AI and Deep Learning

Linear Regression & Back-propagation (1)

Jeju National University

Yung-Cheol Byun

c:\>git <u>clone</u> https://github.com/yungbyun/mllecture.git

여행, 그리고 회귀

회귀(Regression)

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


KBS1 HD

회귀는 자연의 법칙, 섭리, 일종의 규칙

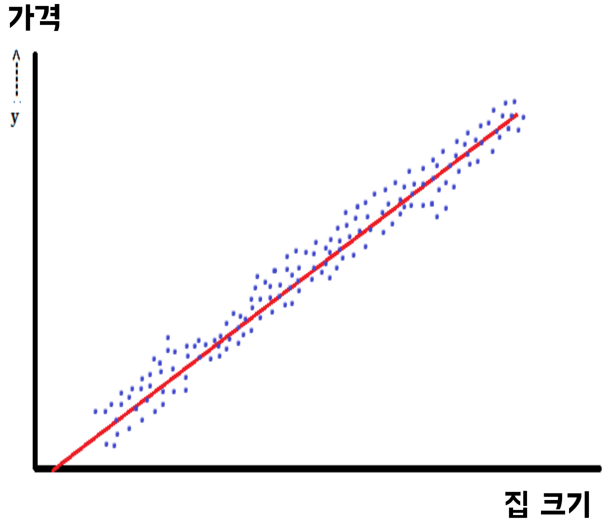
- 연어는 태어난 곳으로 돌아온다.
- 집은 클수록 비싸다.
- 젊을 때는 강하지만 나이가 들수록 약해진다.
- 남자가 여자보다는 큰 편이다.
- 성적이 좋을 수록 취업이 잘된다.

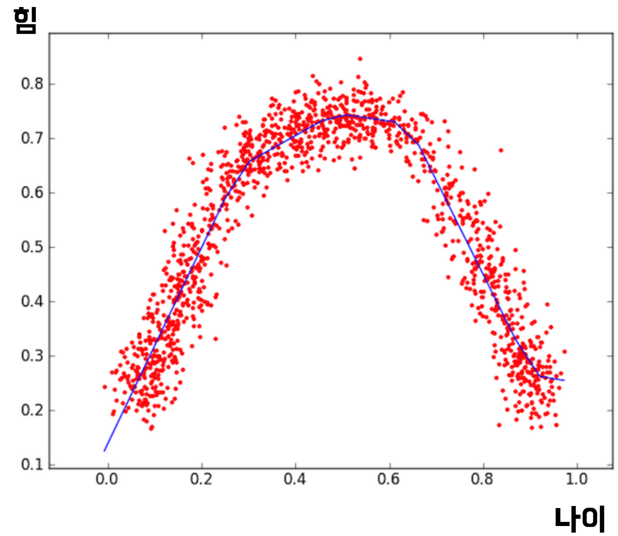
반드시 그런 것은 아니지만 일반적으로 그런 경향이 있다.

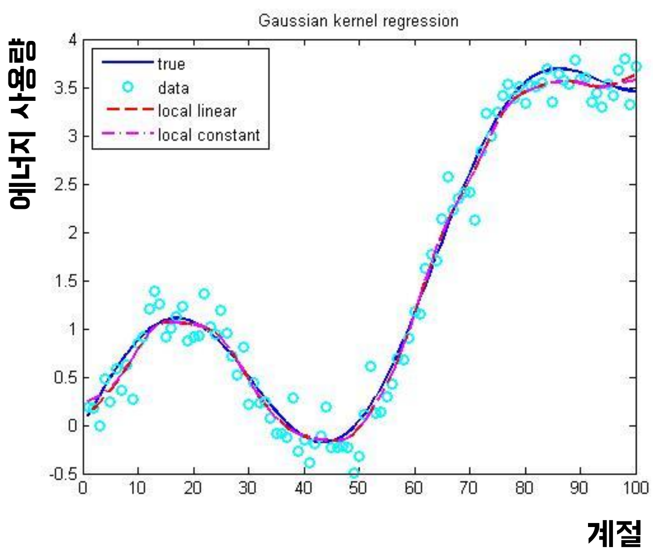
이런 '일종의 규칙'이 회귀이고 이 때문에 '예측'을 할 수 있다.

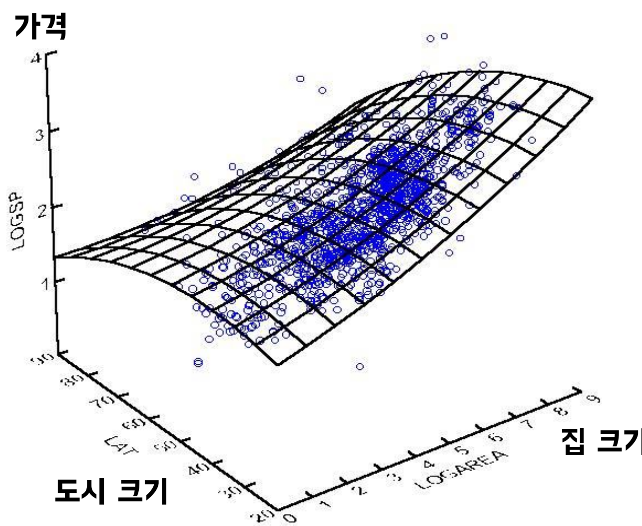
이를 잘 표현하는 말,

회귀(Regression)


가격
y
집 크기


힘
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0.8
1.0
나이


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


가격
LOGSP
LAT
LOGAREA
집 크기
도시 크기

집 크기와 가격의 관계

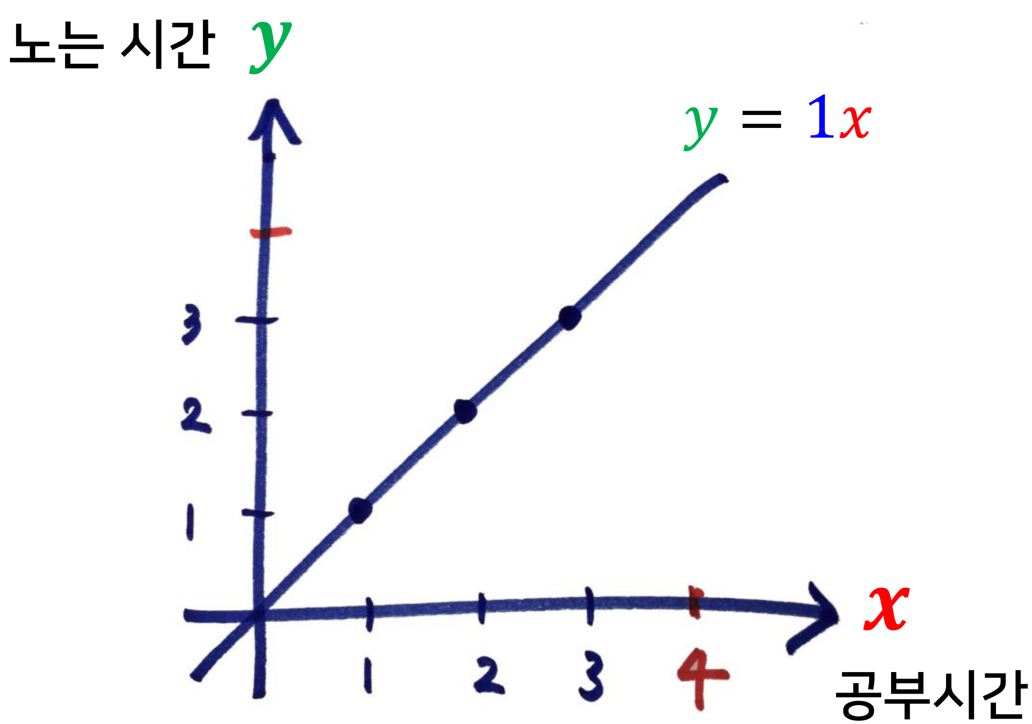
이렇게 회귀는
분포, 그래프로 표현하면
이해하기 쉽다.

