AI and Deep Learning

Linear Regression & Back-propagation (1)

Jeju National University

Yung-Cheol Byun

여행, 그리고 회귀

회귀(Regression)

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

자연의 법칙, 섭리, 일종의 규칙

- 연어는 태어난 곳으로 돌아온다.
- 집은 클수록 비싸다.
- 젊을 때는 강하지만 나이가 들수록 약하다.
- 남자가 여자보다는 큰 편이다.
- 성적이 좋을 수록 취업이 잘된다.

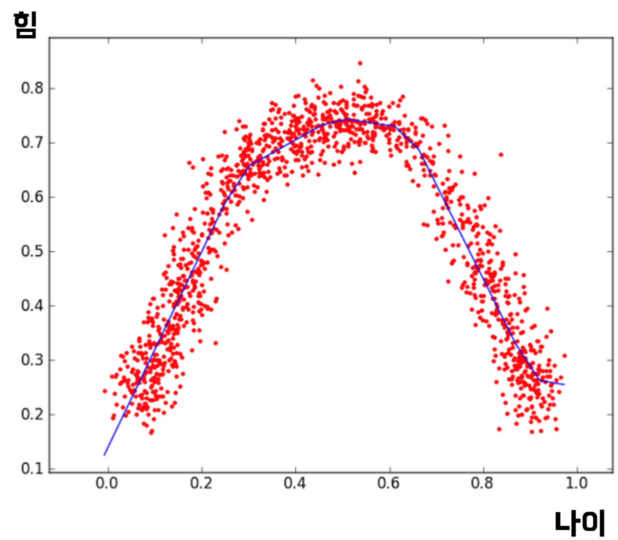
반드시 그런 것은 아니지만 일반적으로 그런 경향이 있다.
이런 '일종의 규칙' 때문에 '예측'을 할 수 있다.

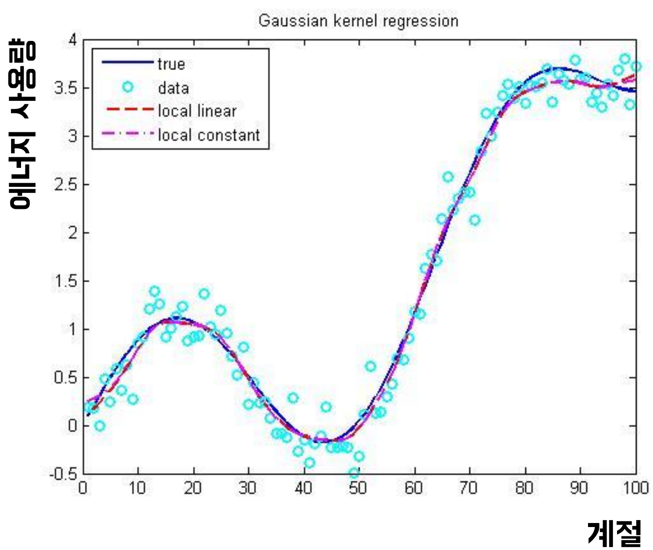
이를 잘 표현하는 말, 용어

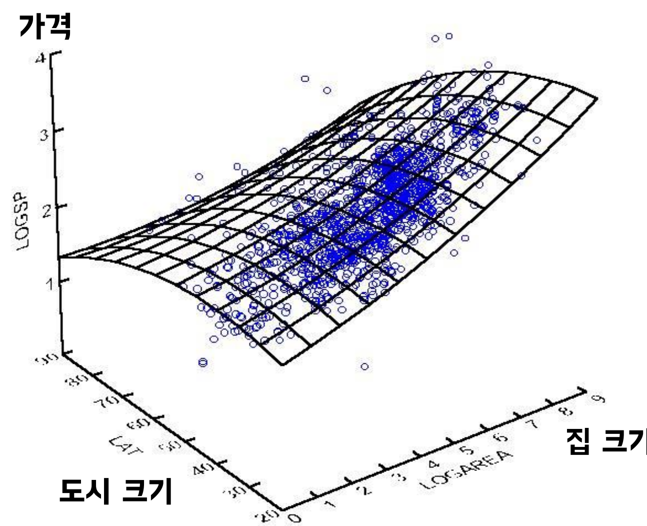
회귀(Regression)

회귀는 그래프로 표현하면 이해하기 쉬움.
(집 크기와 가격의 관계)


가격
y
집 크기


힘
나이

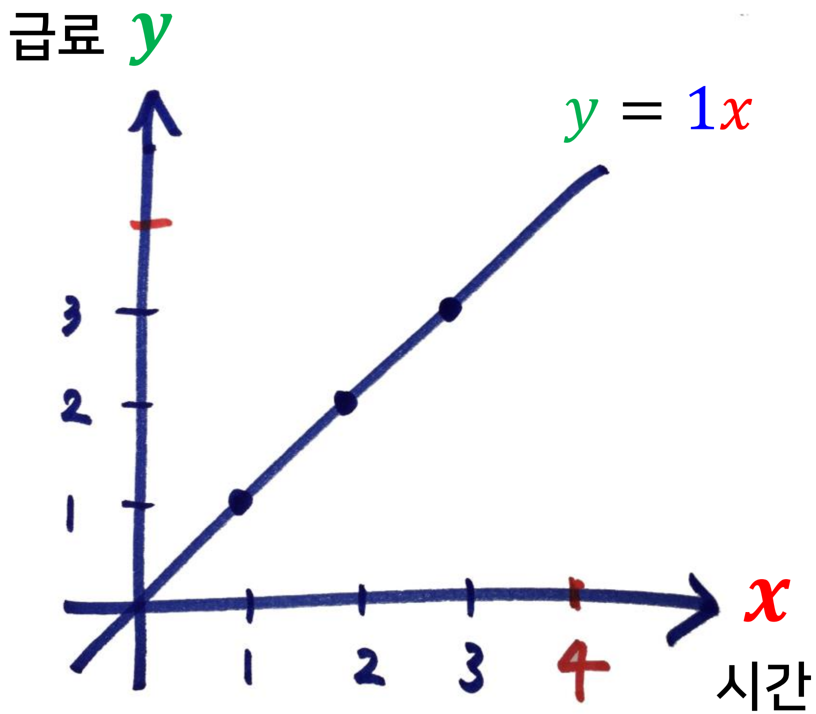

Gaussian kernel regression
true
data
local linear
local constant
에너지 사용량
계절


가격
LOGSP
LAT
LOGAREA
집 크기
도시 크기

Linear Regression

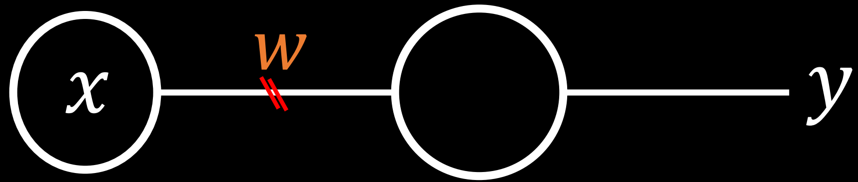
분포 형태가 직선(Line) 모양
일한 시간과 임금의 관계
집의 크기와 가격의 관계

www.desmos.com



급료 y
$y = 1x$
1
2
3
1
2
3
4
x
시간

뉴런의 능력



- $y = wx$
- 하나의 뉴런은 하나의 Linear Regression을 표현

가설(Hypothesis)

