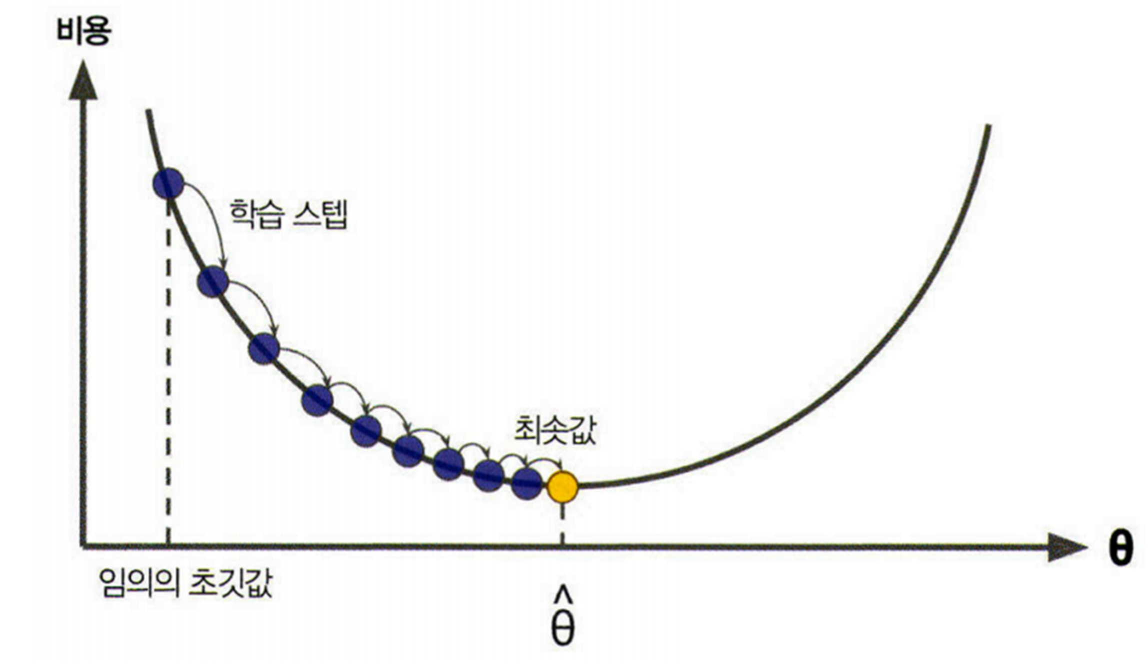
복습 세션
Review Session

3조 이원재 이승재 장윤서 윤희재

목차

경사하강법이란?