Linear Regression

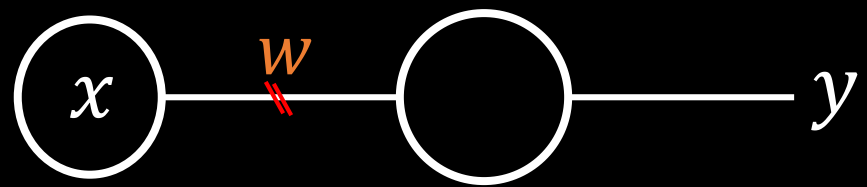
그래프, 분포 형태가 직선(Line)

www.desmos.com

1. 점 (1, 1) 표시
2. 여러 점 (2, 2), (-1, -1), (-2, -2) 추가 -> 분포 확인
3. y = x
4. y = 2x
5. 모든 점 조절
6. y = wx
7. y = wx + 1 (그래프 이동)
8. 모든 점 조절
9. y = wx + b (그래프 회전과 이동)

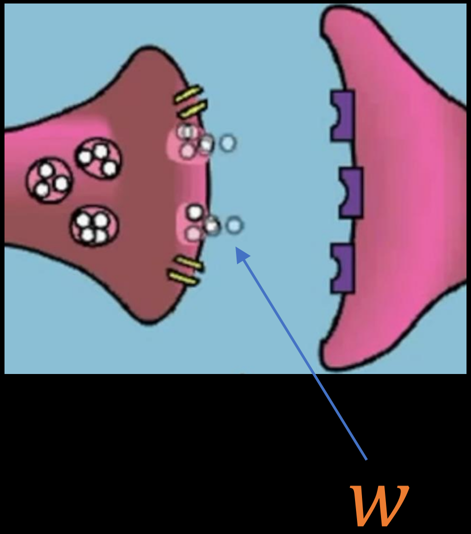

노는 시간 y
y = 1x
3
2
1
1
2
3
4
x
공부시간

뉴런의 능력



- $y = wx$
- 하나의 뉴런은 하나의 Linear Regression을 표현할 수 있다.

신경전달 물질의 양이



맞으면?
적으면?
없으면?

가설(Hypothesis)

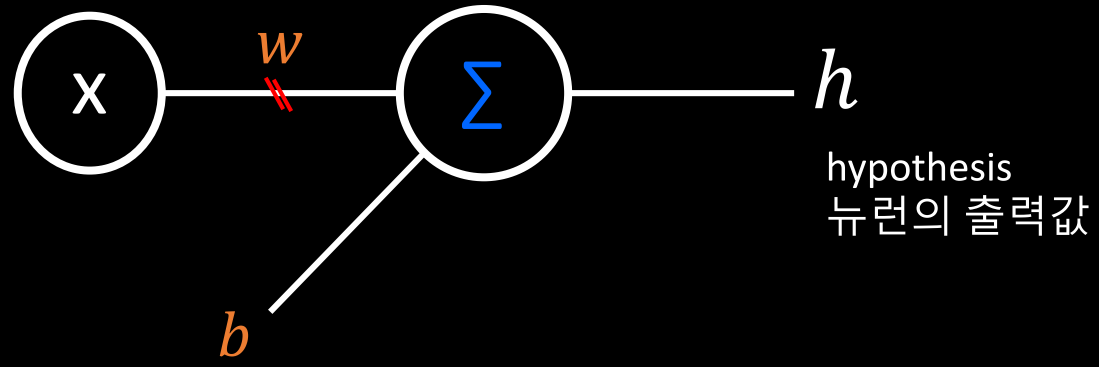
$$h = wx$$

b가 있으면 뉴런은 더 다양한 분포(회귀)를 표현할 수 있음.

$$h = wx + b$$

- 뉴런의 출력을 식으로 표현
- 가설: 증명되지 않았으나 w 조절을 통하여 데이터의 Linear Regression을 잘 표현할 수 있는 것

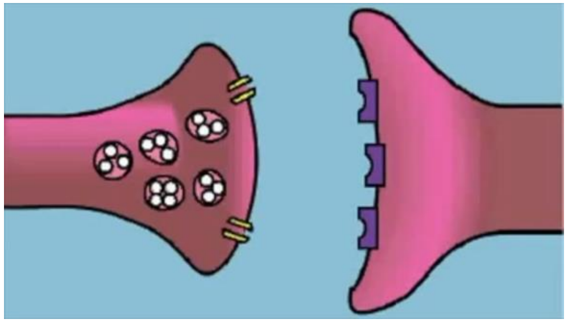
가중치 w와 바이어스 b 역할



어떻게 w를 조정할 것인가?

- 야단치면 된다.
- 그러면 다음에 실수를 덜하도록 가중치 w가 조정된다.
- 모든 점을 지날 경우 차이(오류, 에러 E, 비용, loss)는 0
- 차이가 0이 될 때까지 w를 조정하자.

공부한 시간	w	뉴런 출력	정답	오차(차이)	대가
1	7	7	1	7-1	크게 야단
1	4	4	1	4-1	보통 야단
1	2	2	1	2-1	조금 야단
1	1.5	1.5	1	1.5-1	아주 조금
1	1.3	1.2	1	1.2-1	매우 조금
1	1.1	1.1	1	1.1-1	굳!



개, 돌고래, 아이, 희한하게도 잘못할 경우 야단을 치면 신기하게도 '자동으로' 연결부위 w값(연결강도)이 수정되어 오차(차이)가 줄어듦.

오류 함수

왜 절대값인가?