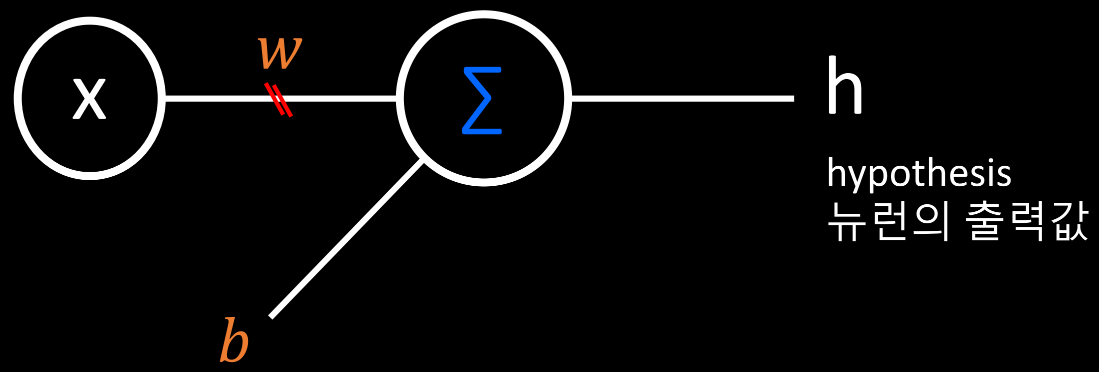
$$h = wx$$

$$h = wx + b$$

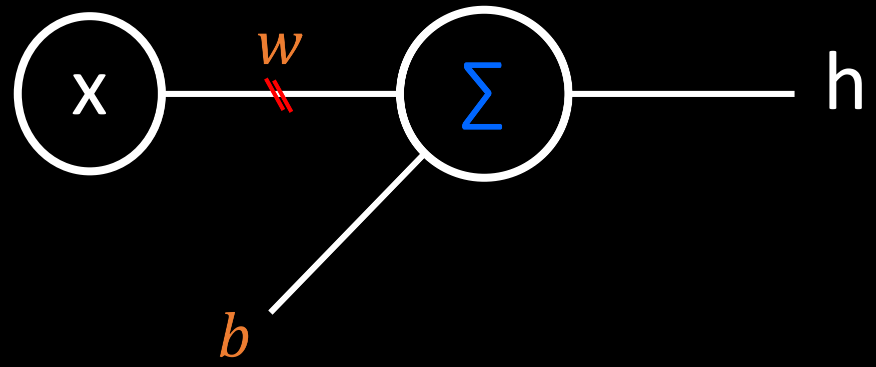
b가 있으면 뉴런은 더 다양한 분포(회귀)를 표현할 수 있음.

- 뉴런의 출력을 공식화
- 아직 증명되지 않았으나 조절을 통하여 데이터를 표현할 수 있는 것 -> 가설(h)
- 뉴런의 시냅스 w 값을 적절히 조절할 경우 Linear Regression 데이터를 잘 표현할 수 있음.

가중치 w와 바이어스 b 역할



(Q) 뉴런과 회귀



어떻게 w를 조정할 것인가?

- 모든 점을 지날 경우 차이(오류, 에러, 비용, loss)는 0
- 차이가 0이 되도록 w 와 b 를 조정하자.

오류 함수

오류(E) = |뉴런이 예상한 값 – 정답|

$$\underbrace{\text{뉴런이 예상한 값}}_{\text{가설(hypothesis)}}$$

$$E = |wx - y|$$

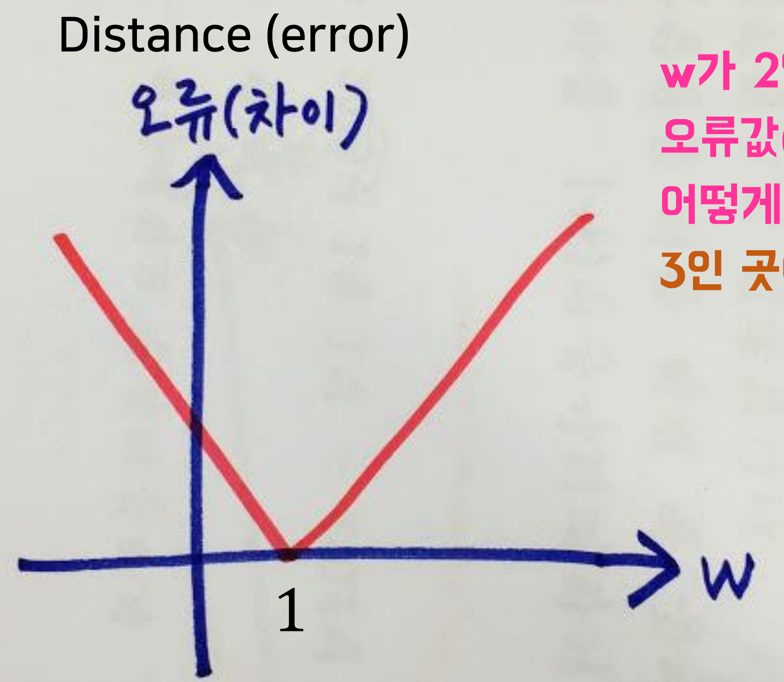
오류 함수

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

x	y
1	1

오류 함수 그래프

$$E = \frac{1}{3}\sum_{i=1}^{3} |wx_i - y_i|$$



w가 2인 곳에서
오류값(E)은
어떻게 달라질까?
3인 곳에서는?

