Differentiation

Principle of Differention
(미분의 원리)

소프트웨어 꼰대 강의

노기섭 교수

(kafa46@cju.ac.kr)

Course Overview

Topic	Contents
01. Orientation 오리엔테이션	Course introduction, motivations, final objectives 과정 소개, 동기부여, 최종 목표
02. Learning in deeplearning 딥러닝 학습	How does the deeplearing learns knowledge from data 어떻게 딥러닝은 데이터로부터 지식을 배우는가?
03. Principle of differentiation 미분의 원리	Basics of differentiation (concepts, notation, operations) 미분 기본지식 (개념, 표기, 연산)
04. Partial differentiation 편미분	Concept & operation of partial differenciation 편미분 개념, 연산
05. Gradient descent 경사 하강법	Concept, interpretation and learning in gradient descent 경사하강 알고리즘 개념, 해석 및 학습
06. Chain rule 연쇄법칙	Concept & operation of chain rule 연쇄법칙 개념 및 연산
07. Matrix differentiation 행렬미분	Partial differentiation in linear system 선형시스템에서의 편미분
08. Back propagation 역전파 학습	The mechanism of back propagation 역전파 학습의 작동 방법
09. Gradient vanishing 기울기 소실	Quick overview on activation function, cause root of gradient vanishing and its counter-measure 활성함수 간단 소개, 기울기 소실 근본원인과 대책

Note

고등학교, 대학에서 배우는 미분을 일일히 설명하지 않습니다.

주로 아래 내용 위주로 공부할게요 ^^

- 미분 개념
- 대표적인 특징
- 딥러닝에 자주 등장하는 함수의 미분

History

Newton (뉴턴 미분)	Leibniz (라이프니츠 미분)
 자연현상 해석 물리현상, 운동을 설명하기 위한 미분 방법 Isaac Newton (영국, 1643~1727)	미분 이론 정립 자연 현상을 미분 이론을 사용해 해석  Gottfried W. Leibniz (독일, 1646~1716)
변수: $time$	변수: x
$f'(x)$	$\frac{dy}{dx}$ 수학적으로 좀 더 실용적, 효율적임 딥러닝에서 많이 사용