$$오류(E) = |\underbrace{뉴런이\ 출력한\ 값}_{가설(hypothesis)} - 정답|$$

$$E = |wx - y|$$

오류 함수

x	y
1	1

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

desmos

점 (1, 1) 표시
y = wx
E = w * 1 – 1
E = | w * 1 – 1 |

오류 함수

x	y
1	1
2	2
3	3

만일, 데이터가 3개라면

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

오류를 모두 더해서 평균

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

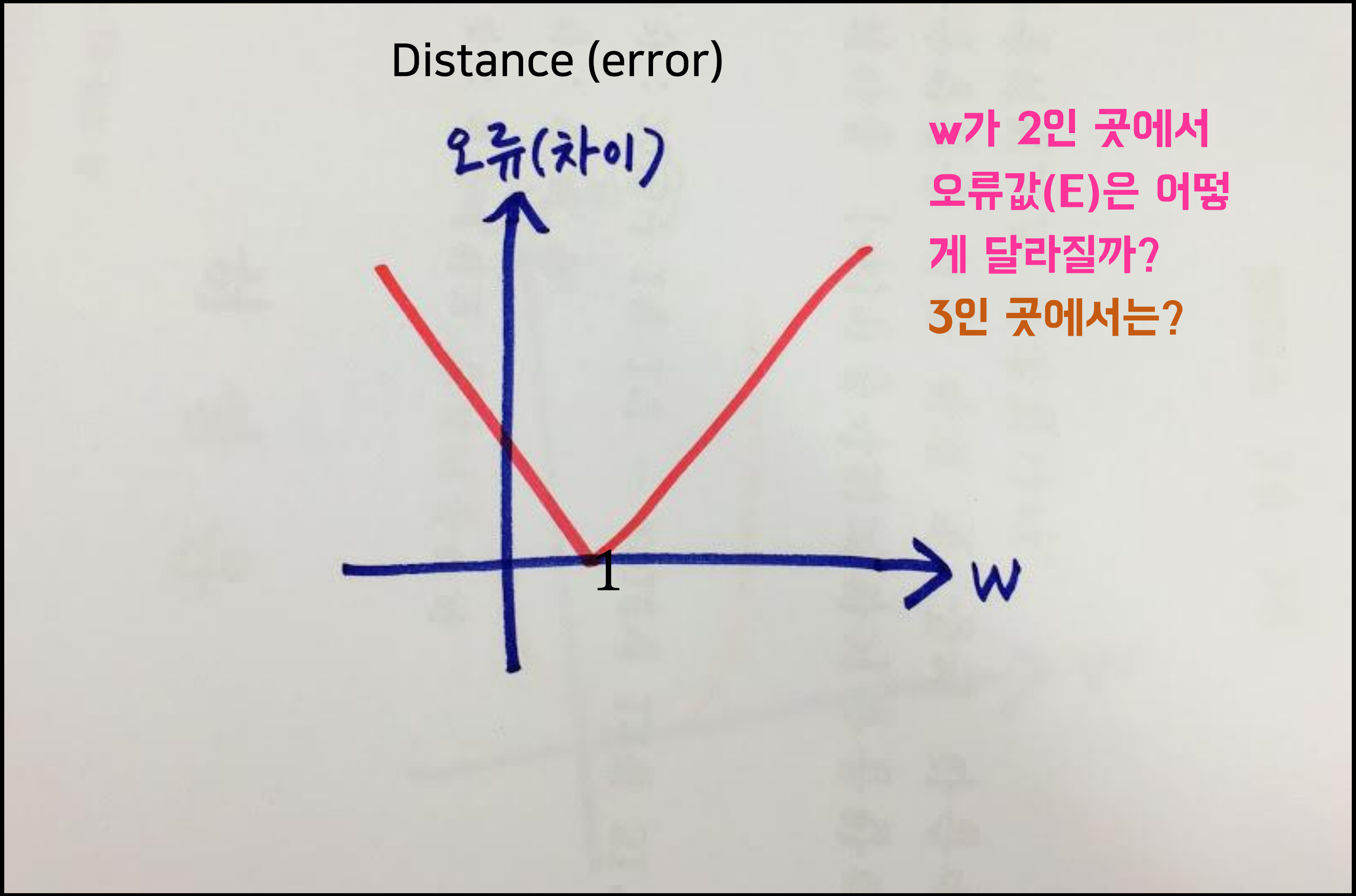
점 (2, 2), (3, 3) 추가

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

(w, E) 오류 그래프 그리기

오류 함수 그래프

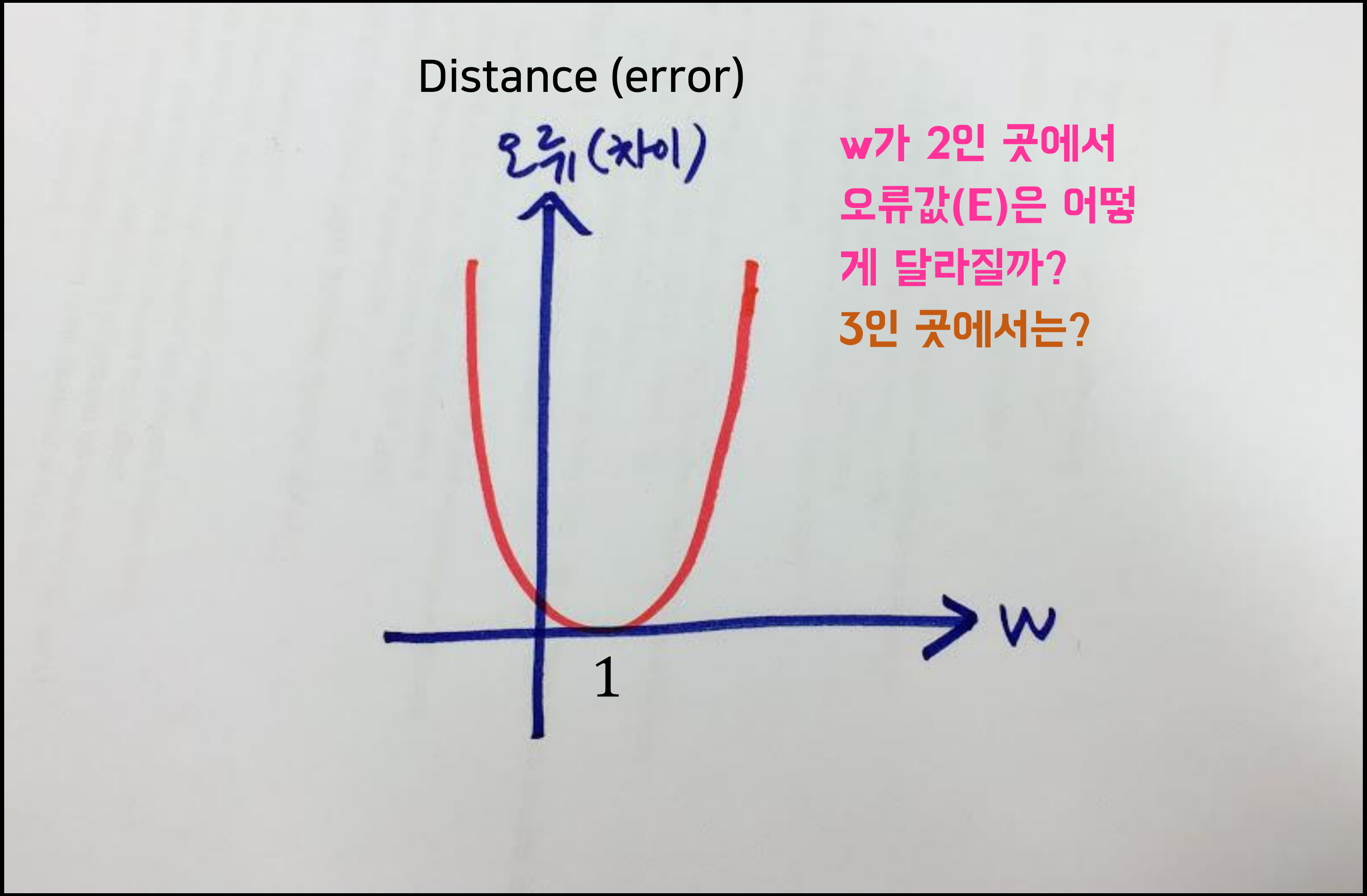
$$E = \frac{1}{3}\sum_{i=1}^{3}|wx_i - y_i|$$