오류 함수

만일, 데이터가 3개라면

x	y
1	1
2	2
3	3

$$E = \frac{1}{3}\sum_{i=1}^{3}|wx_i - y_i|$$

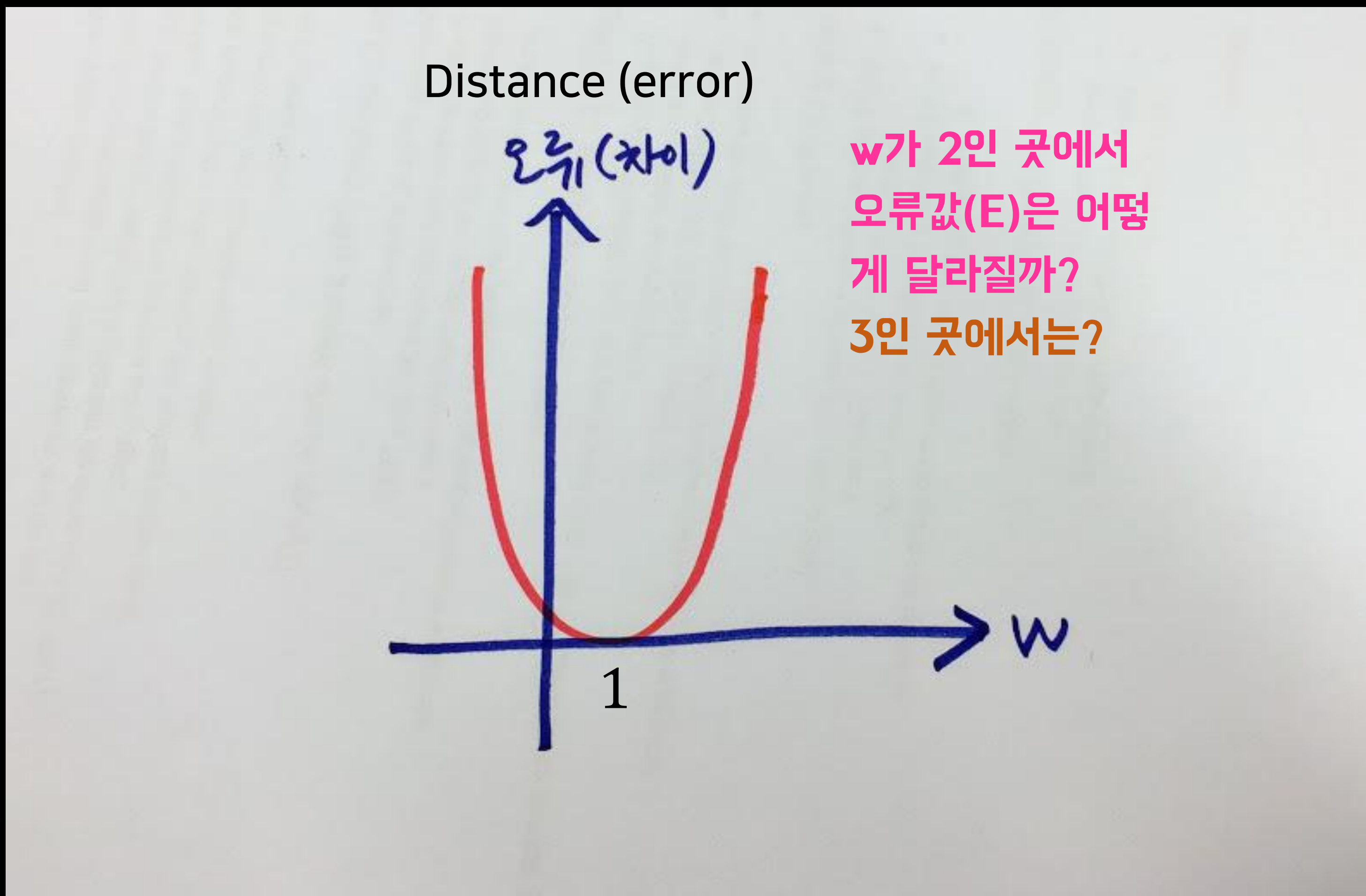
모두 더해서 평균

오류 함수 그래프

오류 함수 그래프

Mean Square Error

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



오류 함수

데이터가 m개일 경우

Mean
Square
Error

뉴런이 예측한 값 (가설)