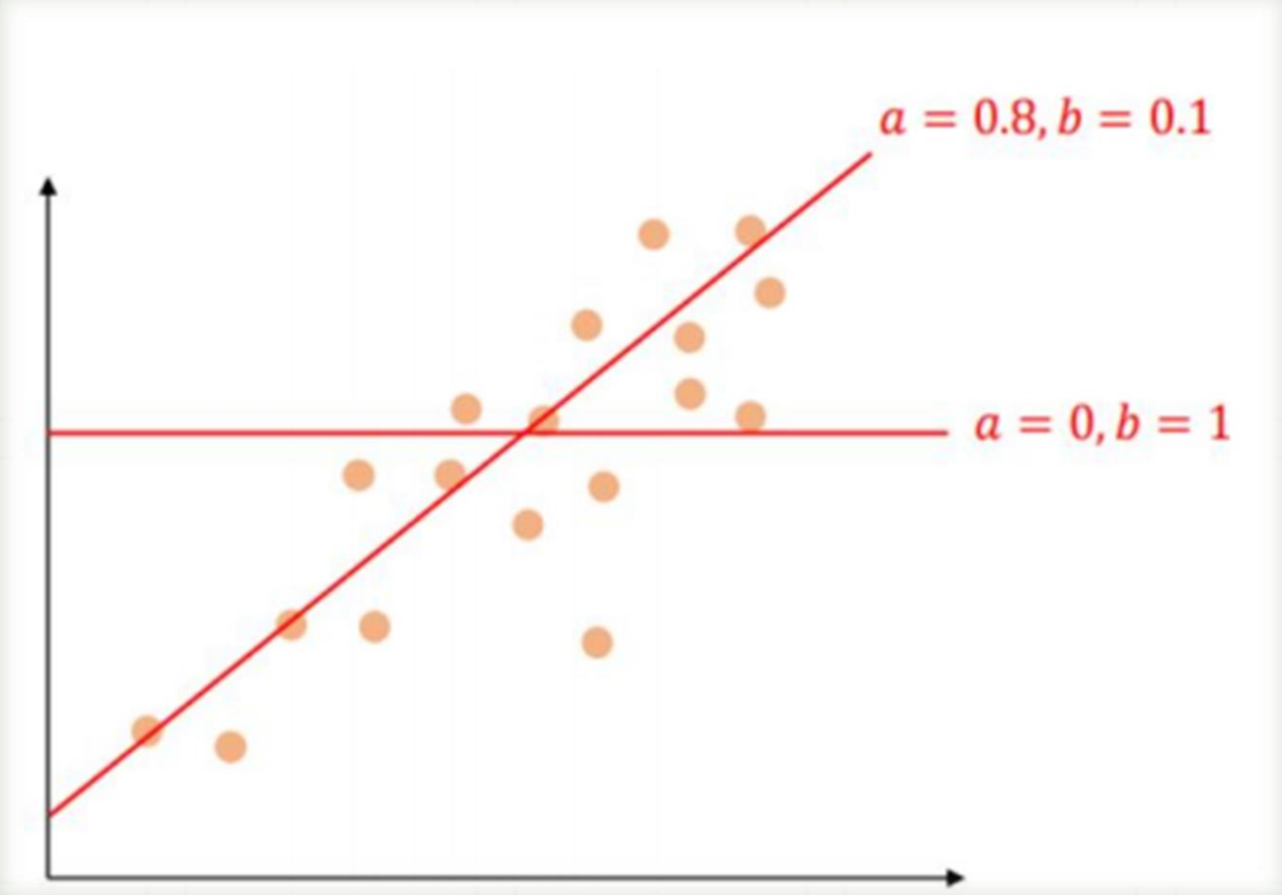
마이너스가 붙은 이유?

$$x_{i+1} : x_i - \alpha \boxed{\frac{\partial f(x)}{\partial x}}$$

기울기 X 학습률

경사하강법은 해당 구간의
최솟값을 찾기 위해서 사용하는 기법이다.
(경사상승법 = 최댓값)

컴퓨터는 어떻게 모델의 오차율을 판단할까?



$y = \underline{a}x + \underline{b}$

a , b = 학습 매개 변수

x = 입력

y = 출력

손실함수

= 알고리즘이 얼마나 잘못하고 있는가를 표현하는 지표

$$E = \frac{1}{2}\sum_{i}(y_i - \tilde{y}_i)^2$$

컴퓨터는 시간이 오래 걸리더라도 결국 답을 찾을 수 있다.

경사하강법 수식 풀이

손실 함수

오차 계산의 편의성을 위해서 제곱을 해준다.

$$\boxed{\frac{1}{n}}\sum(\hat{y}-y)^{\boxed{2}}$$

계산 값이 너무 커지기 때문에 평균을 구한다.

경사하강법 수식 풀이

$$\frac{1}{n}\sum(\hat{y} - y)^2 \quad \left(\hat{y} = wx + b \right)$$

$$= \frac{1}{n}\sum(wx + b - y)^2$$

1. w 대해서 편미분!

2. b 대해서 편미분!

