Partial differential for Deep Learning

딥러닝을 위한 편미분

머신러닝/딥러닝

임 경 태

Week	Chapter	Contents
1	1, 2장	강의 소개, 파이썬 복습
2	1, 3장	파이썬 복습, Numpy, Pandas
3	1, 4장	딥러닝을 위한 미분
4	5장	회귀
5	5장	분류
6	6장	XOR문제
7	7장	딥러닝
8	1~7장	중간고사
9	8장	MNIST 필기체 구현 (팀 프로젝트)
10	9장	오차역전파
11	11장	합성곱 신경망(CNN)
12	12장	순환 신경망(RNN)
13	10장	자율주행 (Collision Avoidanve, Transfer Learning)
14	11장	자율주행 (Load Following)
15	8~12장	기말고사 (or 프로젝트 발표)

CONTENTS

목적 : 딥러닝을 위한 미분 학습

목표 : 편미분, 수치 미분의 개념과 필요한 이유 학습과 파이썬 구현

내용 : 미분의 개념, 편미분, 수치 미분, 연쇄법칙

CONTENTS

미분 개념 – 평균 변화율

평균변화율

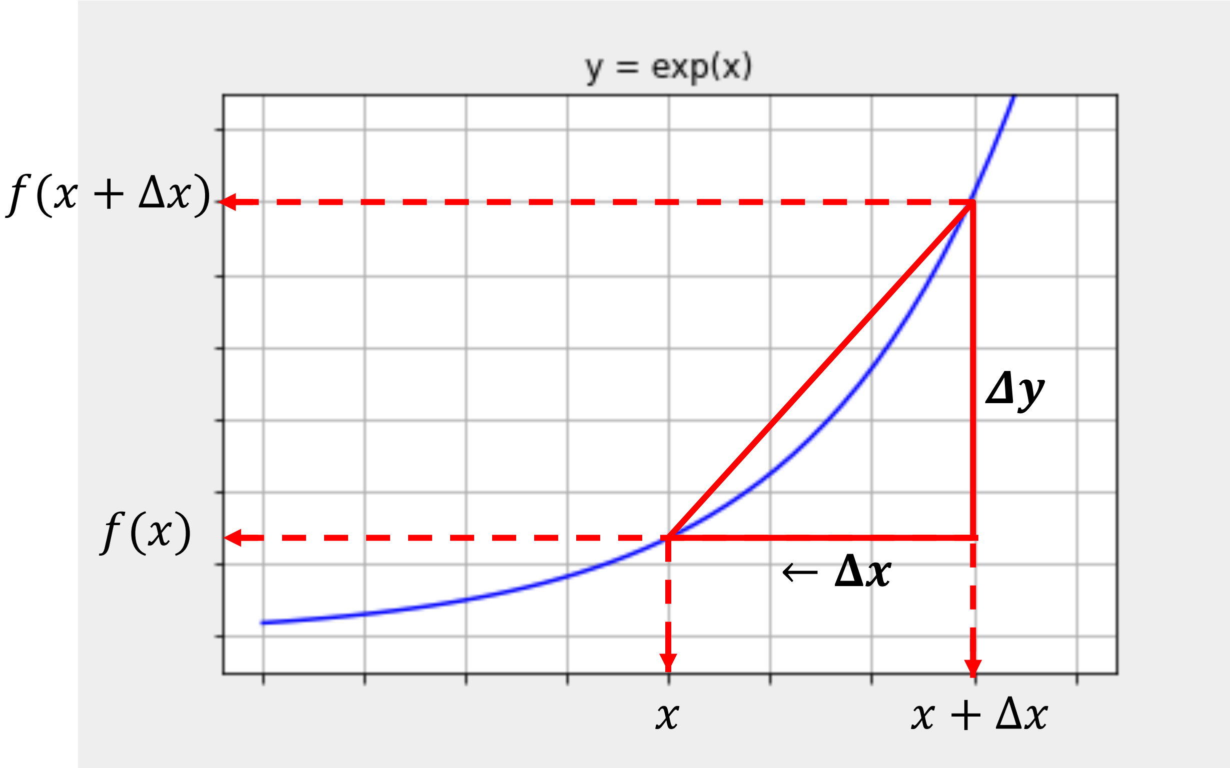
- 구간에 대한 평균적인 변화율