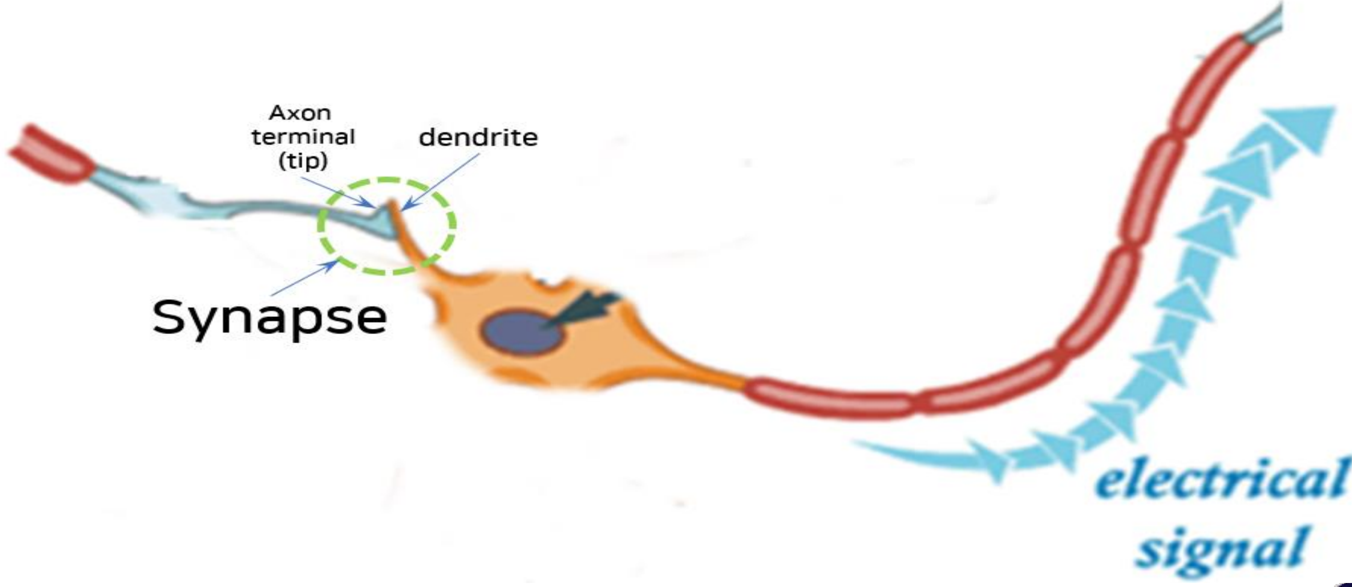
정답

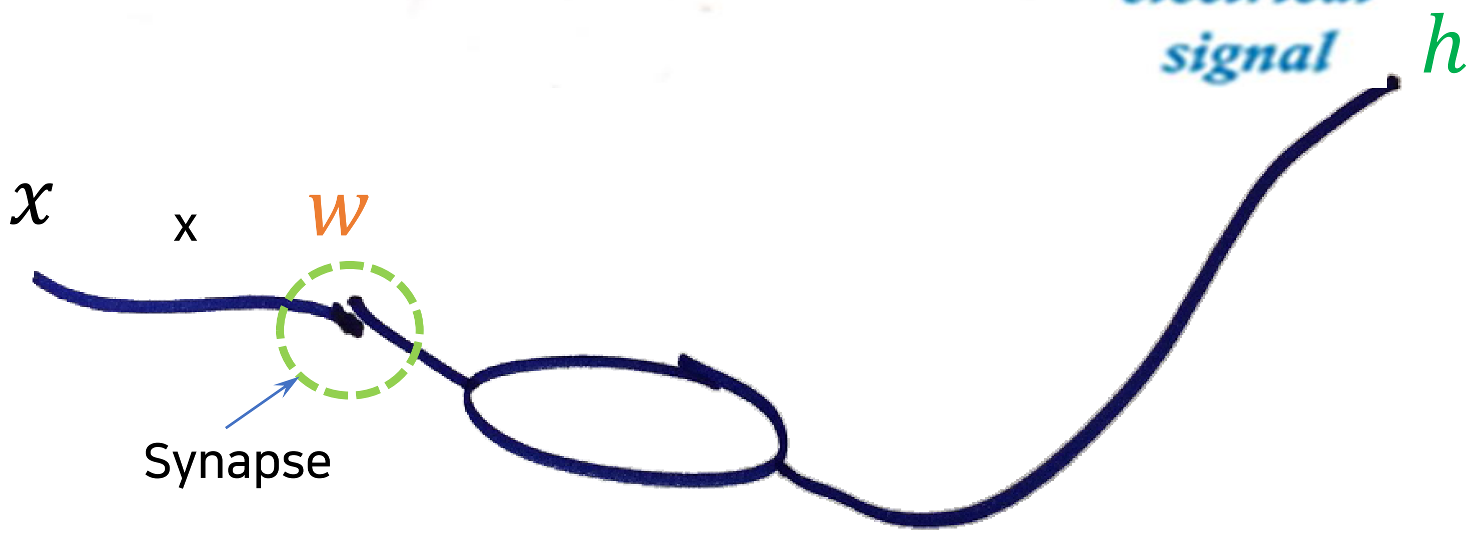
$$E = \frac{1}{m}\sum_{i=1}^{m}(wx_i - y_i)^2$$

(Q) 오류 그래프 모양은?

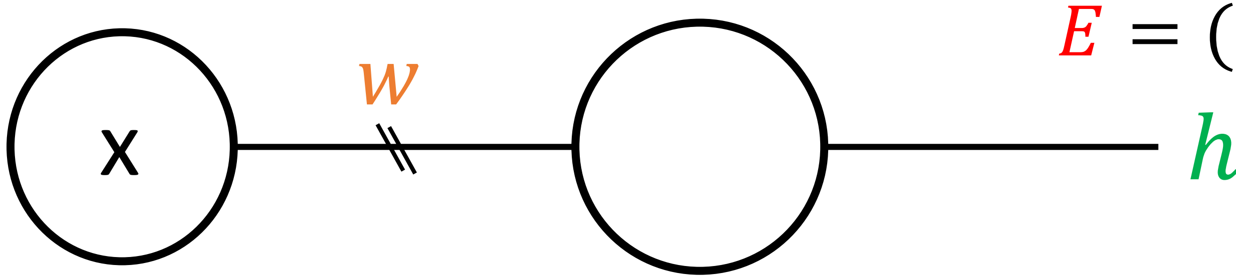
$$E = \frac{1}{3}\sum_{i=1}^{3}(wx_i - y_i)^2$$

x	y
1	1
2	2
3	3

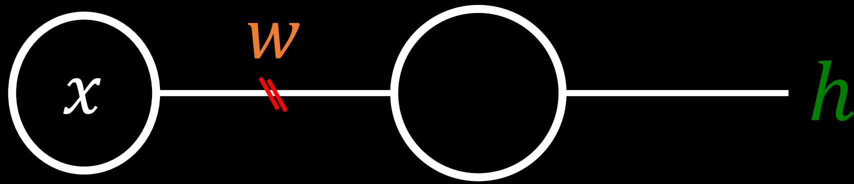

Axon
terminal
(tip)
dendrite
Synapse
electrical
signal


h
x
x
w
Synapse

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

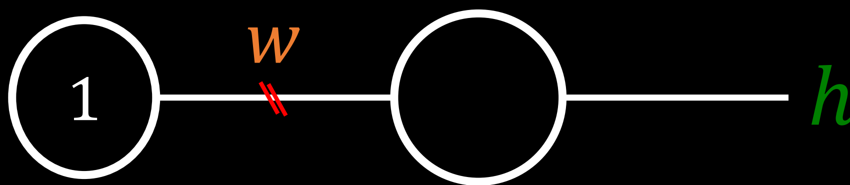

w
x
h

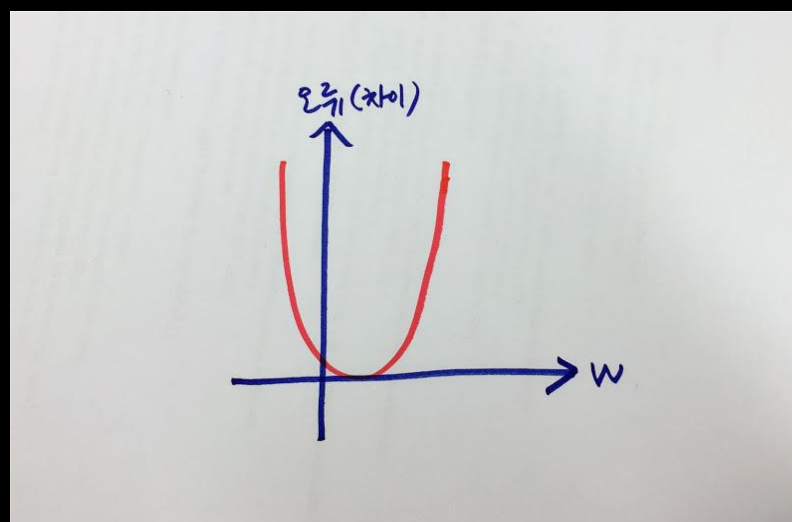
$$E = (h - y)^2$$



x	y
1	1

$$E = (h - 1)^2$$





오류 그래프 생각하기

- w 가 변하면 오류 E도 변한다.
- 오류를 줄이고 싶으면 w를 적절히 바꾼다.
- w 위치에 따라 w 를 조금만 변경해도 오류가 많이 변하는 곳도 있고,
- 어떤 곳에서는 w를 바꾸어도 오류가 거의 변하지 않는 곳이 있다.