W에 대해서 편미분

$$\frac{\partial}{\partial W}\frac{1}{n}\sum(wx+b-y)^2$$

$$=\frac{1}{N}\sum 2(wx+b-y)\times x$$

X LR *(Learning Rate)*

gdw = LR X MEAN X $(\hat{y}-y)\times x$

b에 대해서 편미분

$$\frac{\partial}{\partial b}\frac{1}{n}\sum(wx+b-y)^2$$

$$=\frac{1}{n}\sum 2(wx+b-y)\times 1$$ X LR *(Learning Rate)*

gdb = LR X MEAN X $(\hat{y}-y)$

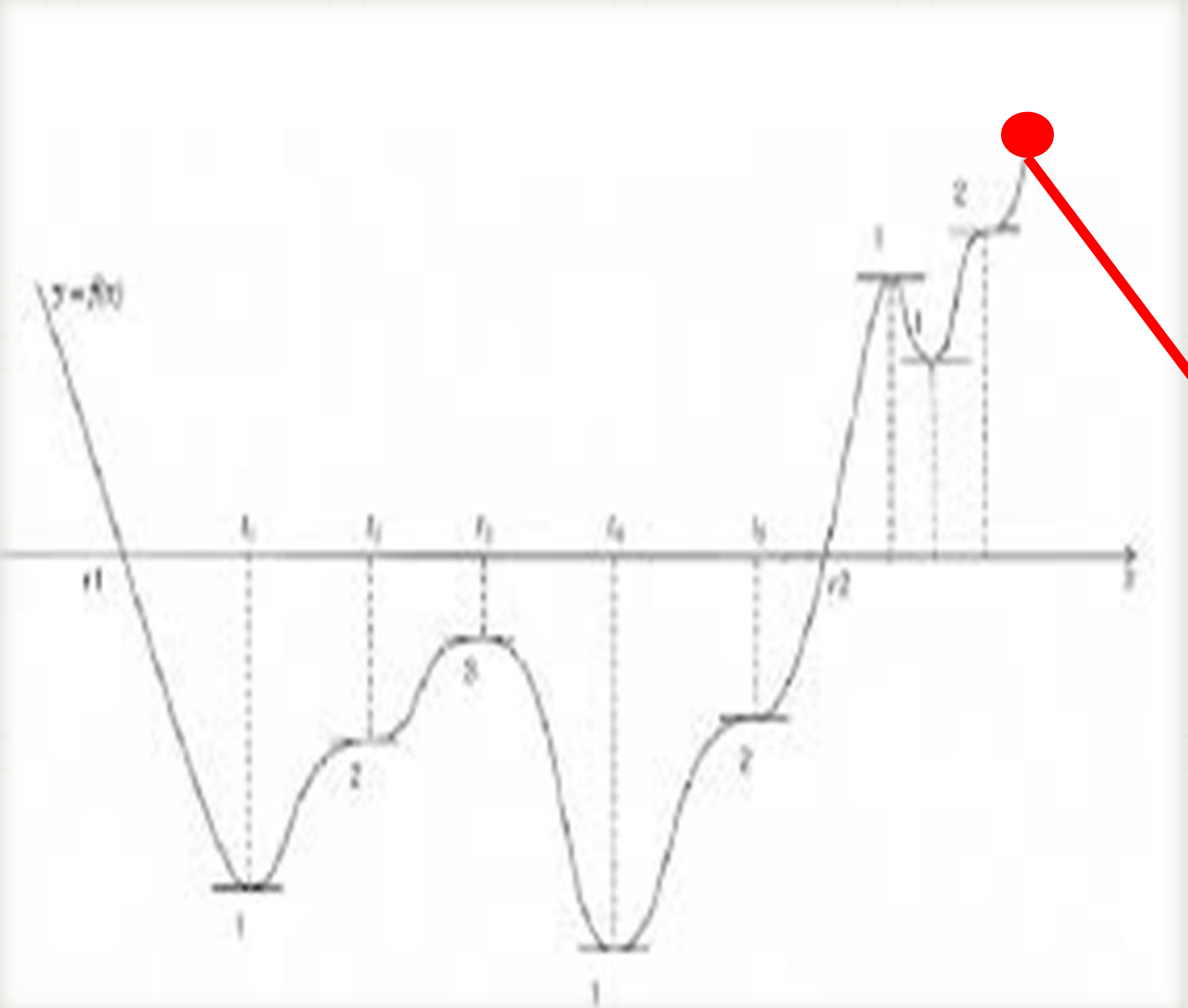
$$\hat{y} = wx + b$$

완성식

$$\hat{y} = (w - gdw)x + (b - gdb)$$

COMPETITION 01

경사하강법을 왜 사용할까?



1. 미분을 하기에는 너무나도 복잡하다.