- $\frac{\Delta y}{\Delta x} = \frac{f(x+\Delta x)-f(x)}{\Delta x}$

- 평균적인 기울기를 나타냄

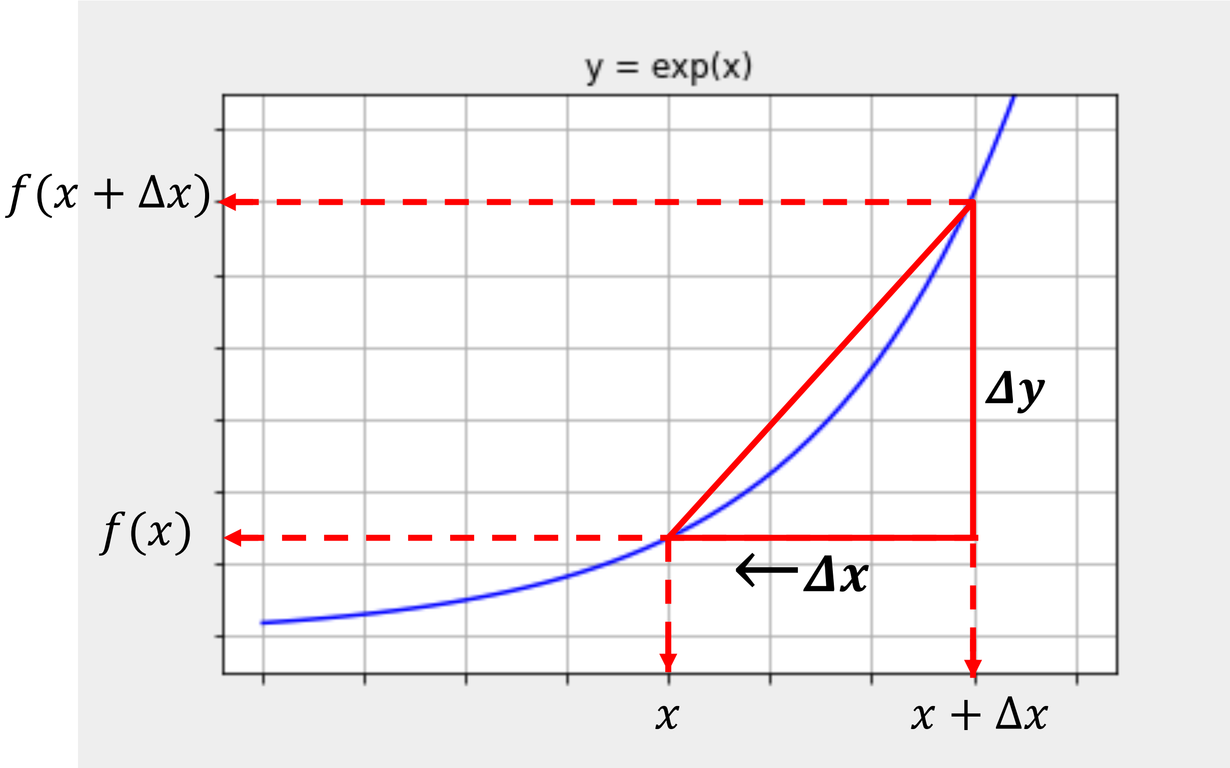
- "평균"이기 때문에 정확하지 않은 정보

→ 순간 변화율 (미분계수)



순간 변화율

- 평균 변화율에서 구간(Δx)을 0^+로 보낸 값
- $\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x}$
- 해석적인 의미로 x 에서 접선의 기울기를 의미함.
- 순간 변화율을 미분계수로 정의함.



미분의 의미

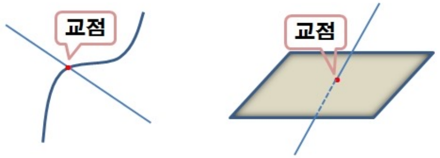
- 입력 x를 현재 값에서 아주 조금 변화(Δx)시키면 함수 $f(x)$의 값이 변하는 정도를 나타냄.

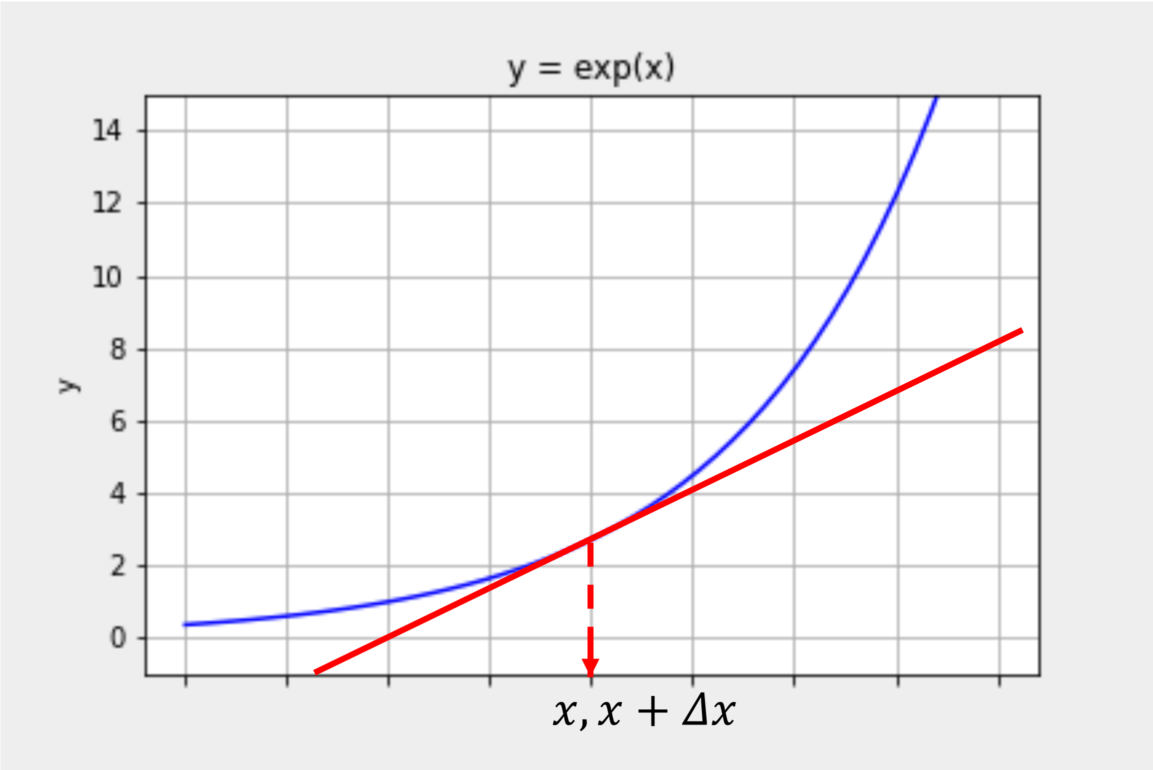
- $f'(x) = \frac{d}{dx} f(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$

- 점 x에서 접선의 기울기를 의미함.