Terms of Differentiation

$$y = f(x)$$

Differentiate: 미분을 하기 위해 잘게 쪼개는 작업

Differential: 차이(difference), 명사로 생각하면 편함

dx: x의 변화량

dy: y의 변화량

$\frac{dy}{dx}$ 차이의 비율

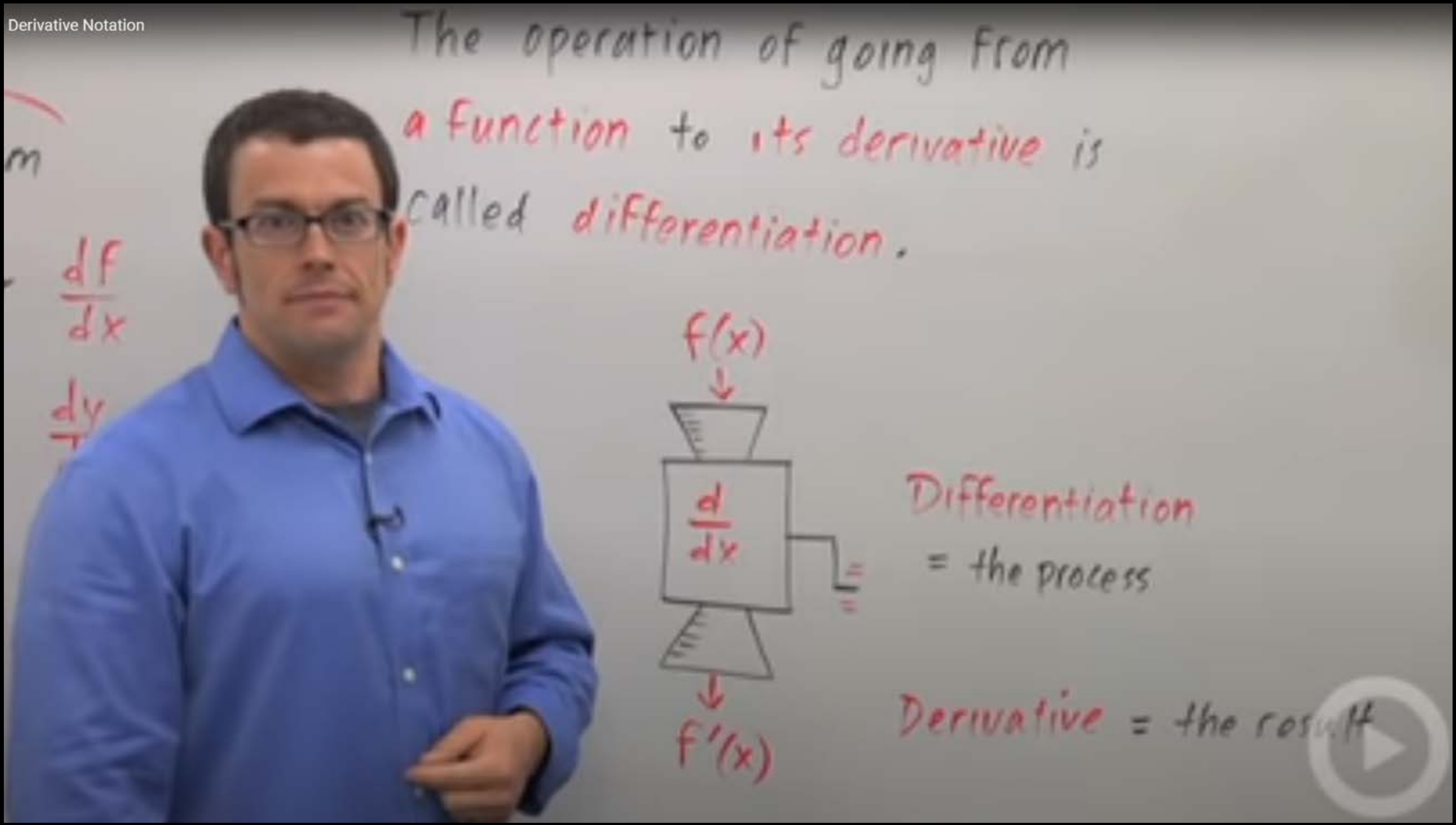
differential of y
with respect to x

Differentiation(미분)
도함수를 구하는 과정
결과는 Derivative

Derivative: Slope of a function at any given point (new function)
(함수의 어떤 점에서 접선의 기울기, 도함수)

$y = x^2 \longrightarrow y' = 2x$

Beautiful Reference in YouTube



이미지 출처: https://youtu.be/3OVbRNEAGVc?si=vzsNbg9xsVi4nsZh

Notation

Newton (뉴턴 미분)	Leibniz (라이프니츠 미분)
$y = f(x)$ $y' = f'(x)$ 매우 불편함 변수가 많을 경우 해석 불가 $y = f(x_1, x_2, \cdots, x_n)$ $y' = ?????$	$\frac{dy}{dx}$ x의 변화에 따른 y의 변화 (발명 당시에는 분수 개념이 맞음) dy, dx 자체가 존재할 수 없음 → 증명 완료 극한의 개념 분수 아님 → 비율도 아님 그러나 여전히 많이 사용됨

연쇄법칙(Chain rule)

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

역함수 미분

$$\frac{dy}{dx} = \frac{1}{\frac{dy}{dx}}$$

대부분 분수와 똑같이 사용

그러나 모든 경우에 분수처럼
사용할 수 있는 것은 아님...