오류 함수 그래프

Mean Square Error

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



오류 함수

desmos

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

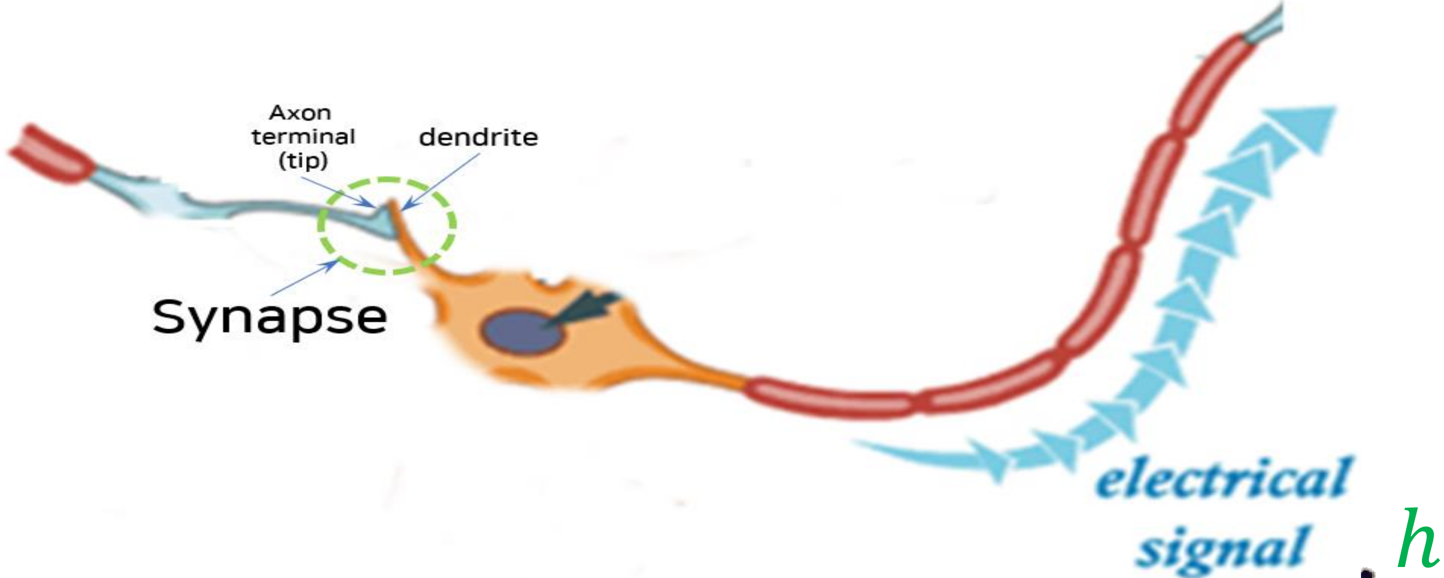
데이터가 m개일 경우

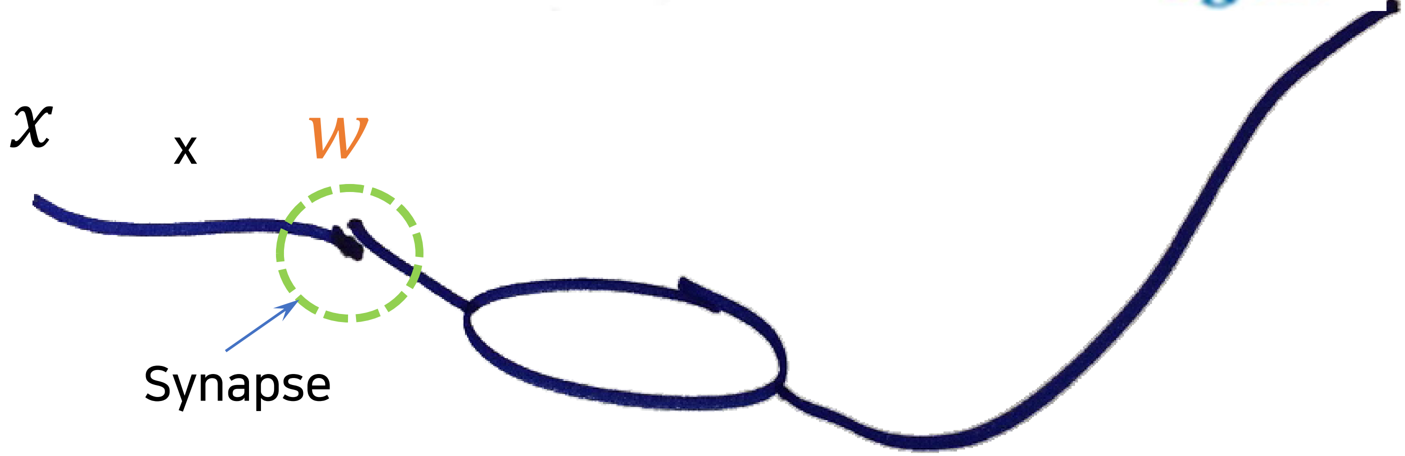
Mean Square Error

뉴런이 예측한 값 (가설)