교점

선과 선 또는 선과 면이 만나서 생기는 점





머신 러닝/ 딥러닝의 최종 목표?

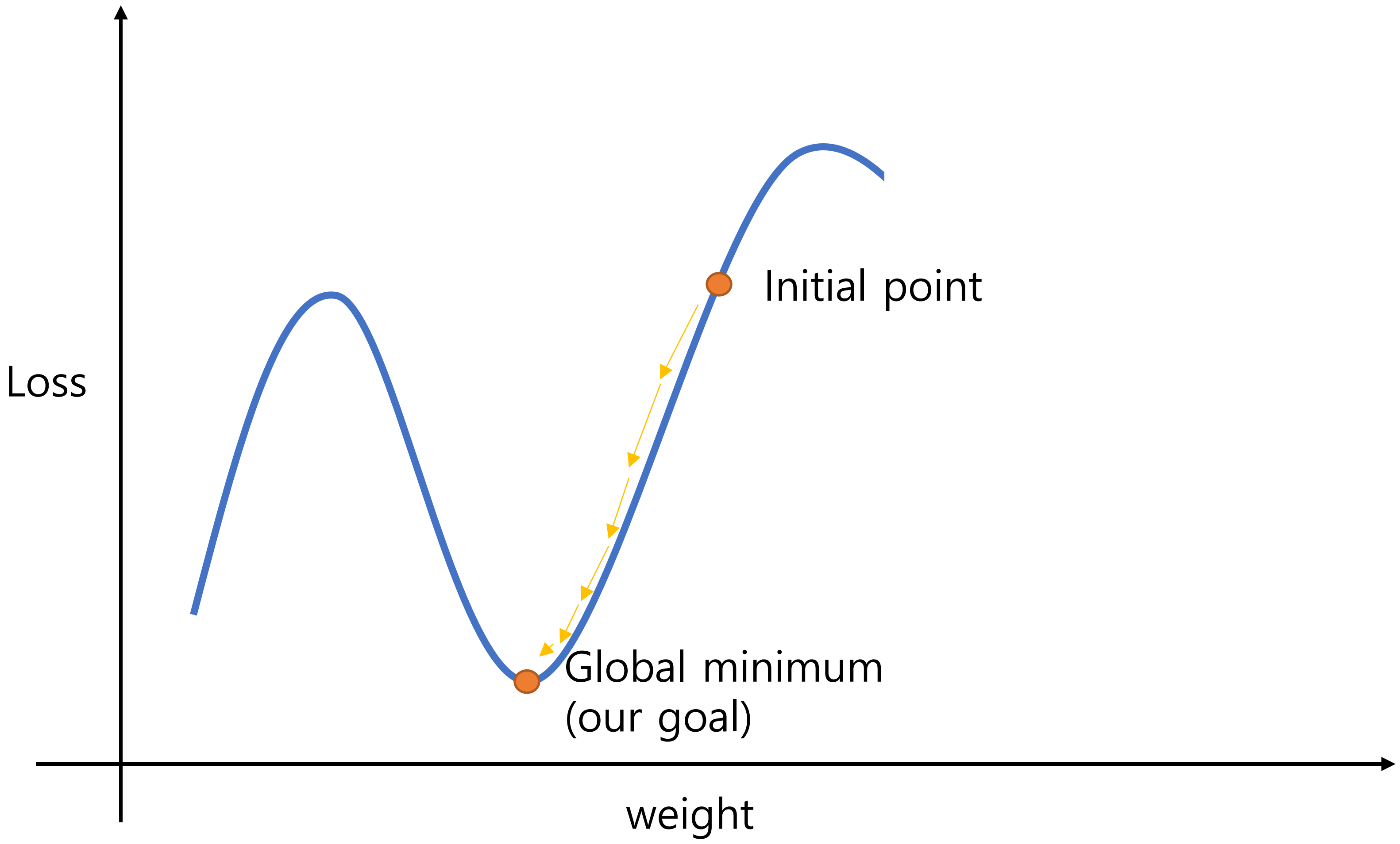
- 손실(loss, cost)을 최소화 하여 가중치의 최적값을 자동으로 찾는 것.

미분이 필요한 이유?

- 가중치를 조금 변화시켰을 때(Δx) 손실함수의 **변화량**을 확인하고 **최적값**(손실을 최소로 하는 가중치)과 가까워지도록 **가중치를 갱신**해야 함.

- 즉, 학습은 **미분**을 통해 이루어진다.

딥러닝에 미분이 필요한 이유 – Gradient Descent 예제



CONTENTS

다변함수?

- 입력으로 받는 값이 한 개 이상으로 이루어진 함수

$$z = f(x) \qquad\qquad z = f(x, y)$$

딥러닝에 편미분이 필요한 이유

- 다변함수의 매개변수(가중치) 개수는 모델마다 다르지만 수백,수천만 가지일 수도 있다.
- 위에서 알아본 미분으로는 해결 불가능.
- 따라서, 각 변수의 변화에 따른 가중치의 변화율을 알기 위해 편미분 사용.

편미분 (Partial derivative)?

- 다변수 함수에서 각 변수에 따른 함수 값의 변화율
- ∂ : “partial”이라고 읽음 (round, del, dee, partial dee라고도 읽음)
- 미분하지 않는 나머지 변수는 상수로 취급한 후 미분

미분 vs 편미분 (Partial derivative)?