(Q) w 예상하기

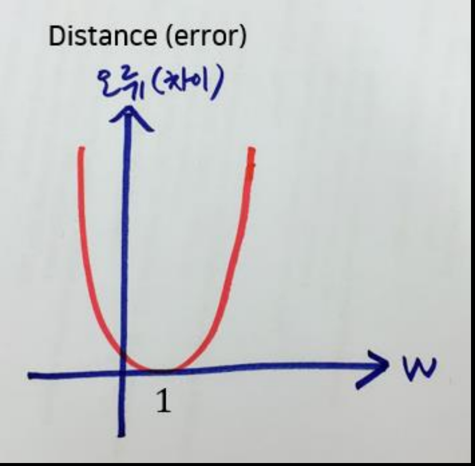
- w 를 조금 증가시켰더니 오류 E가 아주 급격하게 늘어났다. 현재 w 값은?
- w를 조금 증가시켰더니 오류는 급격하게 감소하였다. 현재 w 값은?
- w 를 변경해 보았지만 오류는 거의 변하지 않았다. 현재 w 값은?

현재 w가 어떤 값일 때

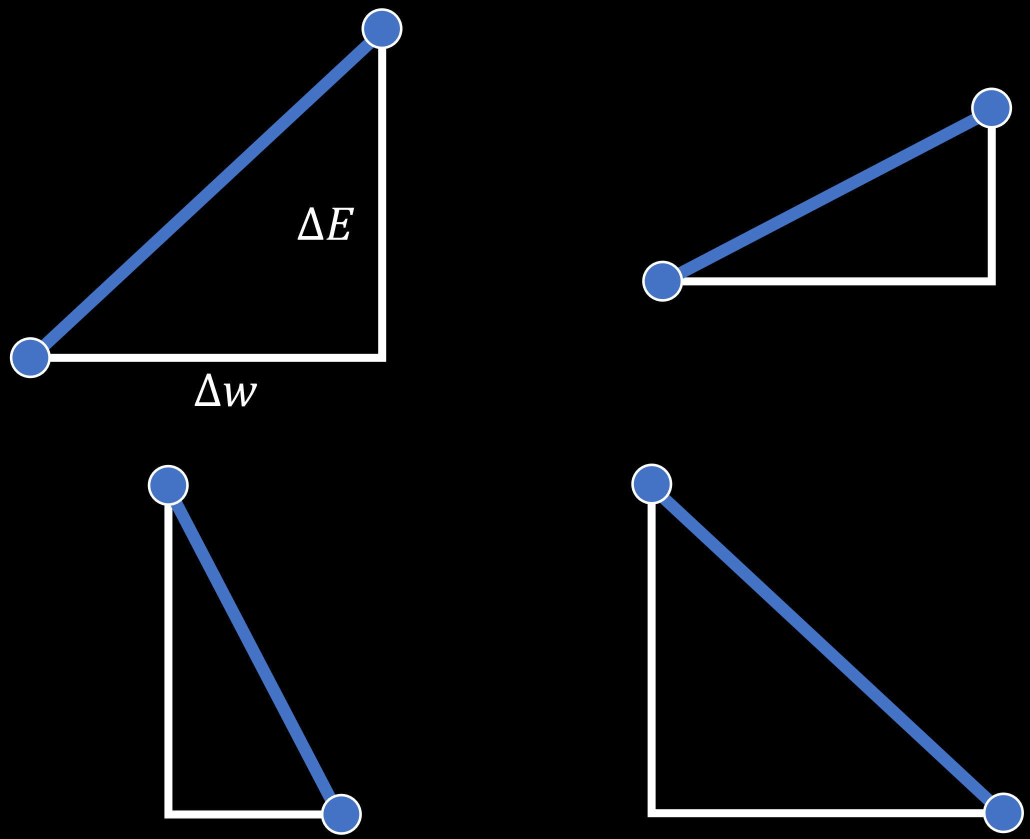
w변화가 오류 E에 미치는 영향

예를 들어, 현재 w가 4인 곳에서는 w가 조금만 늘려도
오류(E)는 아주 크게 늘어난다.
“기울기가 아주 크다”

기울기로 표현된다.

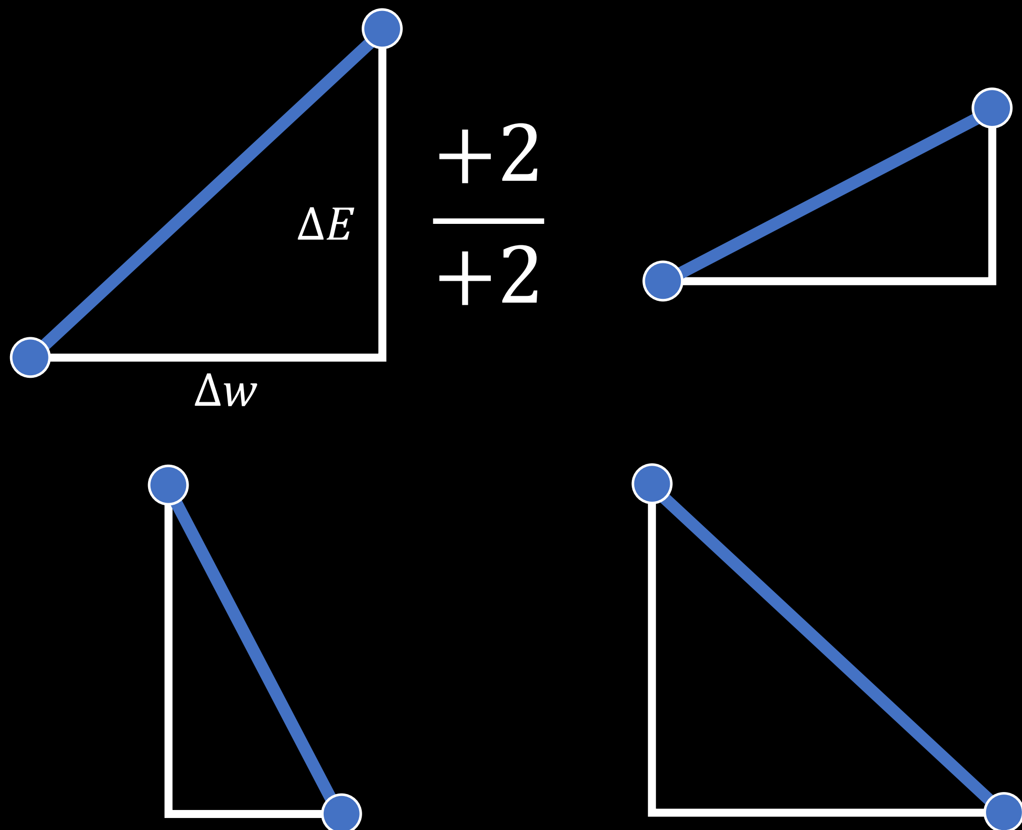


기울기 의미(w와 E 관계)