Notation

Δx: "델타 x" 라고 읽음 수 체계에서 존재하는 값 (숫자)

dx: "디 x" 라고 읽음, x와 차이가 0에 가까운 값
(수 체계에서 존재할 수 없는 값)

분수(숫자임)

$$\boxed{\frac{dy}{dx}} = \lim_{\Delta x \to 0} \boxed{\frac{\Delta y}{\Delta x}}$$

극한의 표현, 분수 아님

[참고] 무한소:
0은 아니지만 한없이 0에 가까운 작은 수
(실제로 실수에서 존재할 수 없는 값)

극한(또는 극한값):

독립 변수(x)가 일정한 값(a)에 한없이 가까워질 때,

함수의 값 $f(x)$ 이 한없이 가까워지는 값 L → "L은 $f(x)$의 극한(값)"

- 존재할 수도 있고(미분가능),
- 존재하지 않을 수도 있음(미분 불가)

함수의 극한 (이차함수 예제)

극한의 정의

$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

Δx를 h로 바꿔서
쓰기도 함

[Toy example]

$f(x) = x^2 + 1$

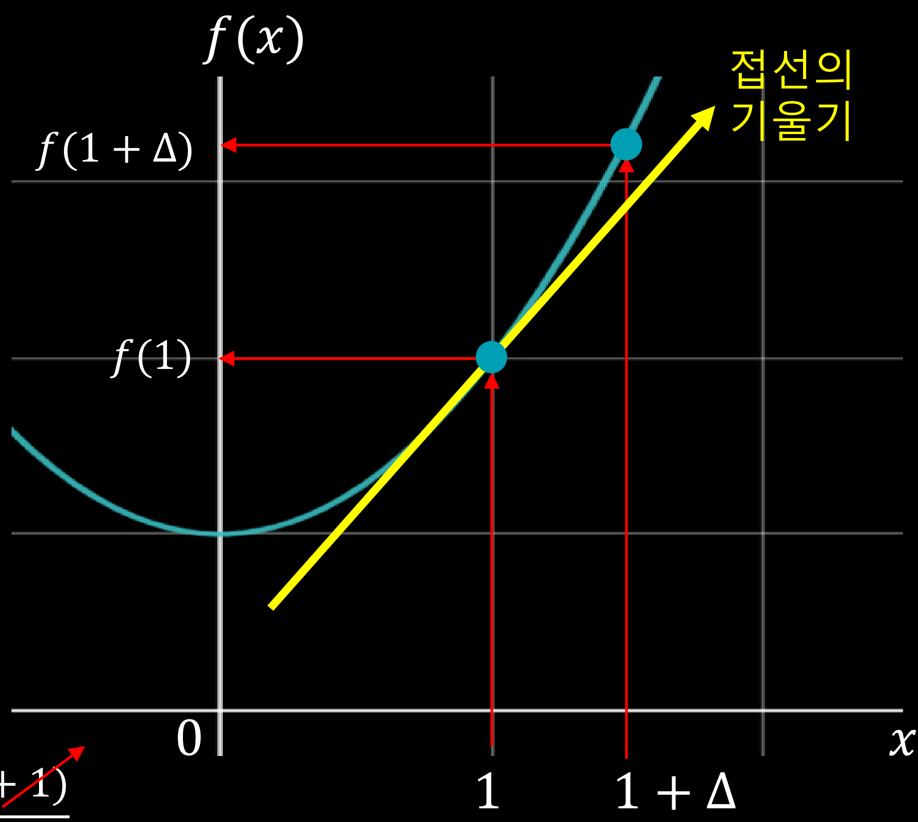
Differential

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{(x+\Delta x)^2 + 1 - (x^2+1)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\cancel{x^2} + 2x\Delta x + \Delta x^2 + \cancel{1} - (\cancel{x^2} + \cancel{1})}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{2x\Delta x + \Delta x^2}{\Delta x} = \lim_{\Delta x \to 0} \frac{\cancel{\Delta x}(2x+\Delta x)}{\cancel{\Delta x}} = \lim_{\Delta x \to 0} 2x + \Delta x$$