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

  e.g., $f(x) = 2x^2 + 4e^x + 15$, $f'(x) = \frac{d}{dx} f(x) = 4x + 4e^x$

- $f_x(x, y) = \frac{\partial}{\partial x} f(x, y) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x, y) - f(x, y)}{\Delta x}$

  e.g., $f(x, y) = 2x^2 + 4xy + y^2 + 15$, $f_x(x, y) = \frac{\partial}{\partial x} f(x, y) = 4x + 4y$

편미분 개요

$$z = f(x, y)$$

- $f_x(x, y)$: $f(x, y)$를 x로 편미분
- $f_y(x, y)$: $f(x, y)$를 y로 편미분

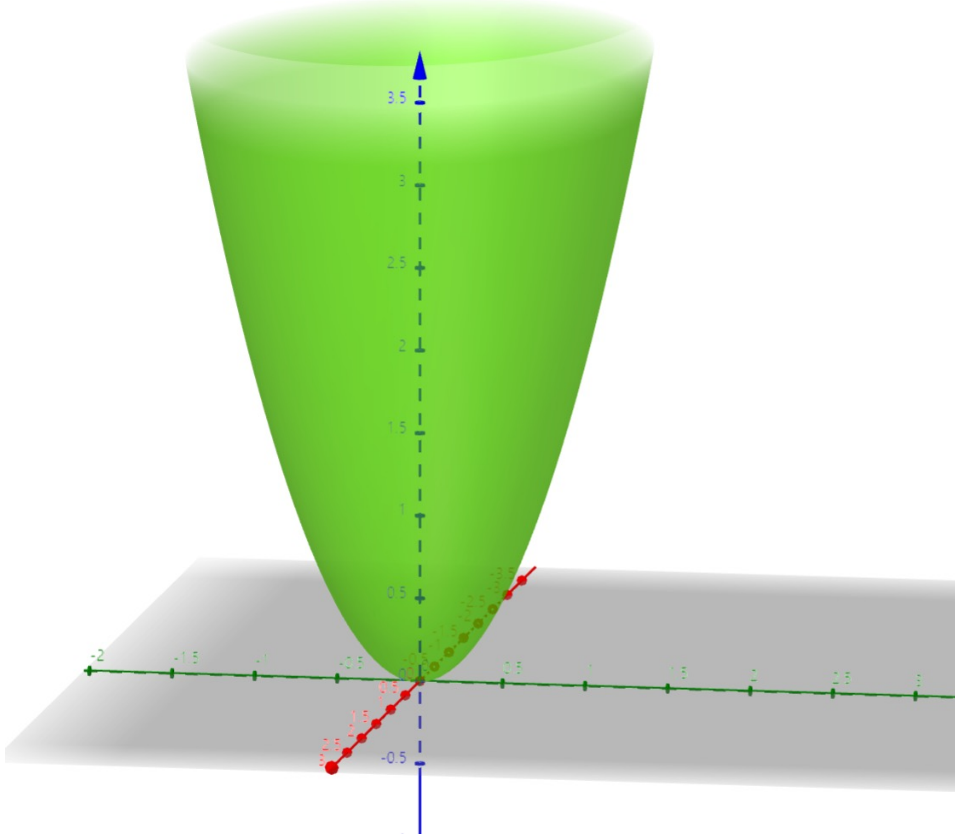
$$f_x(x, y) = \frac{\partial}{\partial x} f(x, y) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x, y) - f(x, y)}{\Delta x}$$

$$f_y(x, y) = \frac{\partial}{\partial y} f(x, y) = \lim_{\Delta x \to 0} \frac{f(x, y+\Delta x) - f(x, y)}{\Delta x}$$