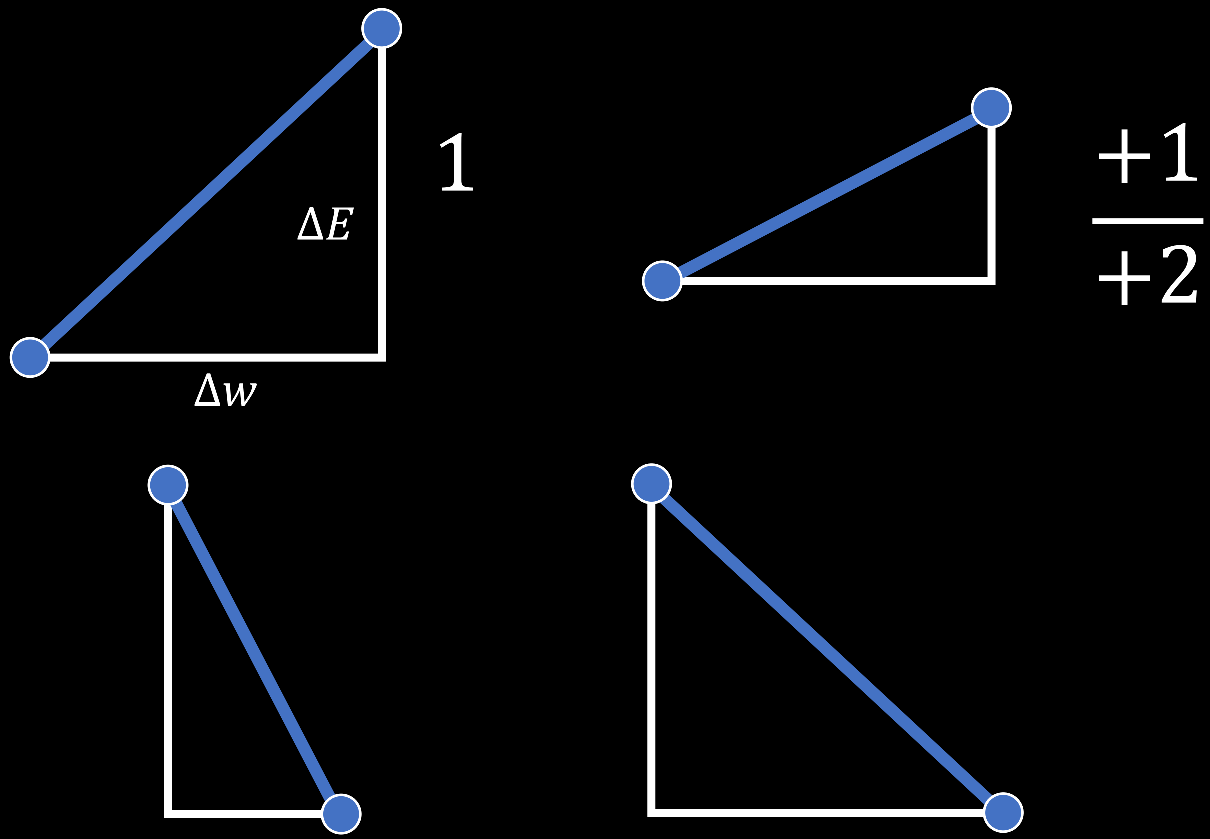
오류 그래프에서 두 점을 연결했을 때의 기울기

기울기 의미(w와 E 관계)



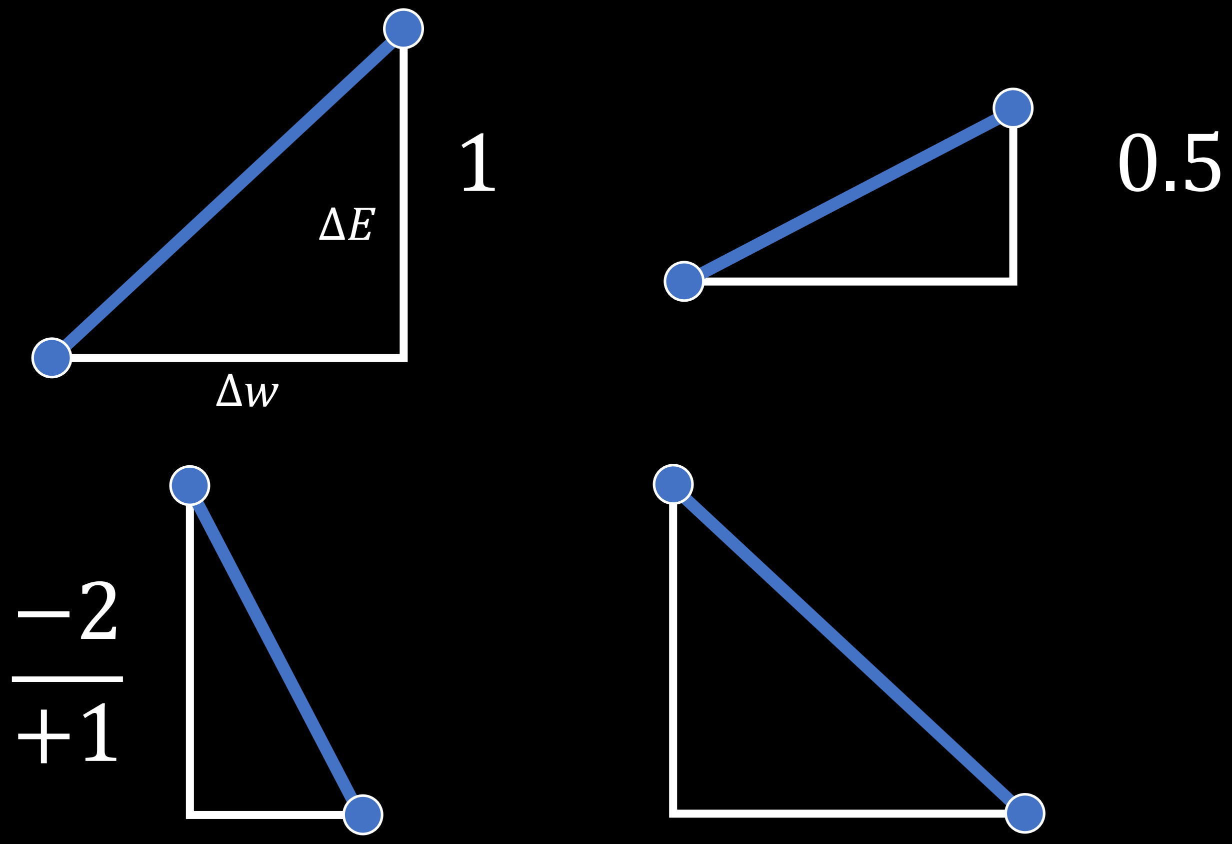
오류 그래프에서 두 점을 연결했을 때의 기울기

기울기 의미(w와 E 관계)