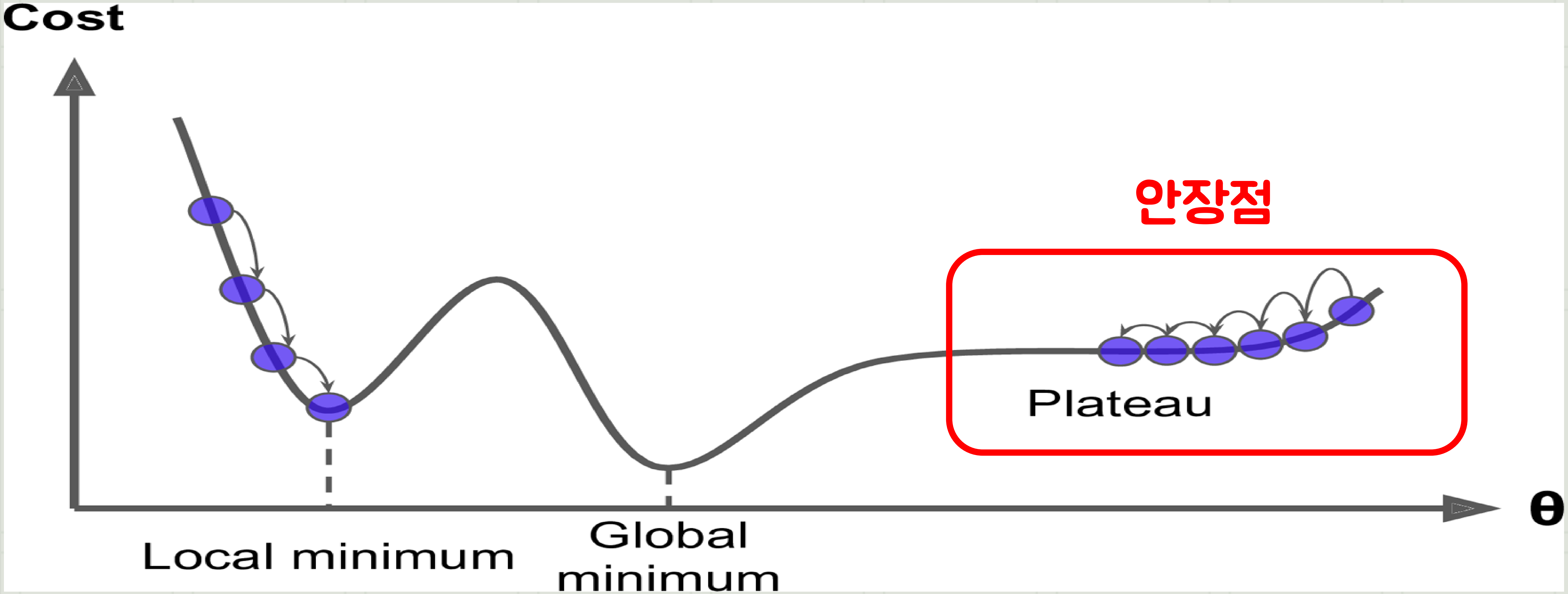
2. 컴퓨터는 lim 개념을 인식하지 못한다.

가장 좋은 점이 아닐까?

X 축을 비용

Y 축을 수익으로 가정

경사하강법의 문제점



기울기가 너무나도 작아서 움직이는 x의 변화량이 미미해진다.