$= 2x$ ← Derivative



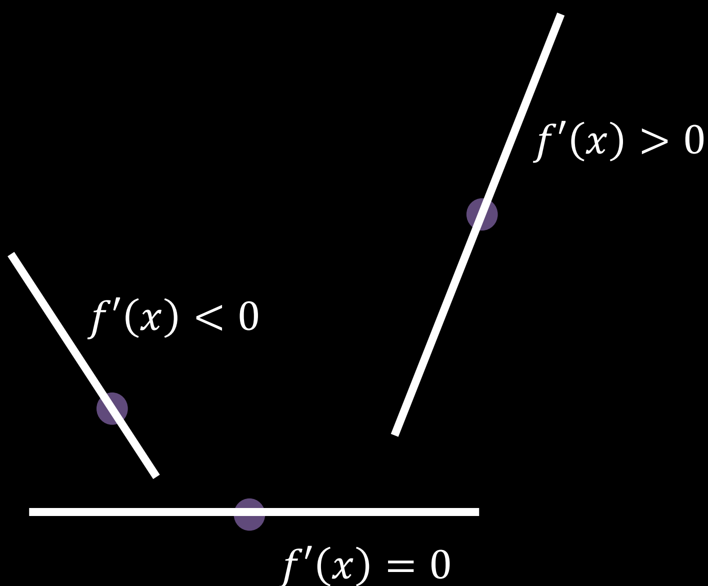
$f'(x) = 2x$

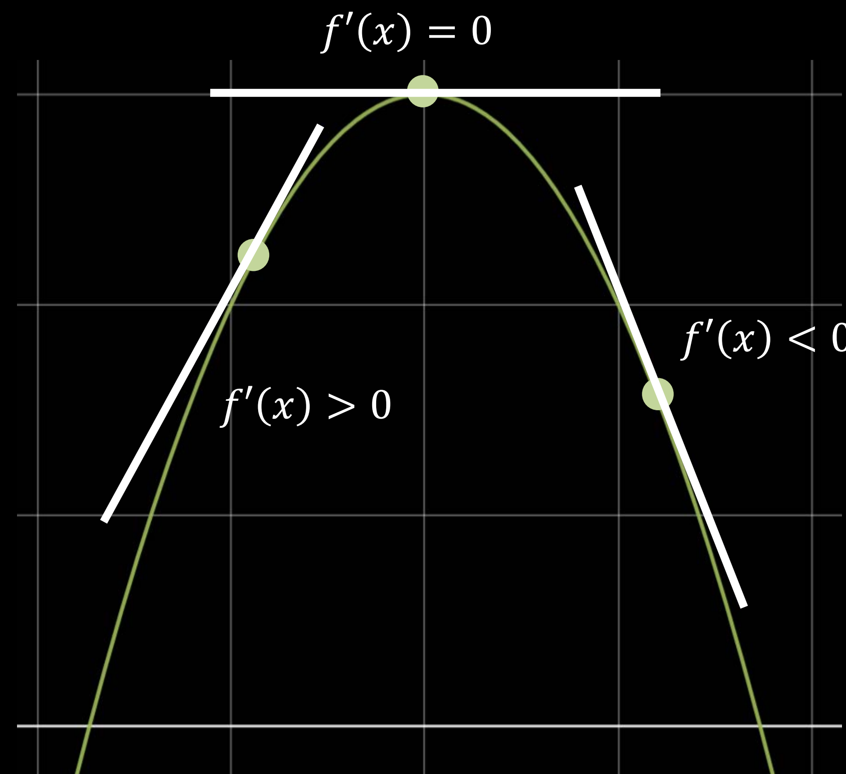
$f'(1) = 2$

미분의 일반화

$Types$	$y = f(x)$	$y' = f'(x) = \frac{dy}{dx}$
상수 미분	$f(x) = c, where\ c\ is\ constant$	$f'(x) = 0$
다항식 미분	$f(x) = x^n$	$f'(x) = nx^{n-1}$
상수배 미분	$f(x) = c \cdot g(x)$	$f'(x) = c \cdot g'(x)$
함수 합 미분	$f(x) = g(x) \pm h(x)$	$f'(x) = f'(x) \pm g'(x)$
함수 곱 미분	$f(x) = g(x) \cdot h(x)$	$f'(x) = g'(x)h(x) + g(x)h'(x)$
합성함수 미분	$f(x) = g \circ h(x) = g(h(x))$	$f'(x) = g'\big(h(x)\big) \cdot h'(x)$ $= g'\big(h(x)\big) \cdot \frac{dh(x)}{dx} = \frac{dg}{dh} \cdot \frac{dh}{dx}$