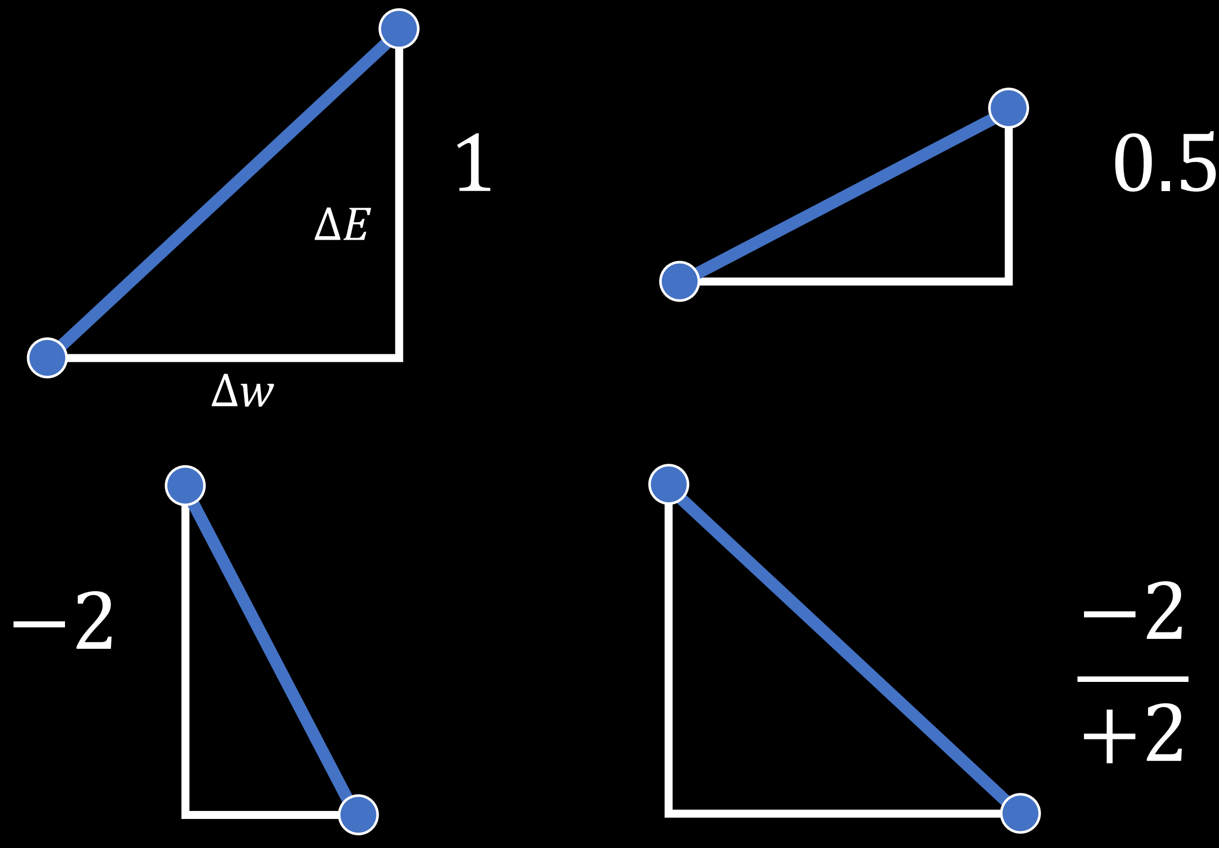
오류 그래프에서 두 점을 연결했을 때의 기울기

기울기 의미(w와 E 관계)