COMPETITION 01

경사하강법의 문제점 해결방안

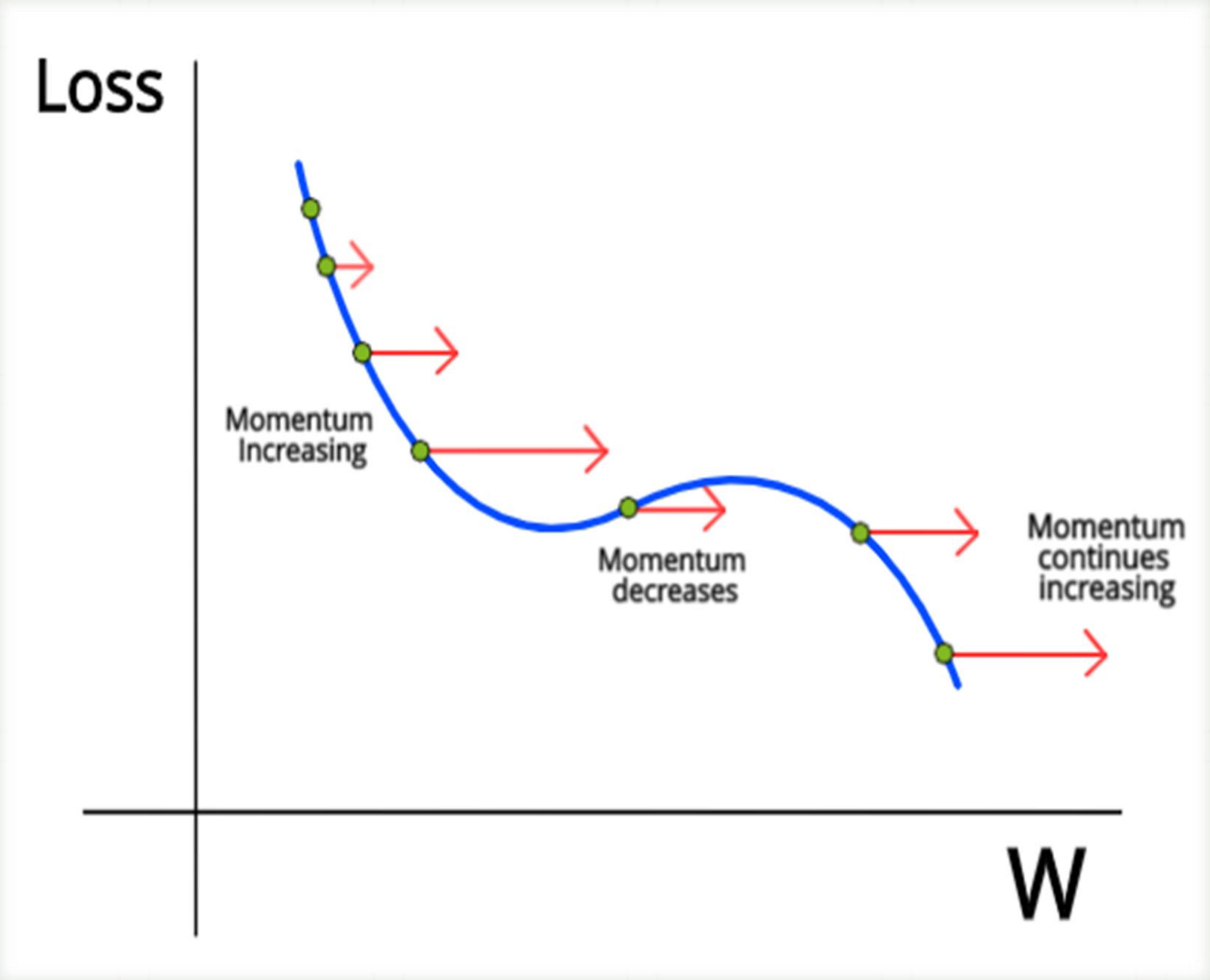
Adam

Momentum

Adaptive Gradient Descent

RMSprop

경사하강법의 문제점 해결방안1 = 모멘텀