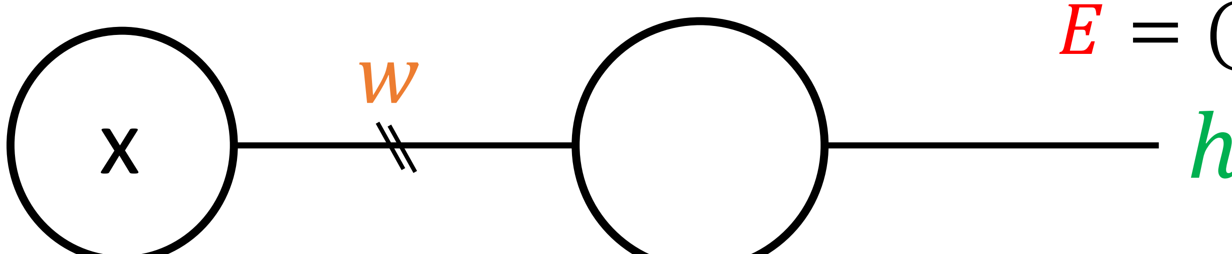
정답

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

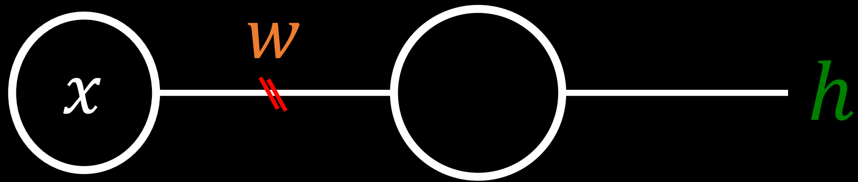

Axon
terminal
(tip)
dendrite
Synapse
electrical
signal


h
x
x
w
Synapse

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

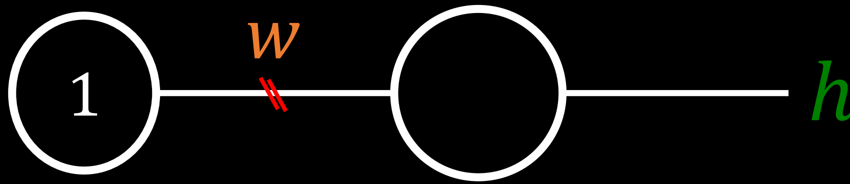

x
w
h

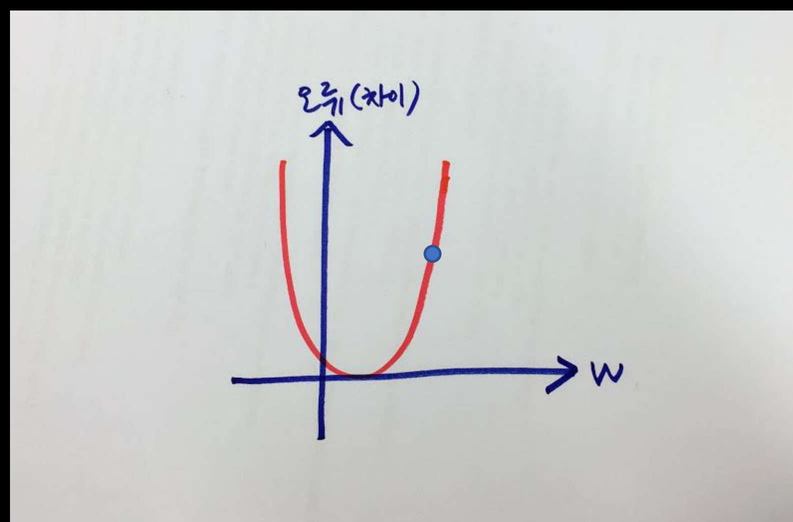
$$E = (h - y)^2$$



x	y
1	1

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





오류 그래프 생각하기

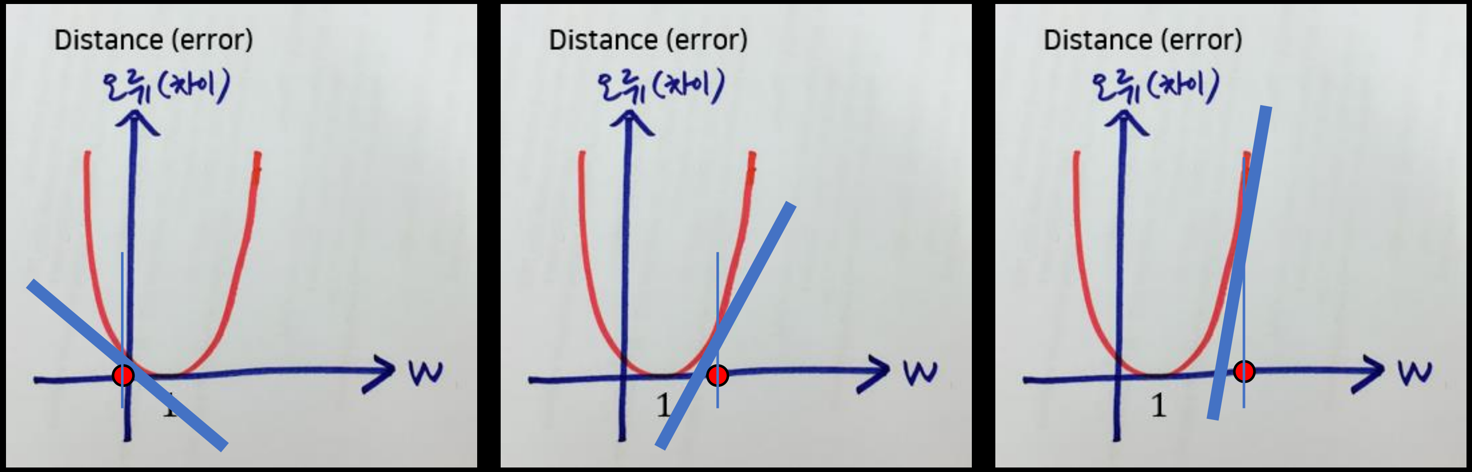
- w 가 변하면 오류 E도 변한다.
- 오류를 줄이고 싶으면 w를 적절히 바꾼다.
- w 위치에 따라 w 를 조금만 변경해도 오류가 많이 변하는 곳도 있고,
- 어떤 곳에서는 w를 바꾸어도 오류가 거의 변하지 않는 곳이 있다.

(Q) w 예상하기

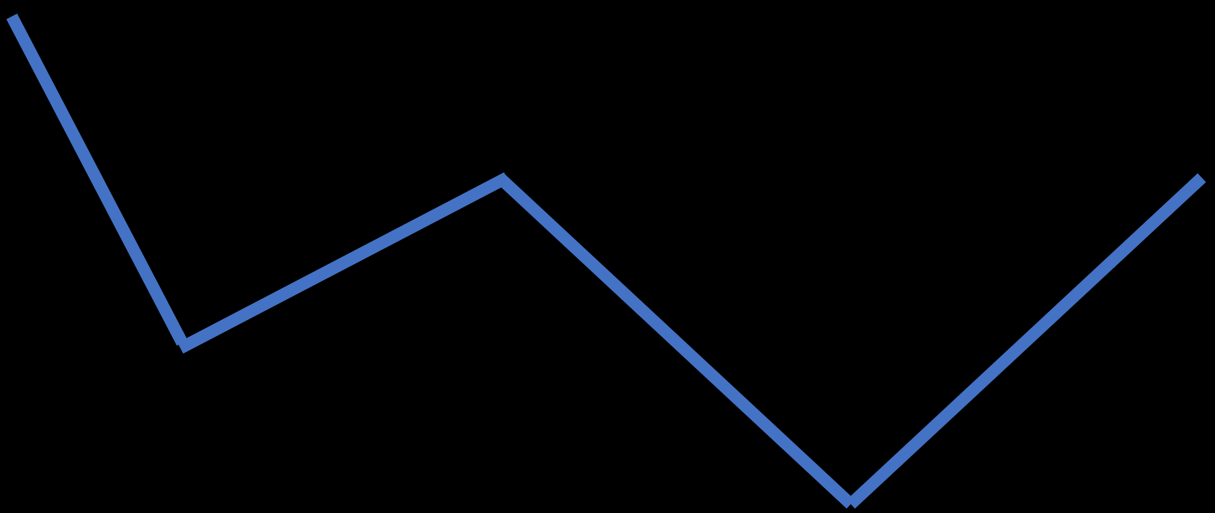
- w 를 조금 증가시켰더니 오류 E가 아주 급격하게 늘어났다. 현재 w 값은?
- w를 조금 증가시켰더니 오류는 급격하게 감소하였다. 현재 w 값은?
- w 를 변경해 보았지만 오류는 거의 변하지 않았다. 현재 w 값은?