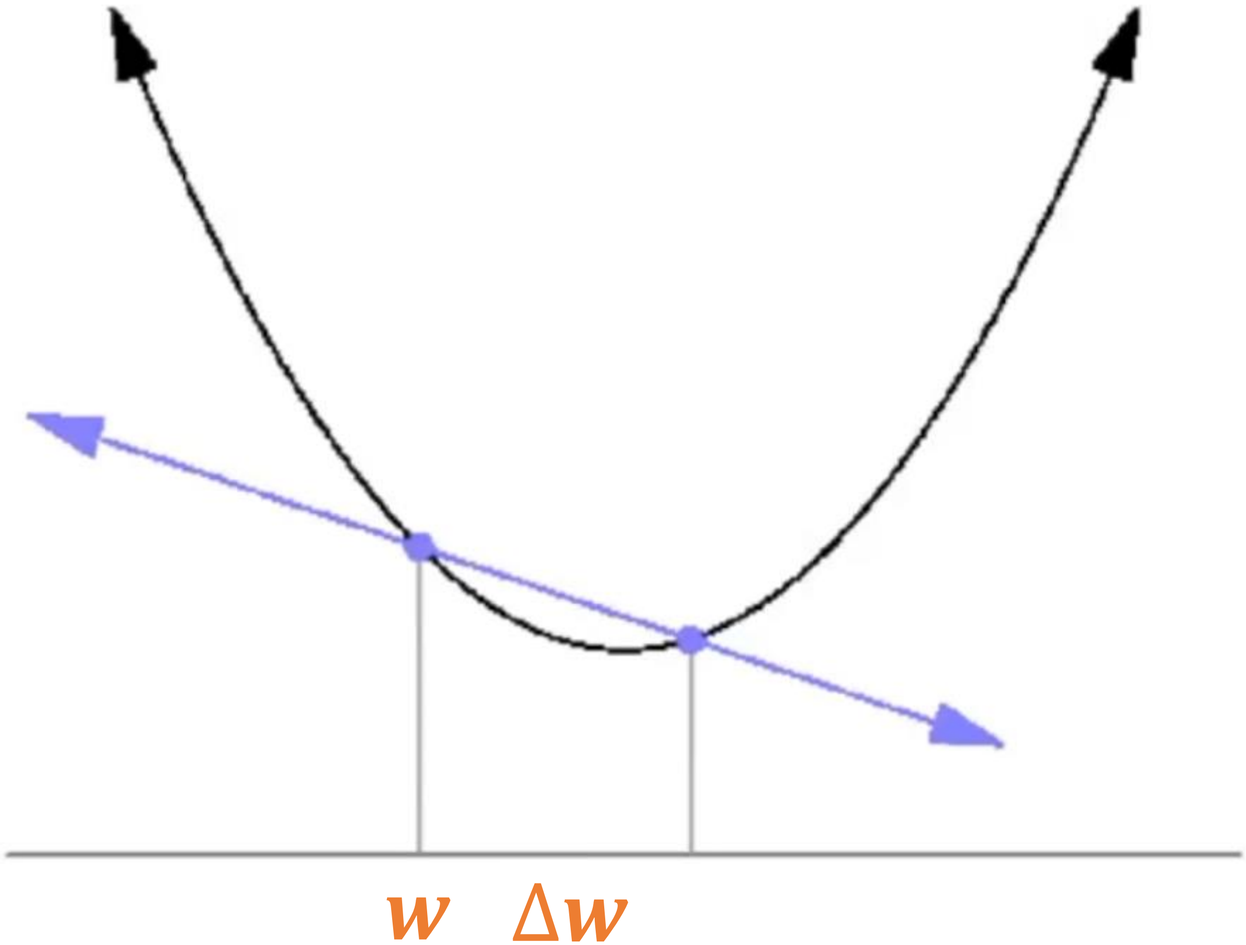
오류 그래프에서 두 점을 연결했을 때의 기울기

기울기 의미(w와 E 관계)

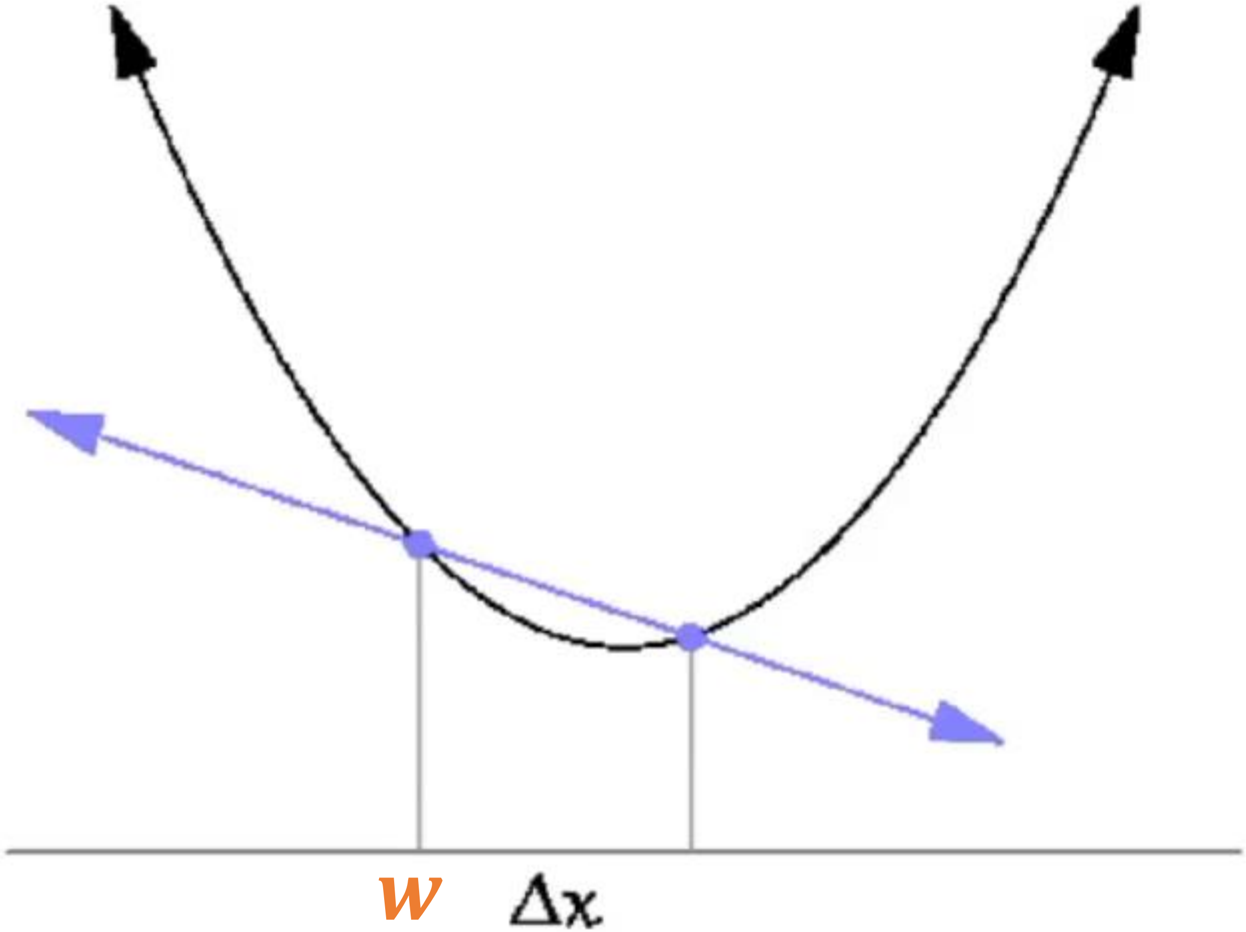


오류 그래프에서 두 점을 연결했을 때의 기울기

오류 그래프에서 기울기



오류 그래프에서 두 점을 연결했을 때의 기울기



w
Δx

w를 아주 조금만(*) 변경할 때
오류 E가 얼마나 증감하는 지 = 기울기 =
w의 변화가 오류 E에 미치는 영향

Numerical differentiation

(Q) 미치는 영향 구하기

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

데이터 (x, y)가 (1, 1)일 때
w=3인 지점에서 w변화가 오류 E에
미치는 영향(기울기)를 구하라.

(Q) 미치는 영향 구하기

$$E = (w1 - 1)^2$$

w: 3 -> E: 4

w: 3.00001 -> E: 4.00004

w가 0.00001 증가할 때 E는 0.00004 증가

w가 1증가할 때 E는 4 증가함을 의미

따라서 미치는 영향(기울기)=4

기울기=미치는 영향

기울기에 대한 생각

- 기울기가 4 → w가 1만큼 증가하면 오류 E는 4만큼 증가
- 따라서 오류를 줄이고 싶으면? w를 감소시켜야 함
- 기울기가 -5 → w가 1만큼 증가하면 오류는 5만큼 감소 뜻
- 따라서 오류를 줄이고 싶으면? w를 증가시켜야 함

w를 변화시키면서(튜닝)
오류 E를 계속해서 줄이는 것
(학습, Learning)

어떻게 '자동으로'

- 오류 E를 최소화하는 w 값을 찾을까?

요 약

- 리그레션(회귀)을 이해하였다.
- 가설과 오류 그래프를 이해하였다.
- 오류 그래프를 해석할 수 있다.
- 가중치가 오류에 미치는 영향을 알았다.
- 기울기의 의미를 알았다.