편미분 해석적 의미

$$z = f(x, y) = x^2 + 2y^2$$

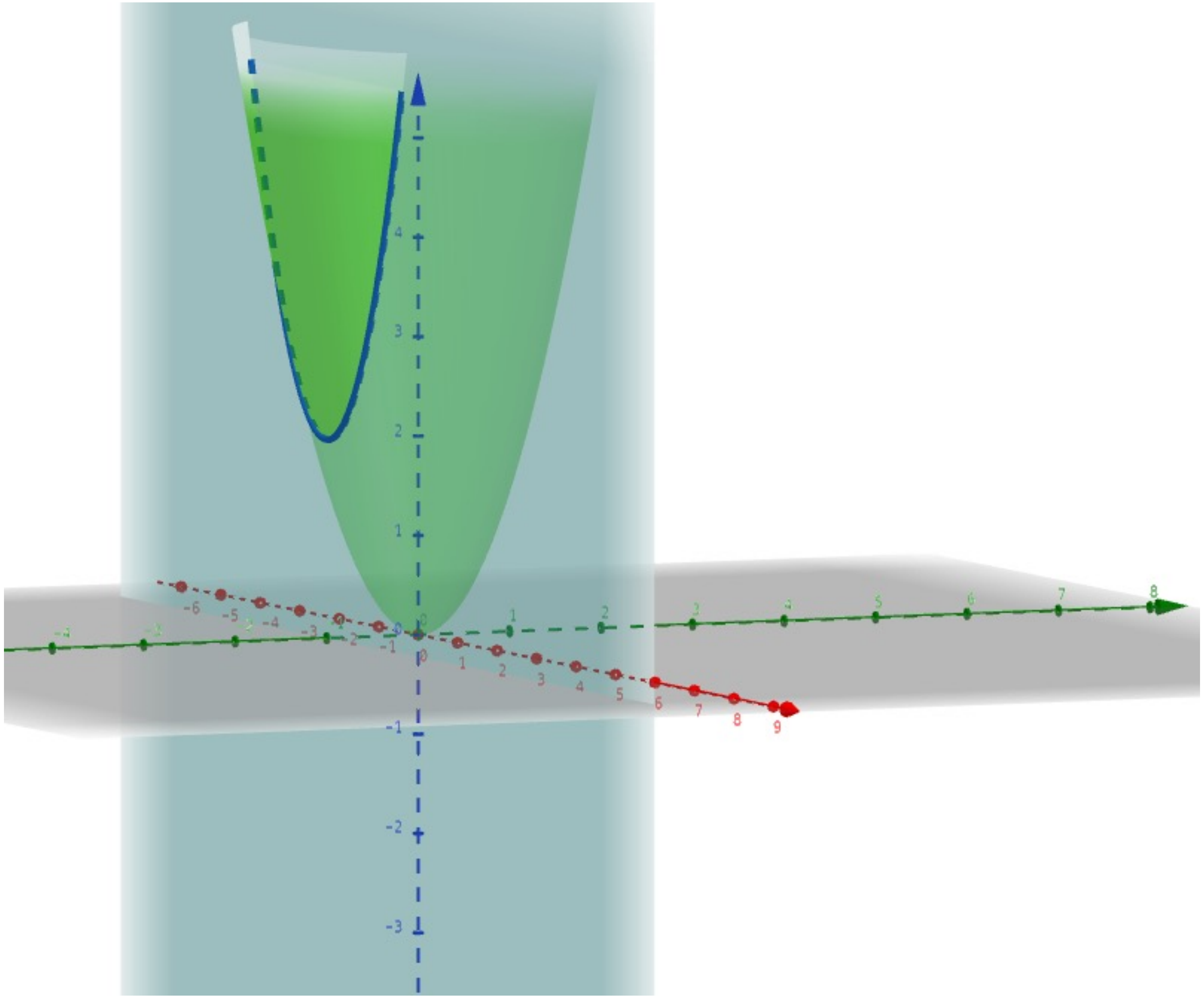
$$f_x(2, -1) = ?$$



편미분 해석적 의미

$$z = f(x, y) = x^2 + 2y^2$$

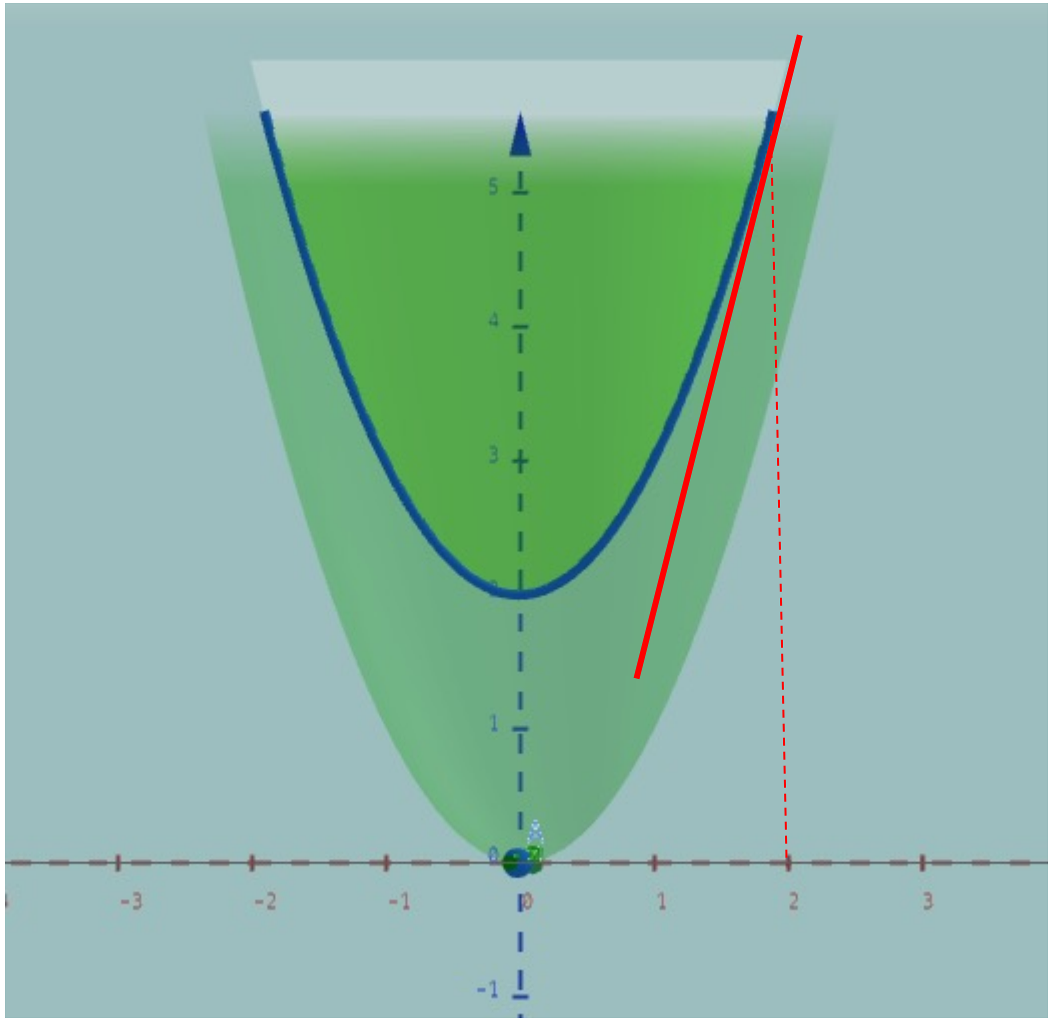
$$f_x(2, -1) = ?$$



편미분 해석적 의미

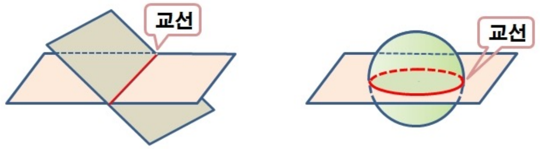
$$f_x(2,1) = ?$$

$f(x, y)$와 $y = 1$의 교선의 $x = 2$에서 접선의 기울기

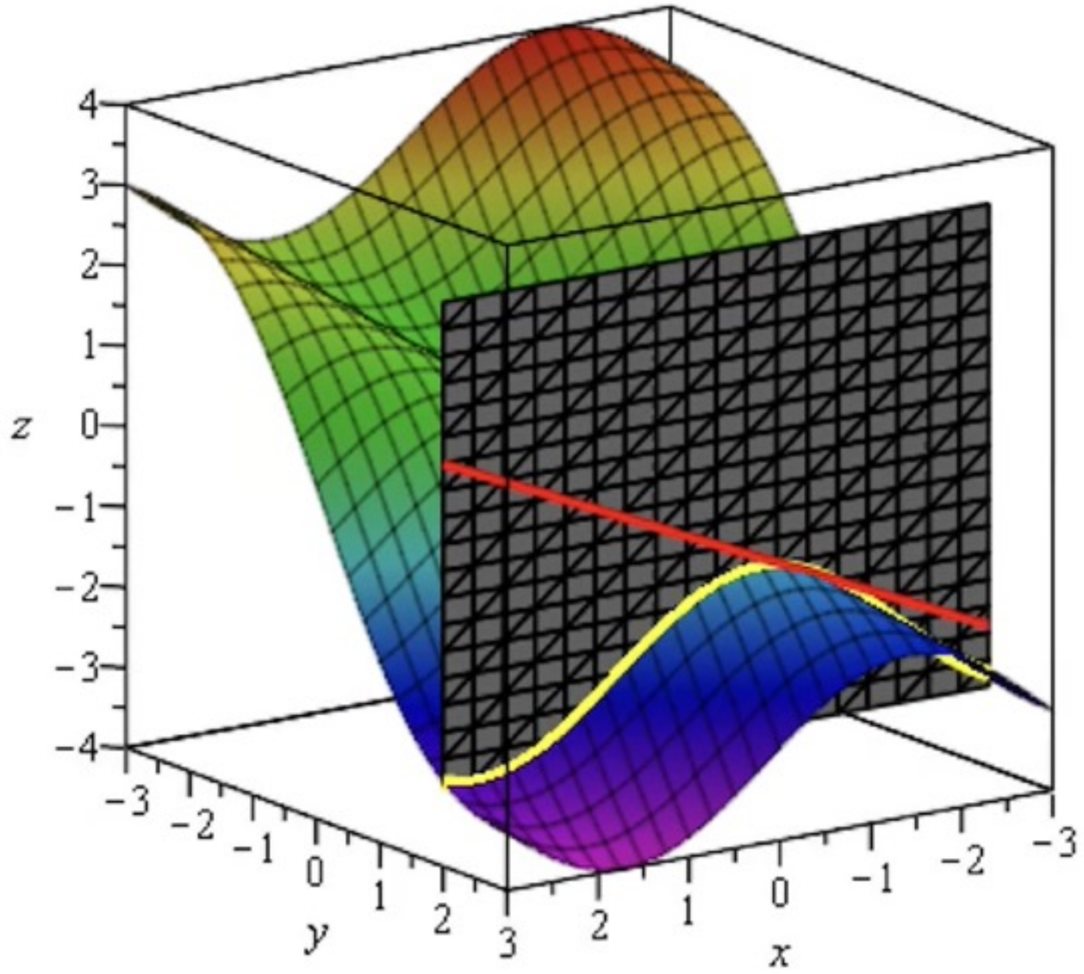


- **교선**

  면과 면이 만나서 생기는 선

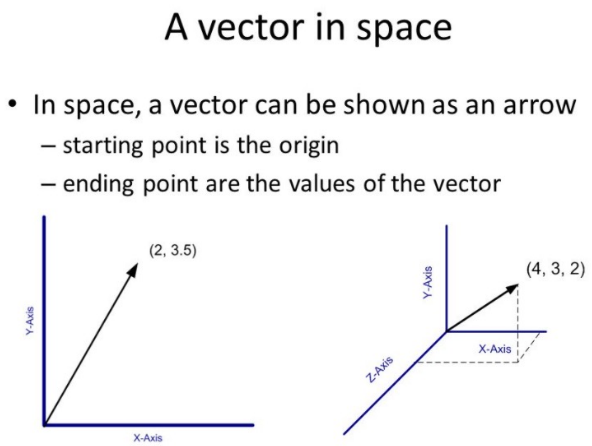


편미분 해석적 의미



$$f_x(-1, 2) = ?$$

Gradient



- ∇ : “**gradient vector**” or “nabla” or “del”
- 기울기 벡터

$\mathbb{R}^3$에서 gradient연산자는 $\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}$ 로 정의된다.

($\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ 는 $\mathbb{R}^3$ 단위 직교 기저)

→ $\nabla f(x, y, z) = \left\langle \frac{\partial f(x,y,z)}{\partial x}, \frac{\partial f(x,y,z)}{\partial y}, \frac{\partial f(x,y,z)}{\partial z} \right\rangle$

Gradient

- Gradient란 최대의 증가율을 나타내는 벡터.

- Gradient가 최대의 증가율을 나타내기 때문에 Gradient의 반대 방향을 따라가면 Local minimum에 가장 빠르게 도달한다.

