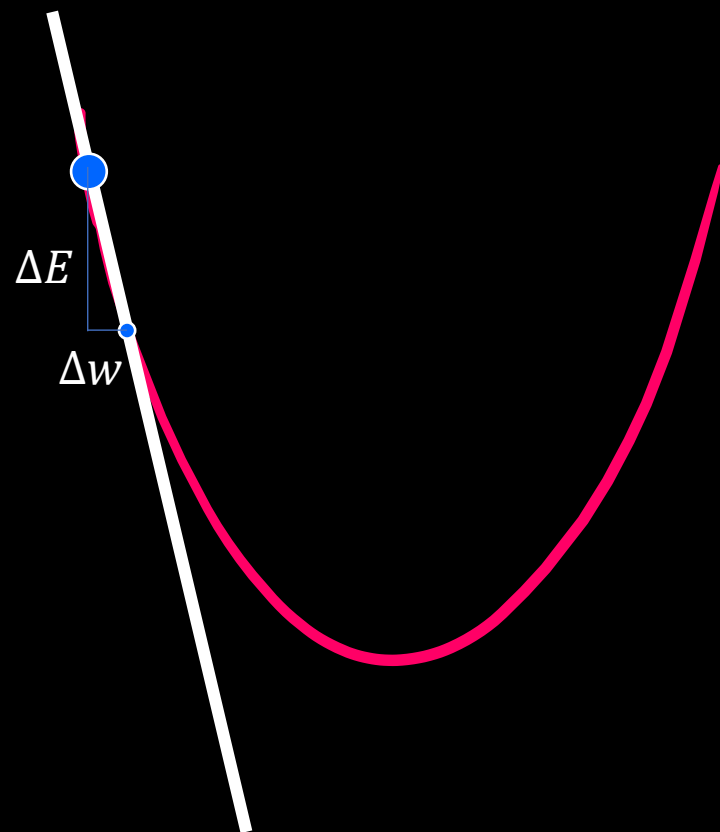




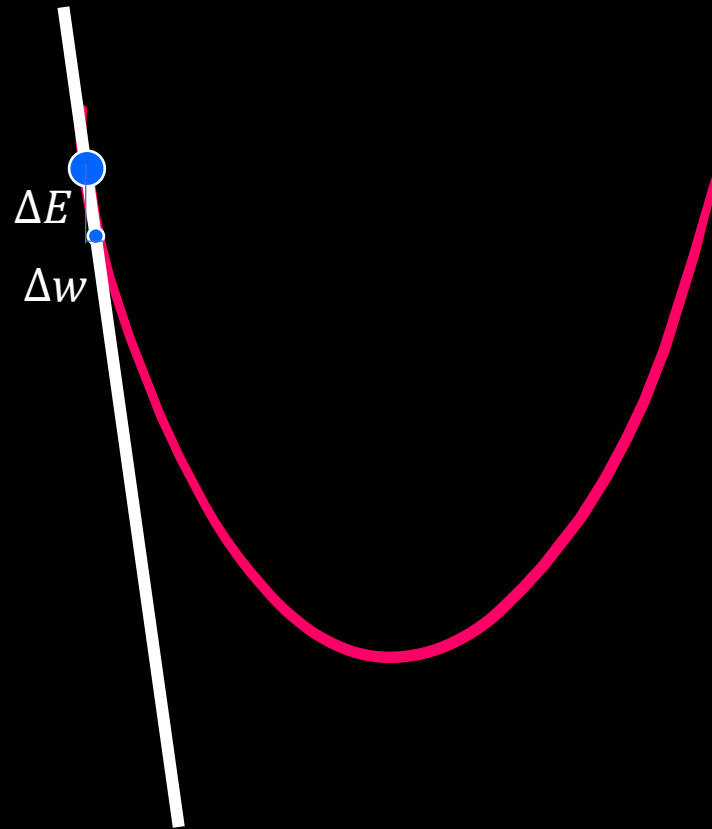


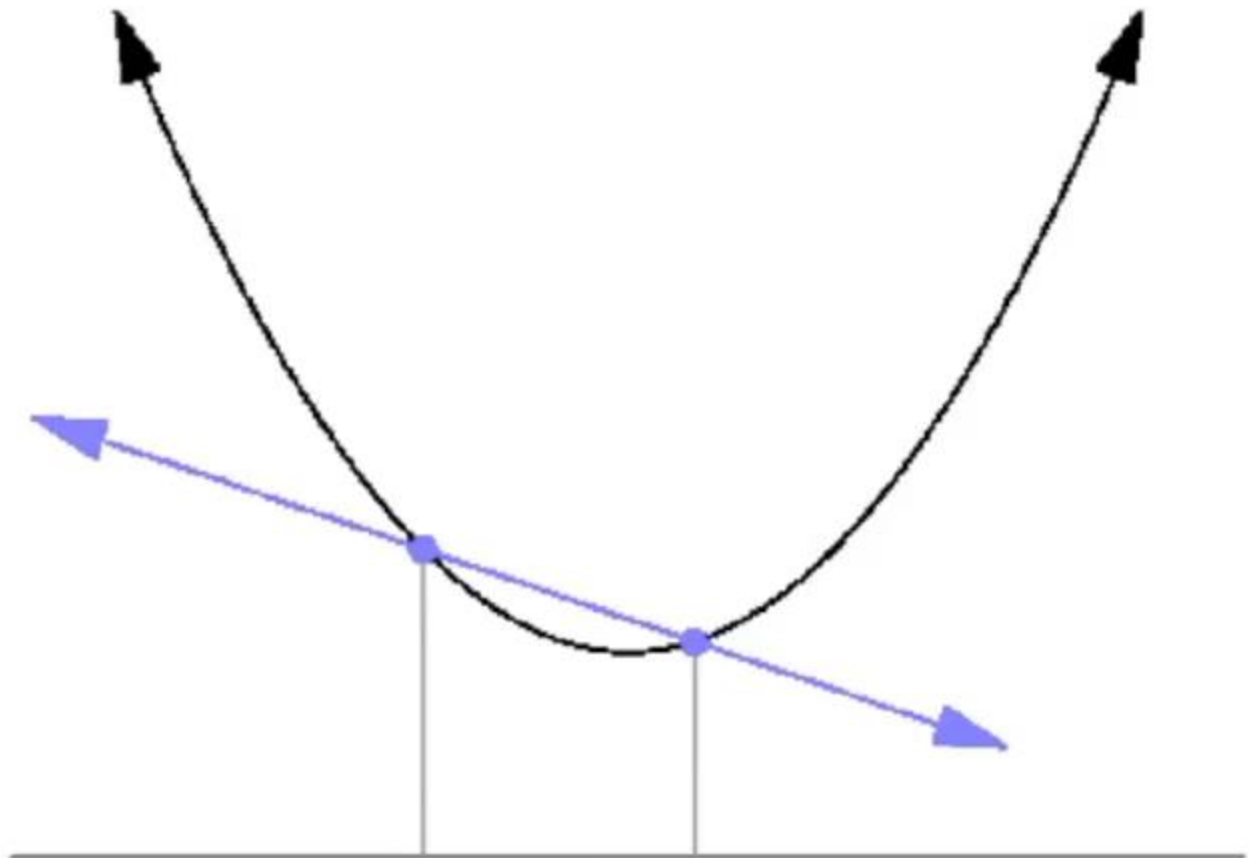






접선·Tangent line



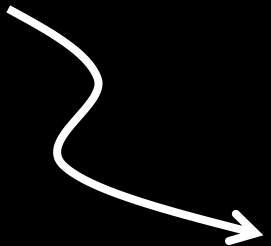


w

Δx

$$\Delta x = \Delta w$$

접선으로 정확한
기울기를 구하기 위함.


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

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

아주 잘게 자른 선, 미분

Numerical
differentiation

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



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

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

$$= \text{미분}$$

$$w = w - \alpha \cdot \text{기울기}$$

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

제공 에러 장점

- 오류 최저점, 즉, w 가 1인 지점에서 멀리 떨어진 곳에서는 급경사,
- 오류 최저점에 가까운 곳은 완만한 경사
- 기울기 값의 크기에 따라 w 가 어디에 있는지 알 수 있음.
- 모든 곳에서 기울기 계산 가능(미분가능)

절대값 오류에서는?

- 현재 w 값에 관계 없이 항상 같은 기울기
- 따라서 항상 같은 경사하강 속도
- 오류가 0, 혹은 거의 0이 되는 w 값으로 수렴하기 어려울 수도 있음.
- 기울기 값만 보고 w 가 최저점에서 얼마나 떨어져 있는지 알 수 **없음**.
- w 가 1일 때는 기울기 구할 수 **없음**.