현재 w가 어떤 값일 때

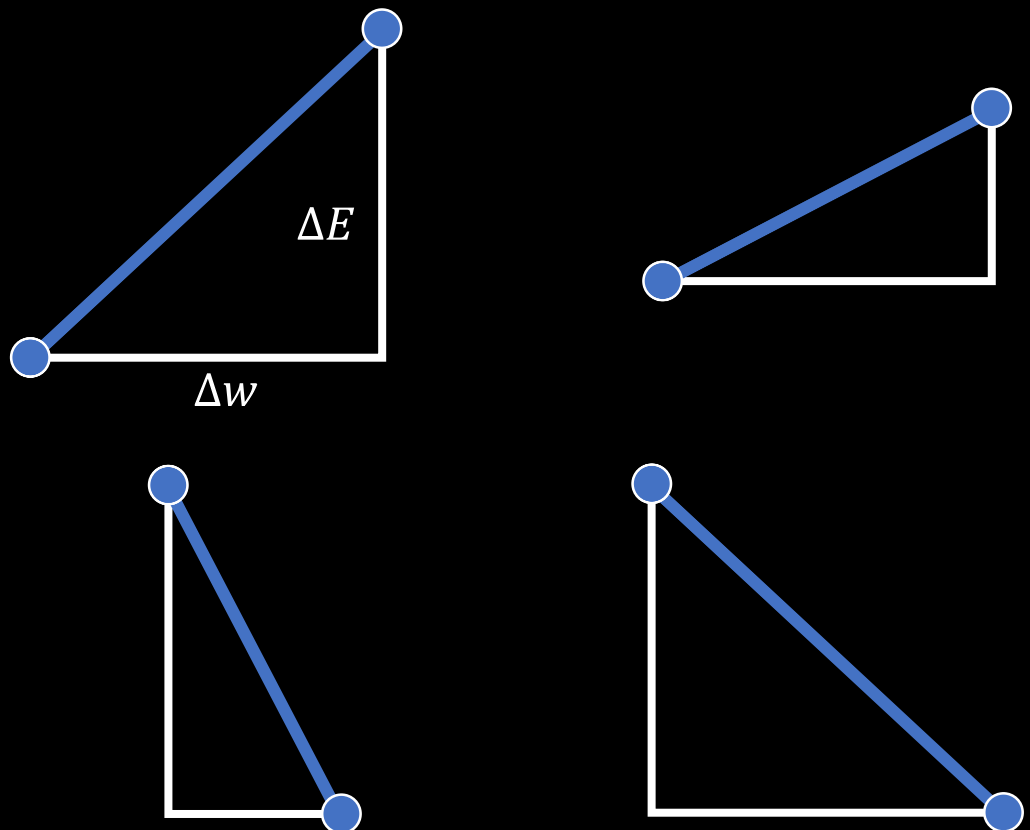
w변화가 오류 E에 미치는 영향



기울기로 표현

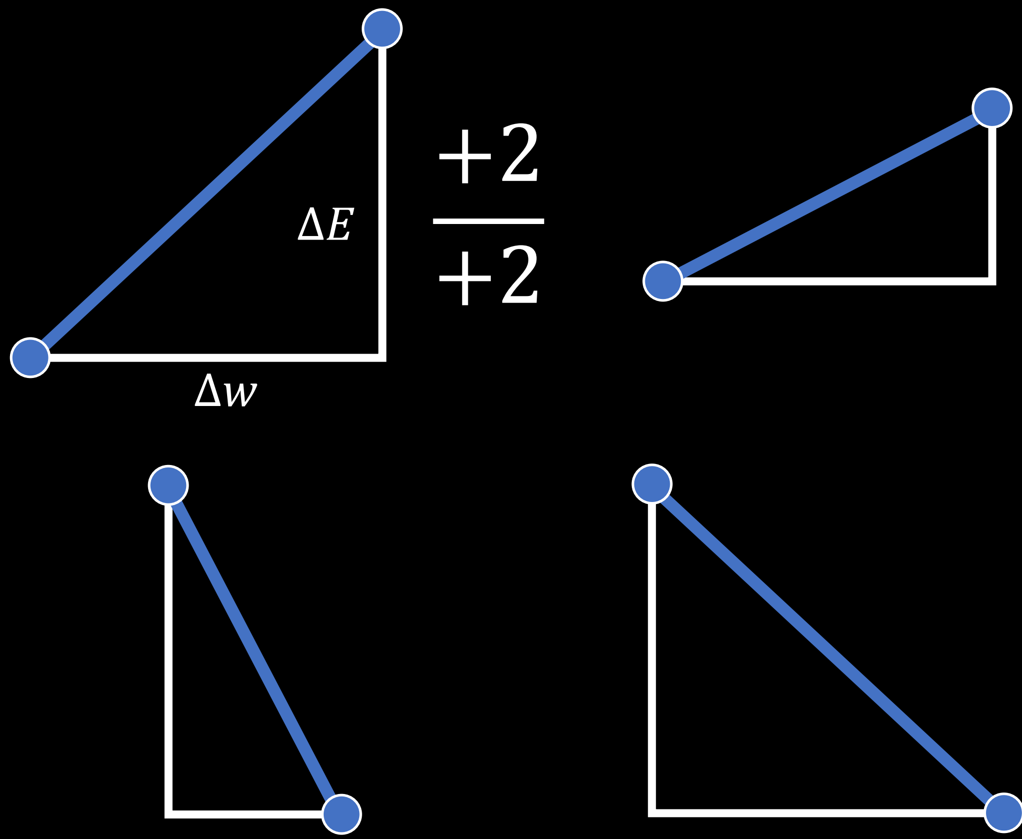


기울기 의미(w와 E 관계)



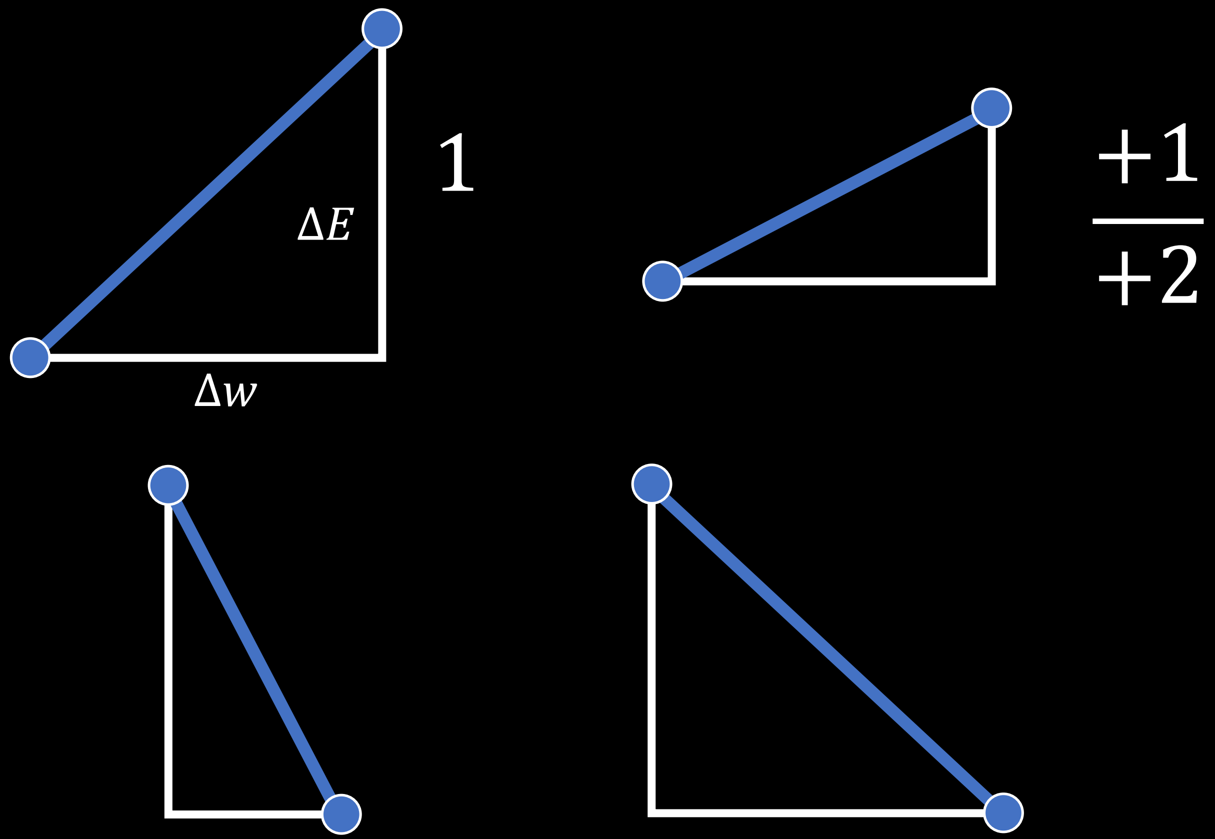
오류 그래프에서 두 점을 연결했을 때의 기울기

기울기 의미(w와 E 관계)



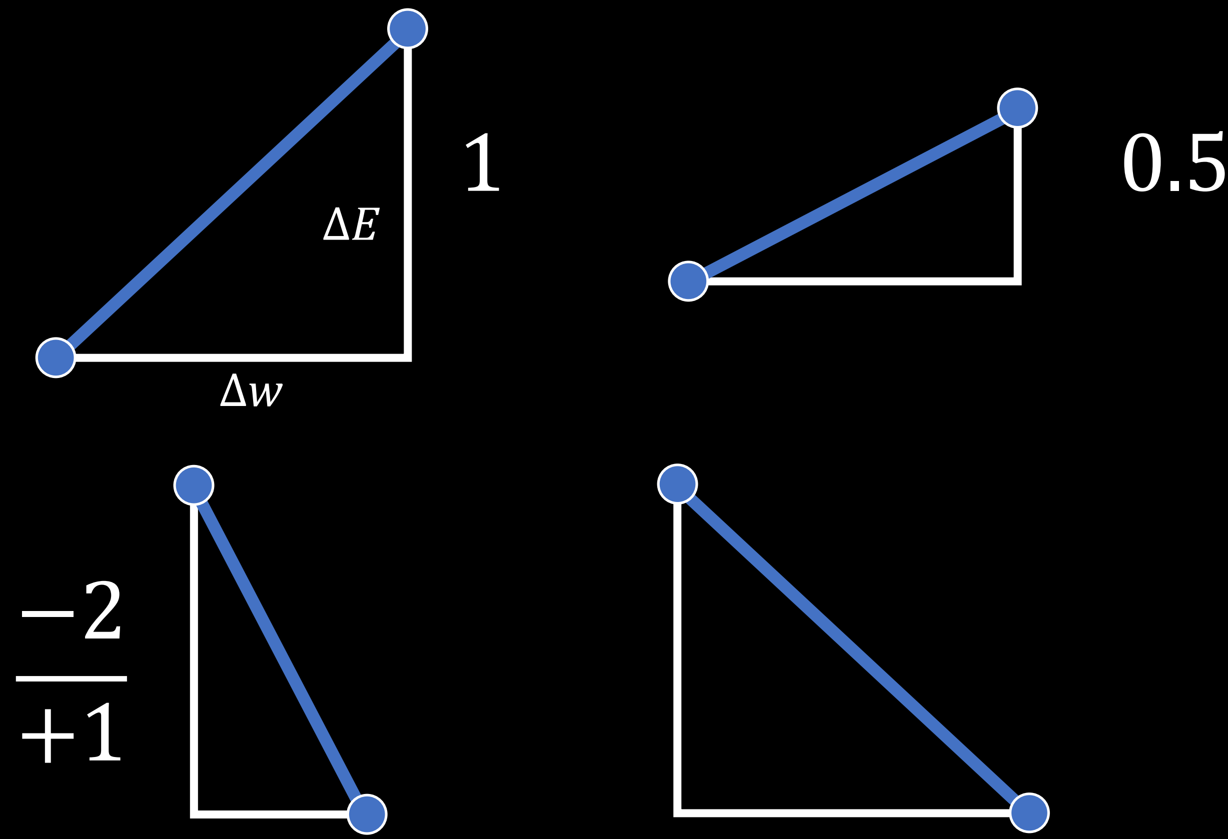
오류 그래프에서 두 점을 연결했을 때의 기울기

기울기 의미(w와 E 관계)