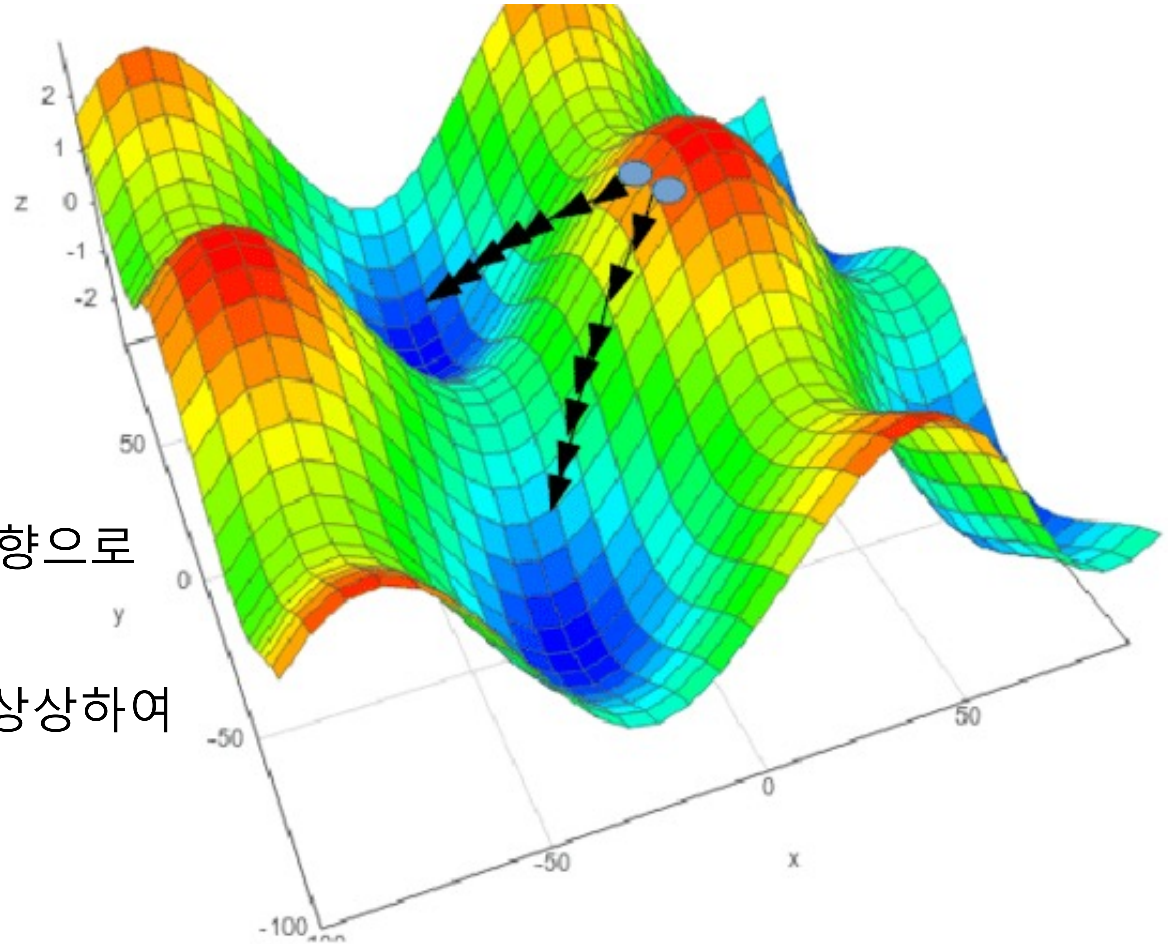
- 즉 벡터 $\nabla f(x, y, z) = \left\langle \frac{\partial f(x,y,z)}{\partial x}, \frac{\partial f(x,y,z)}{\partial y}, \frac{\partial f(x,y,z)}{\partial z} \right\rangle$ 는 $f(x, y, z)$에서 가장 가파른 방향을 나타냄.

Gradient

- $\mathbb{R}^3$에서의 Gradient descent

- 딥러닝에서 가중치의 수가 차원을 의미하는데
보통 3차원 이상이다.
하지만 3차원 이상은 상상할 수 없다.
n차원 에서도 3차원과 같이 Gradient의 반대 방향으로
학습을 하면 Local minimum에 도달한다.
우리가 상상할 수 있는 $\mathbb{R}^3$ 에서의 Gradient 를 상상하여
$\mathbb{R}^n$ 에서도 같은 결과가 나온다고 생각하자.

화살표 : Gradient 반대 방향



이미지 출처 https://www.commonlounge.com/discussion/f5e5b0b3bba44e5daadb93044e8fa648/history

CONTENTS

연쇄 법칙 Chain rule

- Chain rule은 합성함수 미분에 사용됨.

- Let $y = f\left(g(x)\right),\ t = g(x)$
- $\dfrac{dy}{dx} = \dfrac{d}{dx} f\left(g(x)\right) = f'\left(g(x)\right)g'(x) = \dfrac{dy}{dt} \times \dfrac{dt}{dx}$

$$\therefore \frac{dy}{dx} = \frac{dy}{\cancel{dt}} \times \frac{\cancel{dt}}{dx}$$

- 딥러닝에서는 오차역전파를 이해하는데 필요함.

연쇄 법칙 Chain rule

- Let $y = f\big(g(x)\big),\ t = g(x)$
- $\frac{dy}{dx} = \frac{d}{dx} f\big(g(x)\big) = f'\big(g(x)\big)g'(x) = \frac{dy}{dt} \times \frac{dt}{dx}$

$$\therefore \frac{dy}{dx} = \frac{dy}{\cancel{dt}} \times \frac{\cancel{dt}}{dx}$$

$$when\ f(x) = e^{4x^2}, let\ 4x^2 = t, then\ f(x) = e^t$$

$$\frac{df}{dx} = \frac{df}{dt} * \frac{dt}{dx} = \frac{d(e^t)}{dt} * \frac{d(4x^2)}{dx}$$

CONTENTS

수치 미분이란?