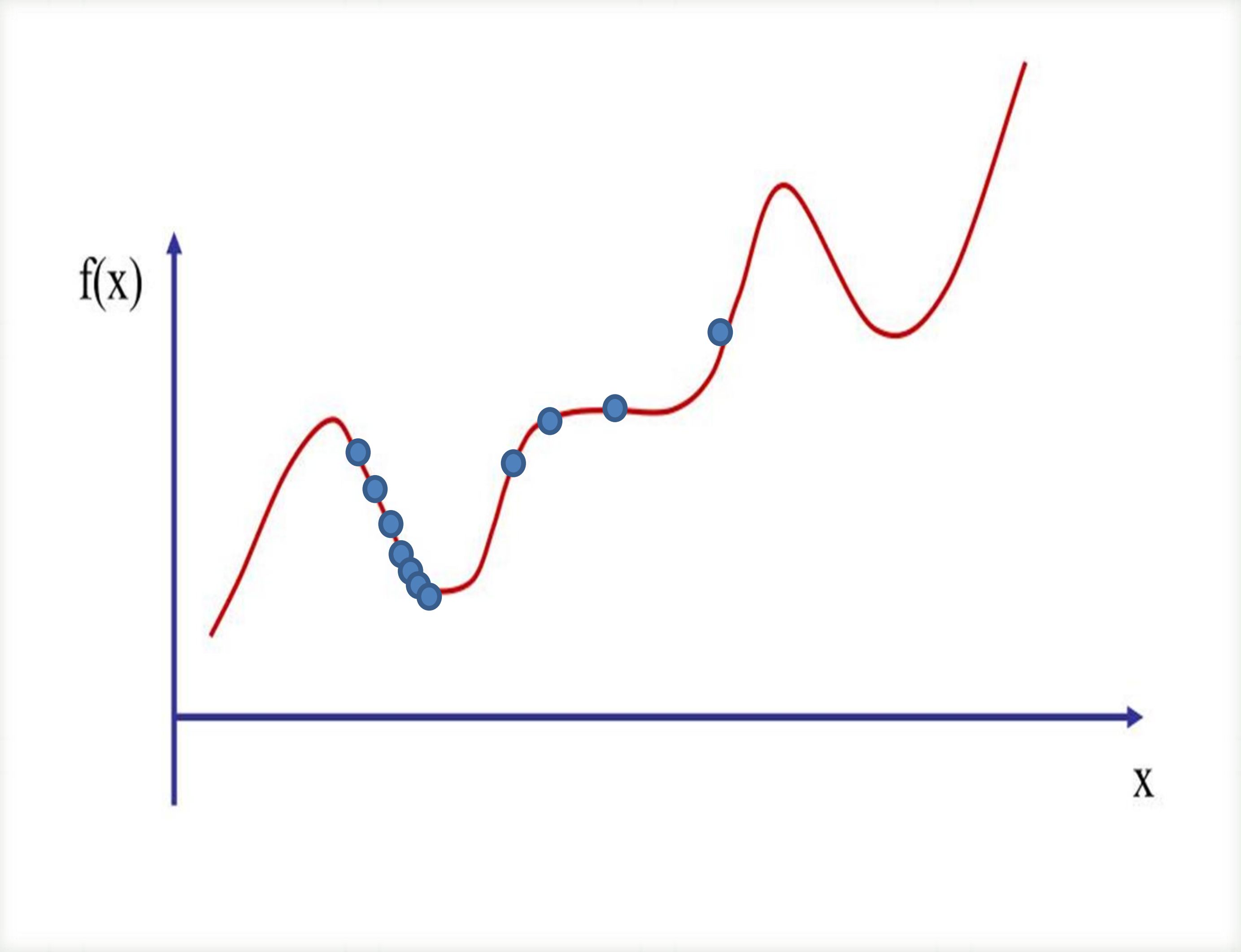


학습률

$$v = \alpha v - \boxed{\gamma} f'(x_1)$$
$$x_2 = x_1 + v$$

α가 제곱으로 계속해서 올라가기 때문에 이동하는 거리가 점점 더 커지게 된다.