여러 데이터

만일, 데이터가 3개라면

x_i	y_i
1	1
2	2
3	3

$$E = \frac{1}{3} \sum_{i=1}^3 (wx_i - y_i)^2$$

각각의 오류를 모두
더해서 평균



점 (2, 2), (3, 3) 추가

$$E = \frac{1}{3} \sum_{i=1}^3 (wx_i - y_i)^2$$

(w , E) 그리기

여러 데이터

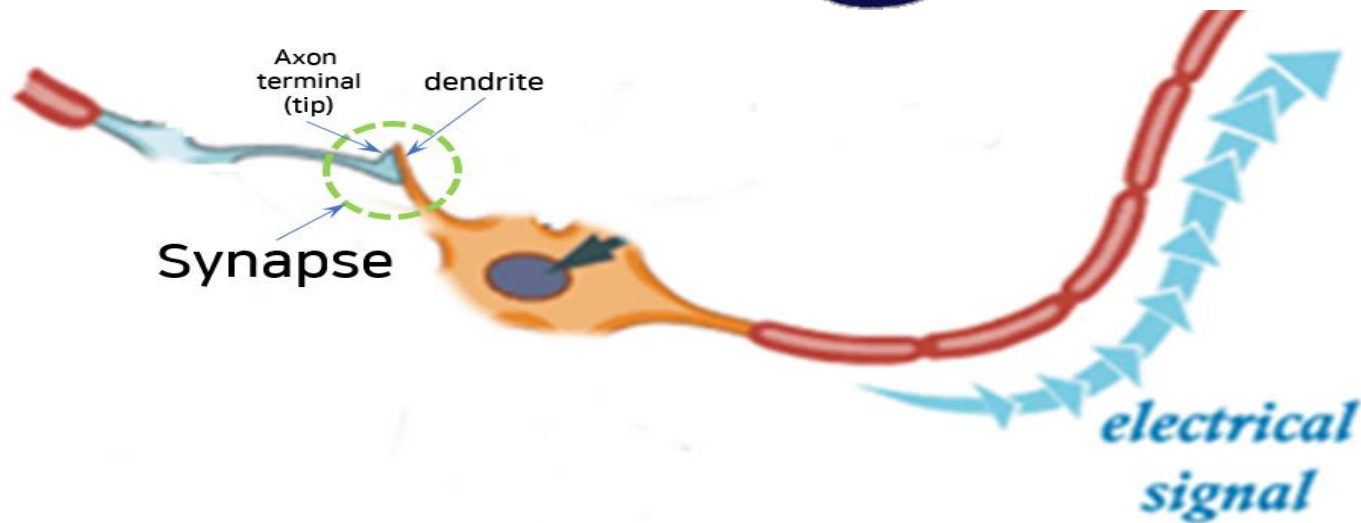
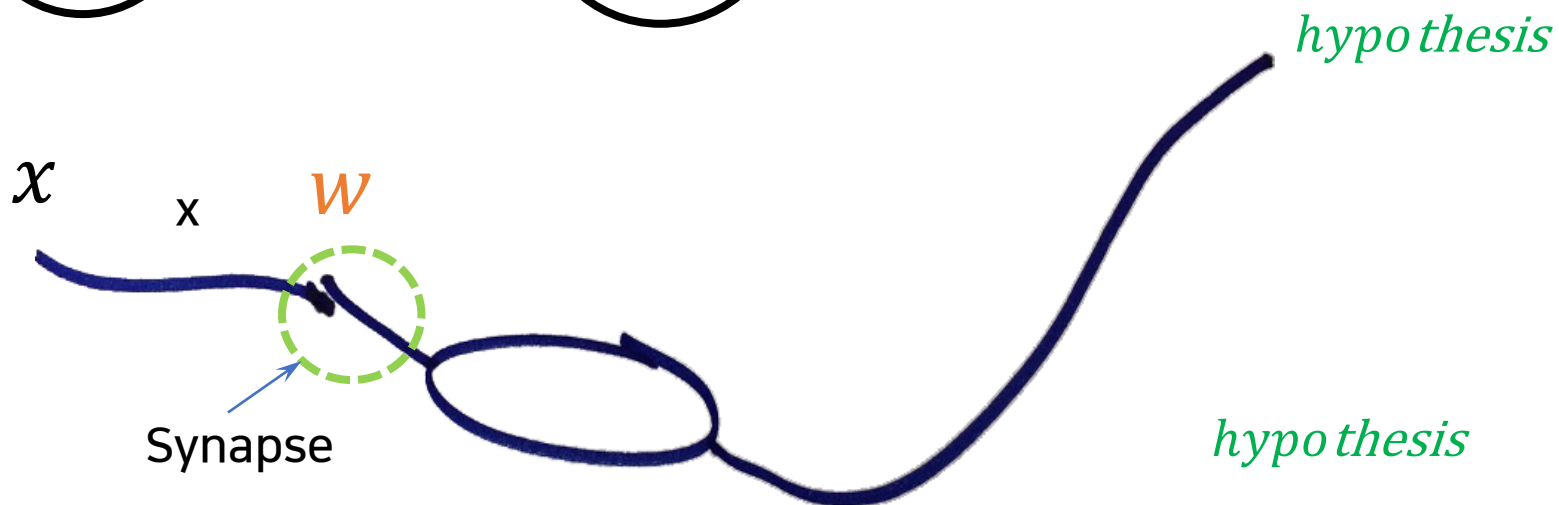
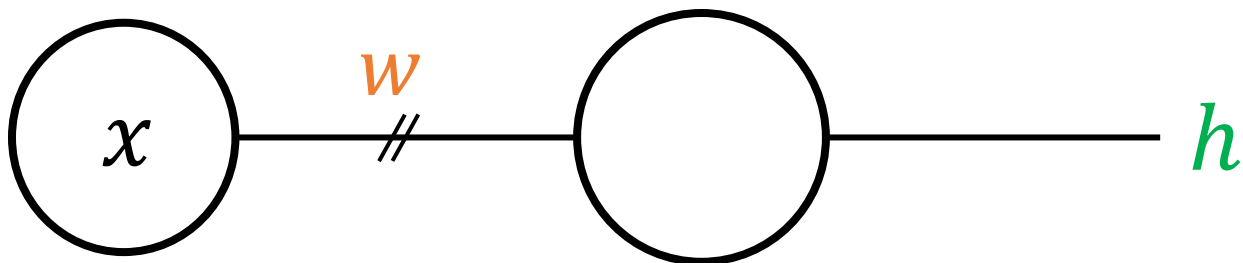
데이터가 m 개일 경우