극대, 극소값





딥러닝에서 자주 사용하는 미분 (1/2) - 지수함수의 미분

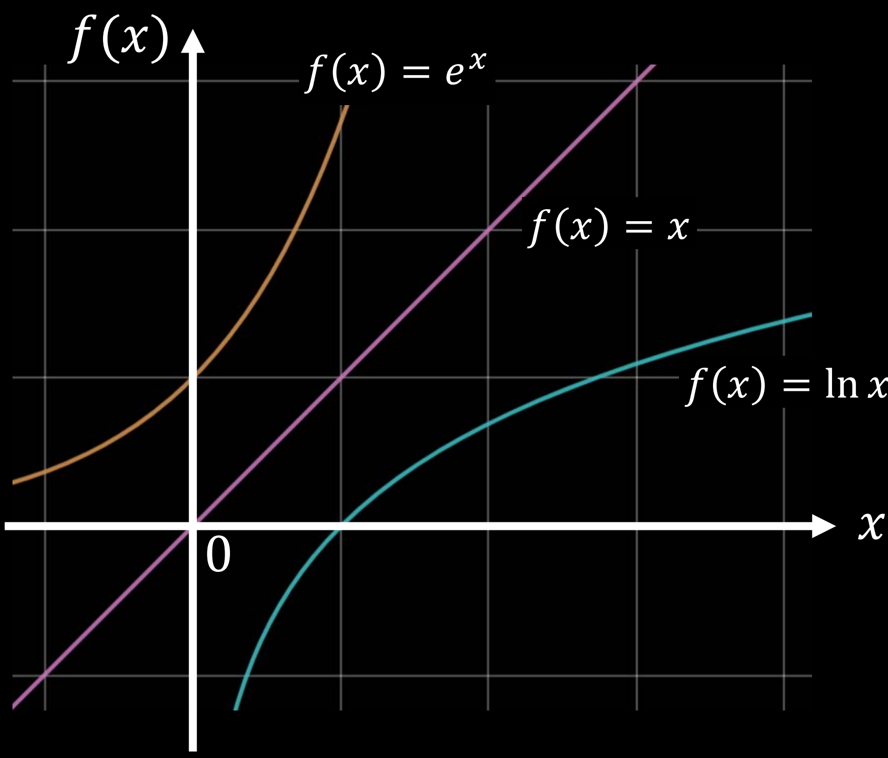
$$f(x) = e^x$$

$$f'(x) = \lim_{\Delta x \to 0} \frac{e^{x+\Delta x} - e^x}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{e^{x+\Delta x} - e^x}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{e^x(e^{\Delta x} - 1)}{\Delta x}$$