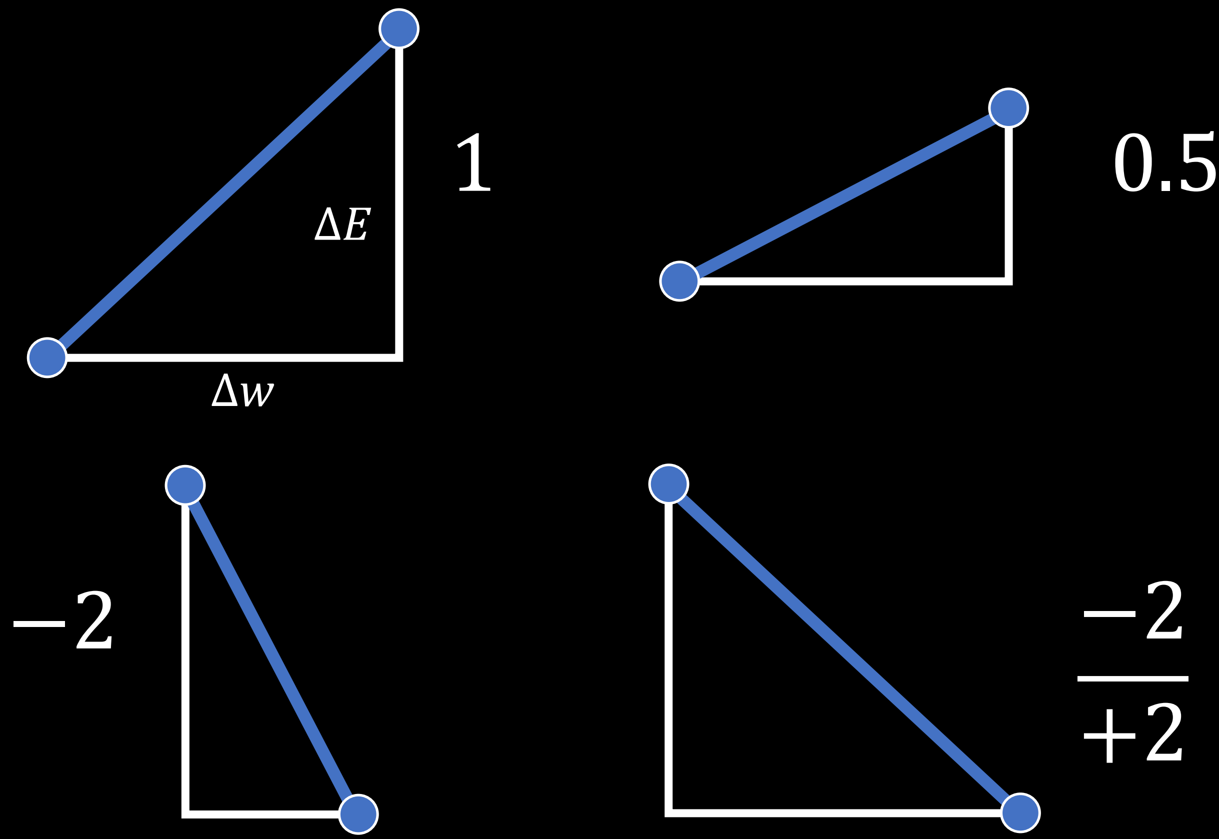
오류 그래프에서 두 점을 연결했을 때의 기울기

기울기 의미(w와 E 관계)

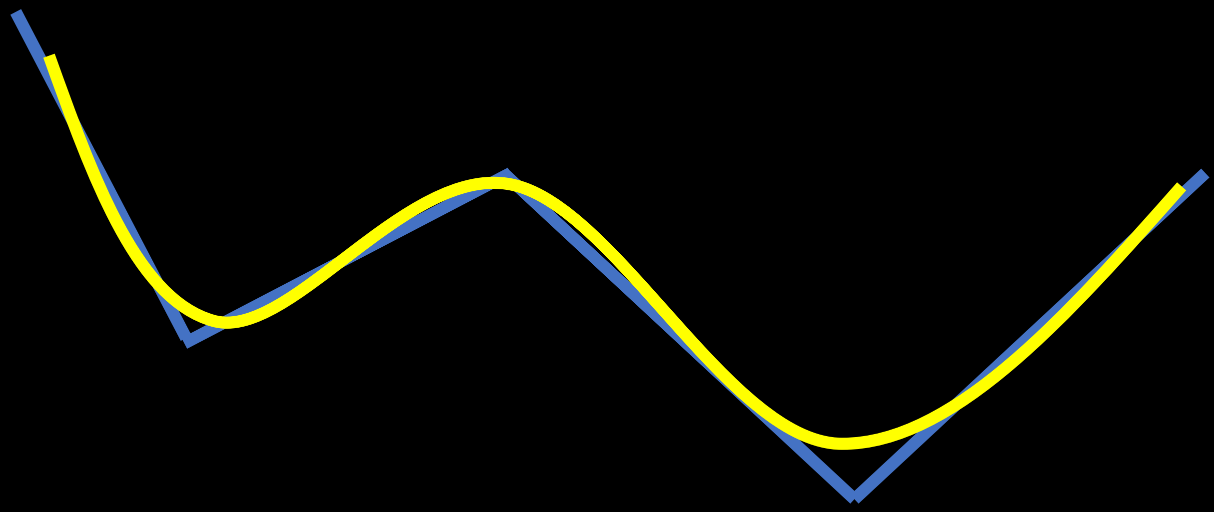


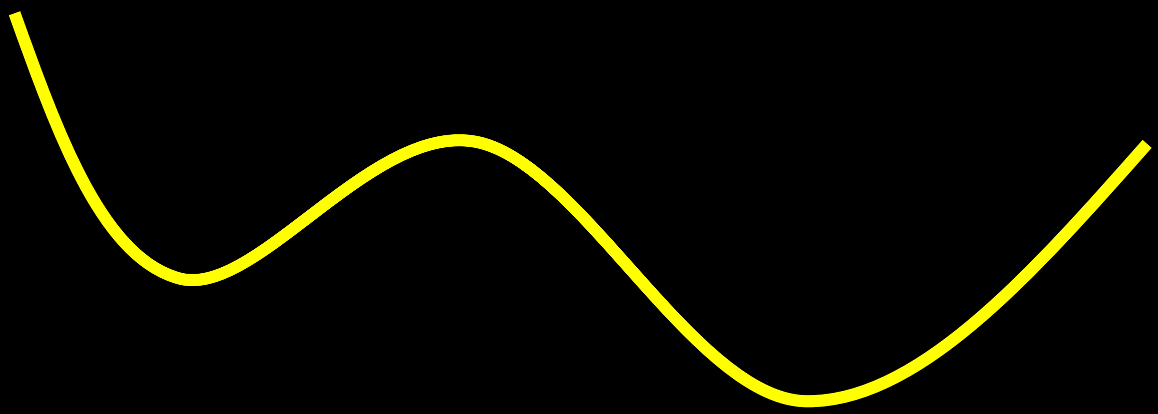
오류 그래프에서 두 점을 연결했을 때의 기울기

기울기 의미(w와 E 관계)