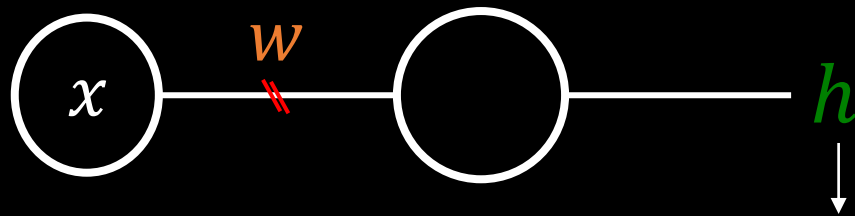
Mean Square Error
평균제곱에러

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

뉴런이 대답(예측)한 값 (가설)

정답



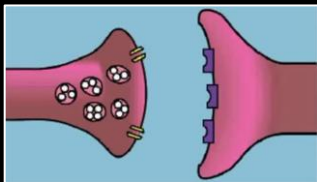


$$E = (h - y)^2$$

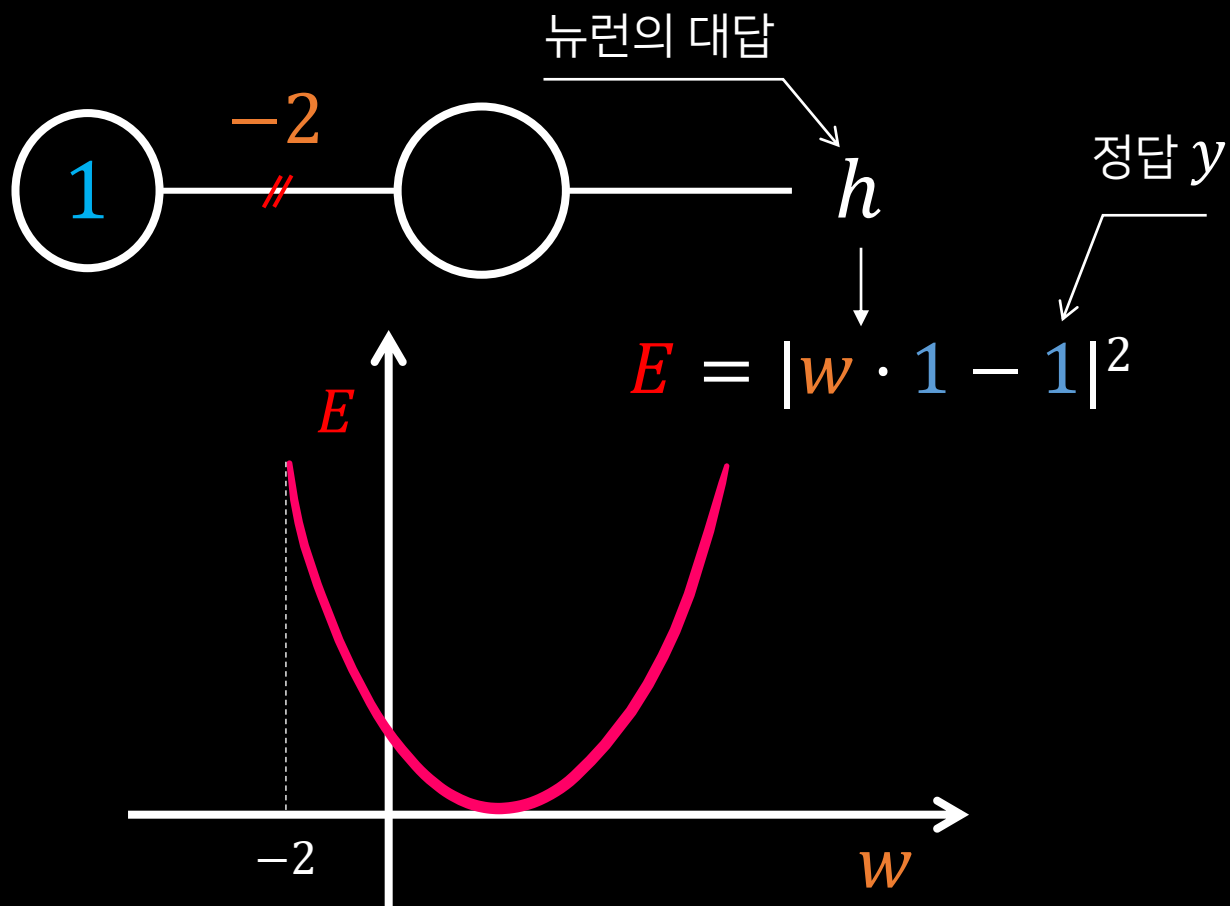
$$E = (w \cdot x - y)^2$$

x	y
1	1

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



$$\frac{\Delta noise}{\Delta dial}$$



기울기의 의미 (기미)

$$\frac{\Delta E}{\Delta w} \quad \frac{8}{1}$$

- 기울기가 크다. w 를 조금만 크게 해도 E 가 많이 늘어난다. $\rightarrow w$ 가 E 에 미치는 영향력이 크다.
- 기울기가 작다. w 를 바꿔도 E 는 별로 변하지 않는다. \rightarrow 미치는 영향력이 작다.
- 결국, 기울기 = 미치는 영향력 $\frac{1}{10}$

(Q) 미치는 영향 구하기

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

데이터 (x, y) 가 $(1, 1)$ 일 때

$w = 3$ 인 지점에서 w 변화가 오류 E 에 미치는 영향(기울기)를 구하라.

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

(A1) 값을 대입

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

w : 3 -> E : 4

w : 3.00001 -> E : 4.00004

w 가 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 기울기 = 미치는 영향 = 4

(A2) 미분

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

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

따라서 $w = 3$ 이면
기울기는 $2(3 - 1) = 4$

학습(Learning)은

- $w = w - a \cdot$ 기울기(미치는 영향)
- 경사하강
- 오류 E 가 감소하도록 w 업데이트
- 파라미터 튜닝

이번 학습에서는

- 회귀가 무슨 의미인지 알 수 있다.
- 뉴런의 대답과 정답 간의 차이를 그래프로 그릴 수 있다.
- 경사하강 방법을 알 수 있다.
- 오류 그래프가 갖는 문제점을 파악할 수 있다.
- 기울기 의미를 이해할 수 있다.