경사하강법의 문제점 해결방안2 = 맞춤형 경사하강법



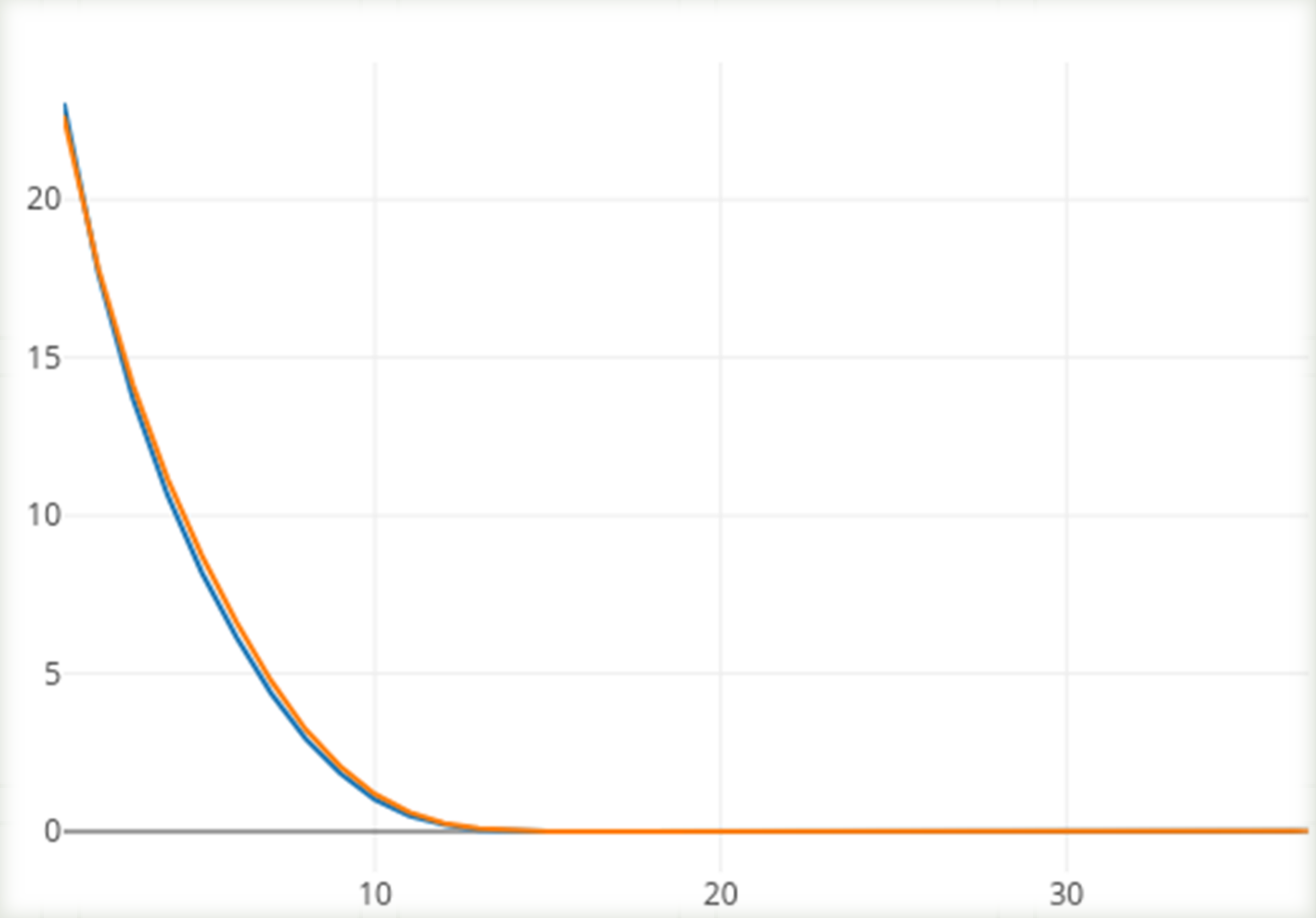
$$h = h + f'(x_1)^2$$

$$x_2 = x_1 - \gamma \frac{1}{\sqrt{h}} f'(x_1)$$

기울기 ▲ h ▲ 학습률 ▼

기울기 ▼ h ▼ 학습률 ▲

경사하강법의 문제점 해결방안3 = RMS Prop



$$h = h\rho + f'(x_1)^2(1-\rho)$$
$$x_2 = x_1 - \gamma \frac{1}{\sqrt{h}} f'(x_1)$$

p(rho) =1 일 경우 기존 데이터 중시
p(rho) =0 일 경우 현재 데이터 중시

COMPETITION 01

GD → Momentum → Adam

GD → Adaptive GD → Adam

Adam

$$v = \alpha v - \gamma f'(x_1)$$
$$h = h\rho + f'(x_1)^2(1-\rho)$$
$$x_2 = x_1 + \alpha v - \gamma \frac{1}{\sqrt{h}} f'(x_1)$$

Thank you for enjoying