$$= e^x \lim_{\Delta x \to 0} \frac{e^{\Delta x} - 1}{\Delta x}$$



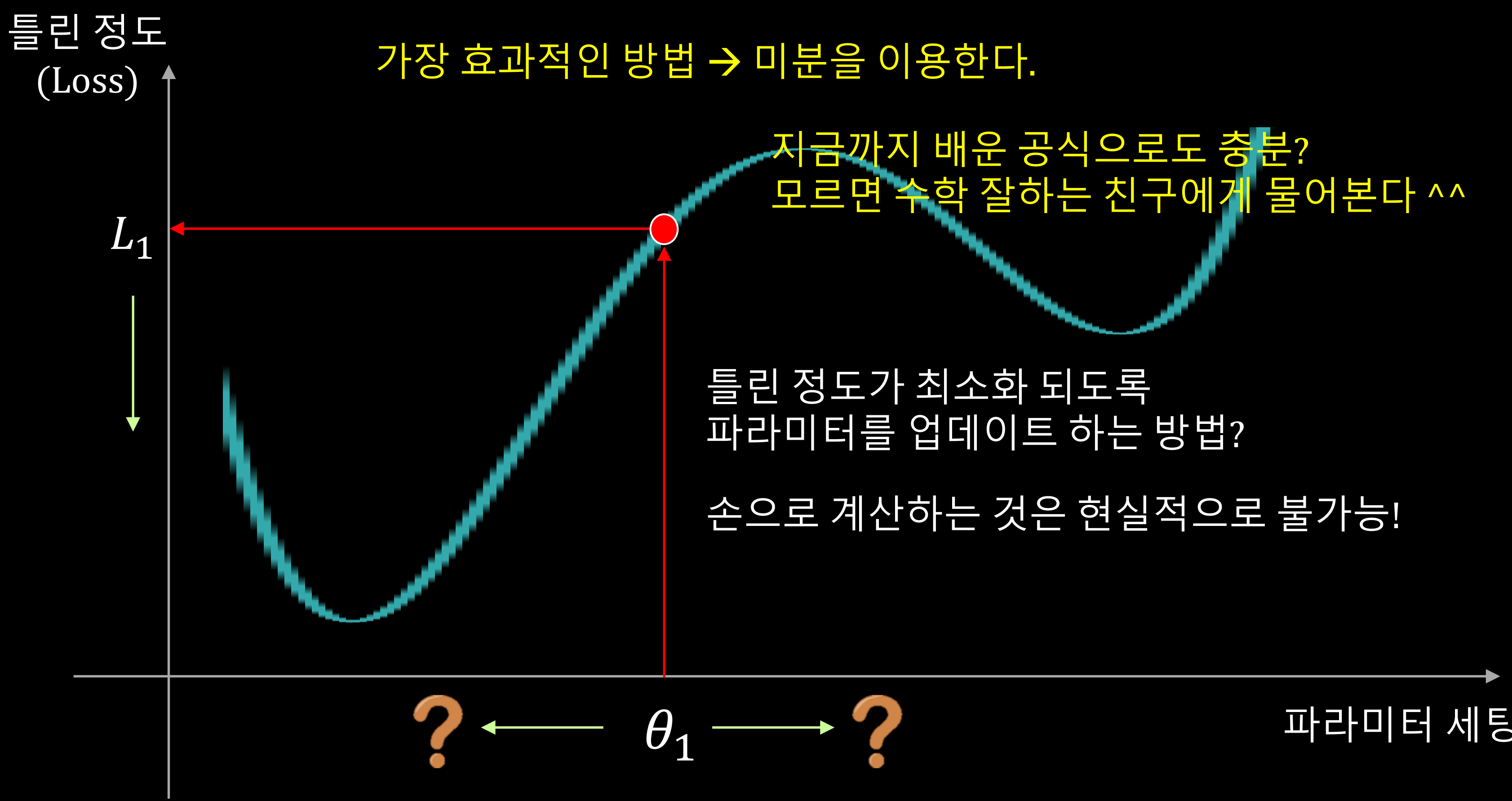
$$let\ t = e^{\Delta x} - 1 \longrightarrow \Delta x = \ln(t+1)$$

$$= e^x \lim_{t \to 0} \frac{t}{\ln(t+1)} = e^x \lim_{t \to 0} \frac{1}{\frac{1}{t}\ln(t+1)} = e^x \frac{1}{\ln(\lim_{t \to 0}(1+t)^{\frac{1}{t}})} = e^x \frac{1}{\ln e} = e^x$$

딥러닝에서 자주 사용하는 미분 (2/2)

$y = f(x)$	$y' = f'(x) = \frac{dy}{dx}$
$f(x) = e^x$	$f'(x) = e^x$
$f(x) = a^x$	$f'(x) = a^x \ln a$
$f(x) = \ln x$	$f'(x) = \frac{1}{x}$
$f(x) = \log_a x$	$f'(x) = \frac{1}{x \ln a}$

마지막 복습 (Final Recap): 딥러닝 어디에 사용하는가?



가장 효과적인 방법 → 미분을 이용한다.

지금까지 배운 공식으로도 충분?
모르면 수학 잘하는 친구에게 물어본다 ^^

틀린 정도가 최소화 되도록
파라미터를 업데이트 하는 방법?

손으로 계산하는 것은 현실적으로 불가능!

어느 방향으로 얼마나 이동해야 하지?

수고하셨습니다 ..^^..