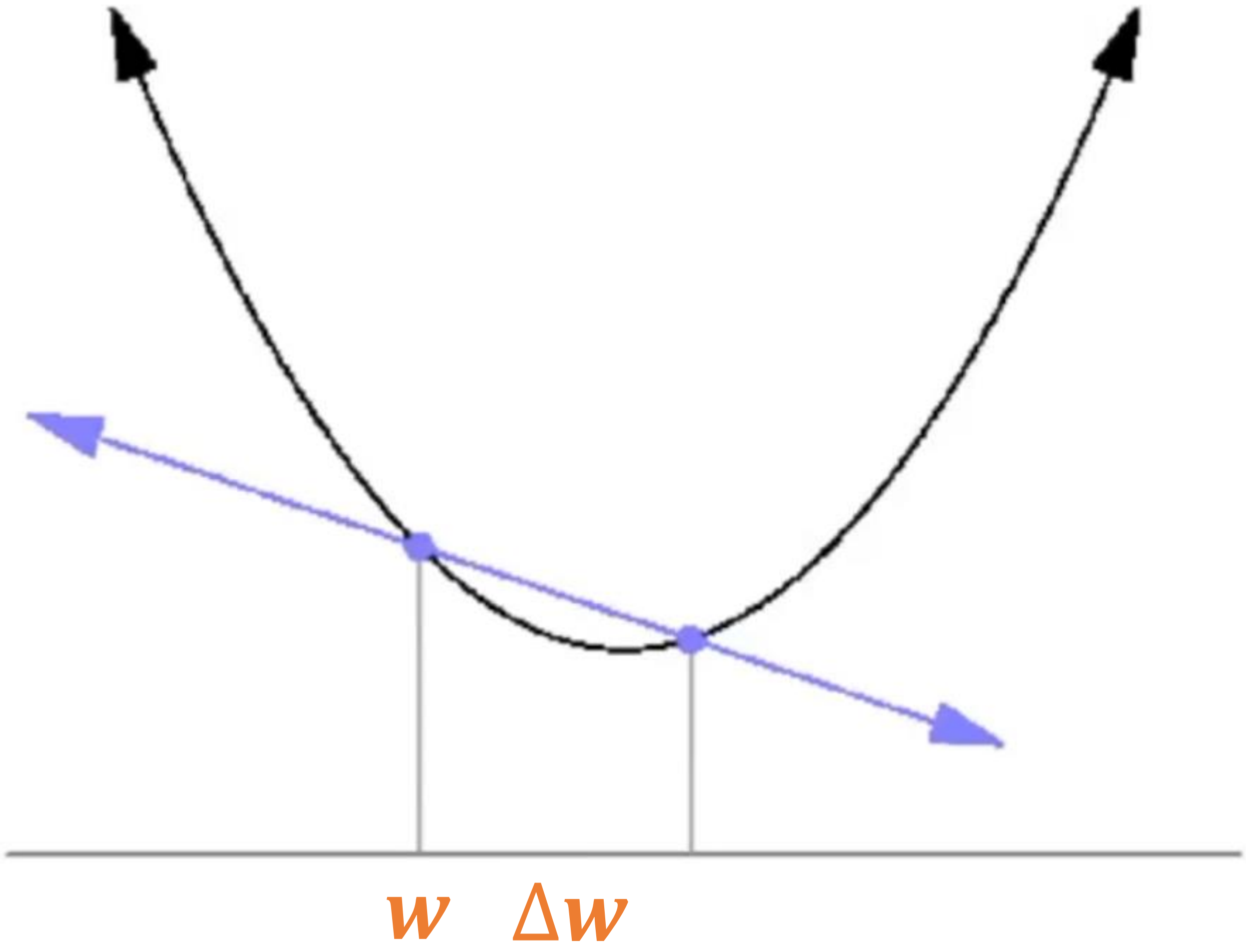


오류 그래프에서 두 점을 연결했을 때의 기울기

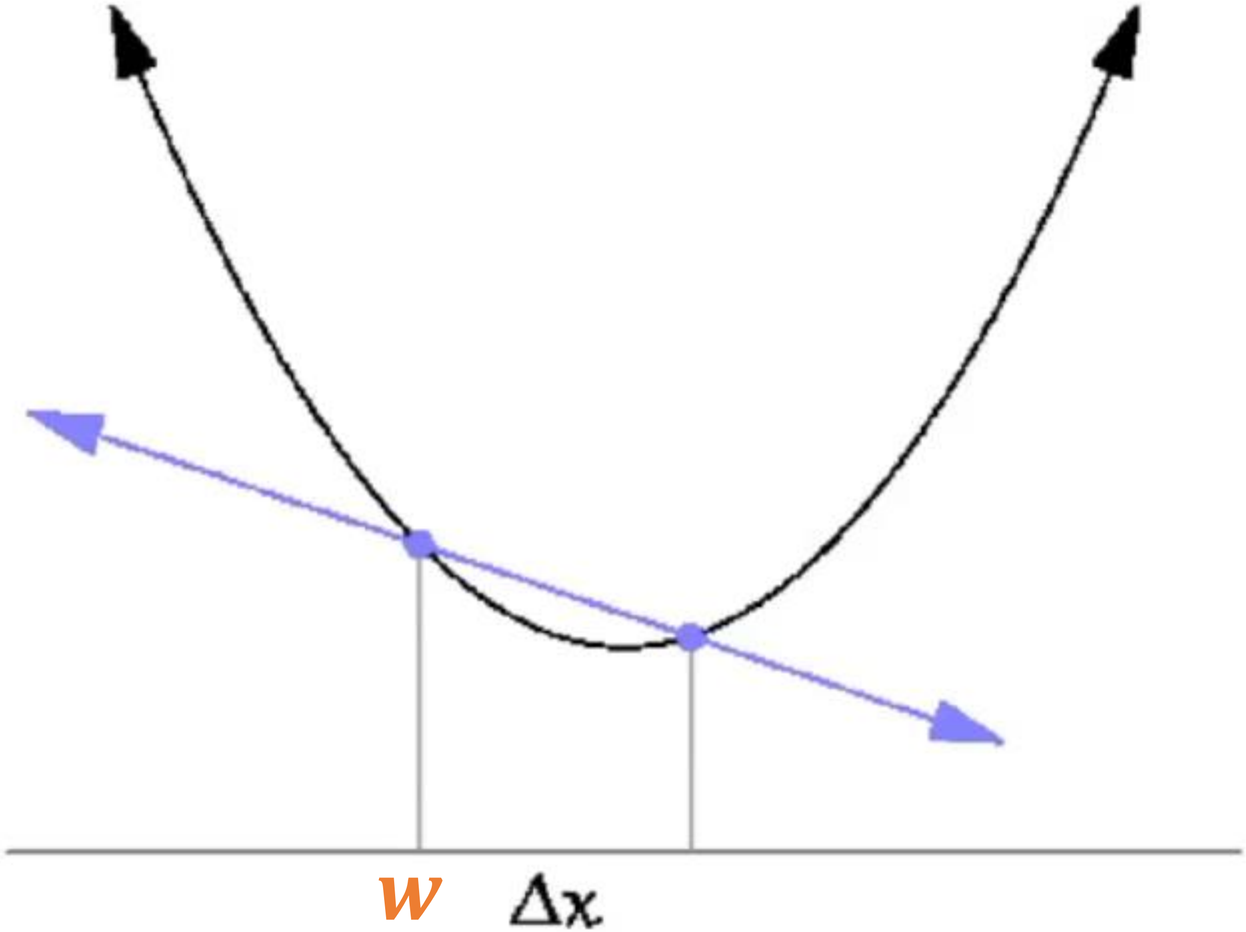




오류 그래프에서 기울기



오류 그래프에서 두 점을 연결했을 때의 기울기


w
Δx

w를 아주 조금만 키울 때(Δw)
기울기가 더 정확하다.

Numerical differentiation

w의 변화가 에러 E에 미치는 영향, 기울기

(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

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

w: 3 -> E: 4

w: 3.00001 -> E: 4.00004

w가 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 기울기 = 미치는 영향 = 4

오류 E가 감소하도록
w를 계속해서 업데이트

학습, Learning, 파라미터 튜닝

어떻게 자동으로

- 오류 E를 최소화하는 w 값을 찾을까?

요 약

- 리그레션(회귀)을 이해할 수 있다.
- 가설과 오류(에러, 스트레스) 그래프를 이해할 수 있다.
- 가중치가 오류에 미치는 영향의 의미를 알 수 있다.
- 기울기의 의미를 이해할 수 있다.