- 함수의 미분이 복잡하거나 계산에 비용이 많이 드는 경우 미분계수의 근사값을 얻기 위해 사용하는 미분 법.

- 전방 차분(forward scheme)
- 후방 차분(backward scheme)
- 중앙 차분(central scheme)

의 세가지 방법이 존재

수치 미분

- 전방 차분(forward scheme)

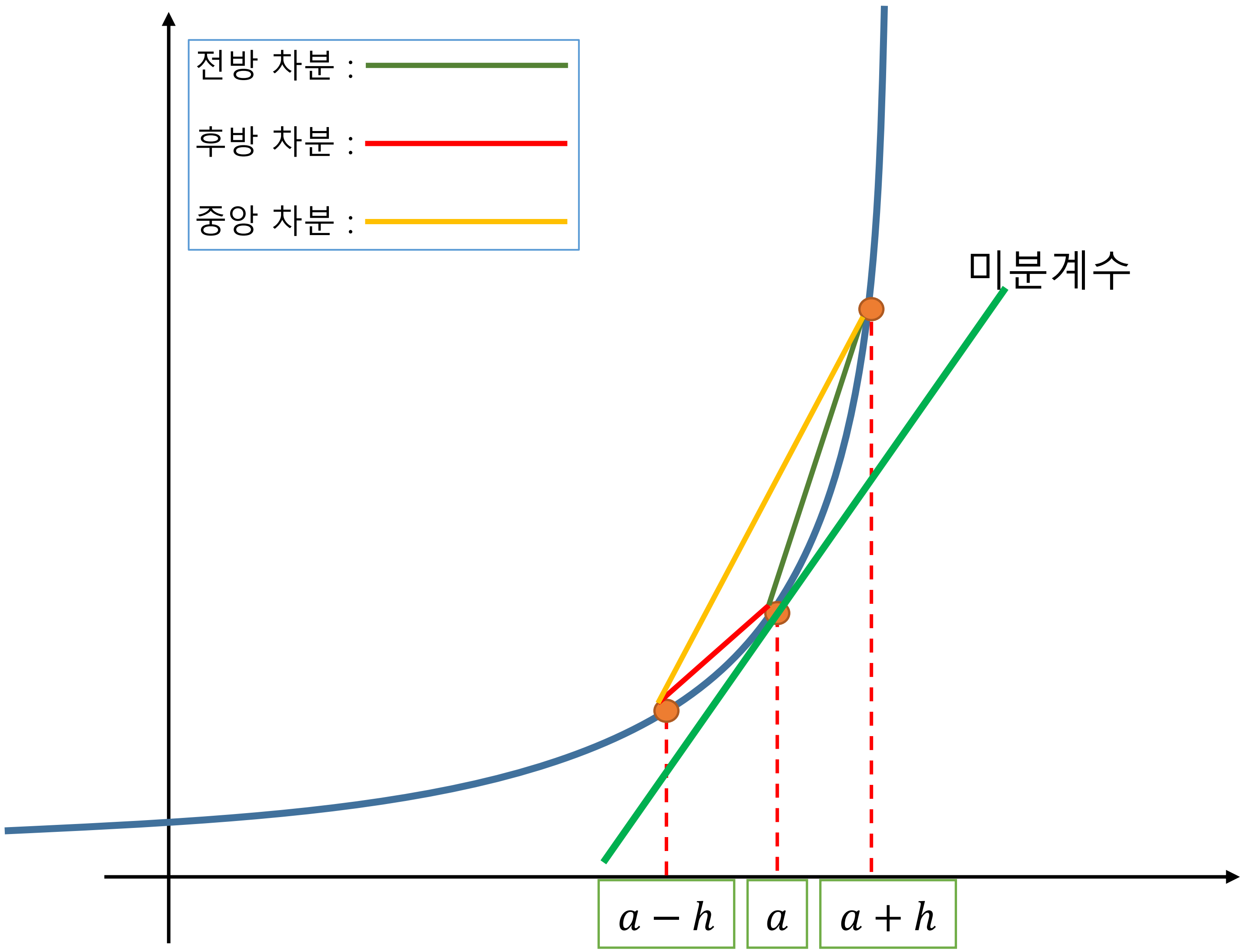
$$\lim_{h \to 0} \frac{f(a+h)-f(a)}{h}$$

- 후방 차분(backward scheme)

$$\lim_{h \to 0} \frac{f(a)-f(a-h)}{h}$$

- 중앙 차분(central scheme)

$$\lim_{h \to 0} \frac{f(a+h)-f(a-h)}{2h}$$



수치 미분

- 중앙 차분이 미분계수에 가장 근접하다. (오차가 작다)

→ 앞으로 수치 미분에는 중앙 차분 사용

수치 미분 – 중앙 차분 오차 증명 (생략 가능)

- For sufficiently small h, h>0

1) 전방 차분 (후방 차분도 같은 방법으로 증명)

By Tayler series,

$f(x) = f(a) + f'(a)(x-a) + \frac{f''(c)}{2!}(x-a)^2, c \in [a, a+h]$

$x \leftarrow a + h$

$f(a+h) = f(a) + f'(a)h + \frac{f''(c)}{2!}h^2$

양변 h로 나눈 후 정리

$\frac{f(a+h)-f(a)}{h} - f'(a) = \frac{f''(c)}{2!}h$

$\frac{f(a+h)-f(a)}{h}$: 전방 차분으로 구한 수치 미분 값, $f'(a)$: 미분계수 값

$\left|\frac{f(a+h)-f(a)}{h} - f'(a)\right|$ = 오차 = $|f''(c)|\frac{h}{2!}$

∴ Error of forward scheme = $|f''(c)|\frac{h}{2!}$

수치 미분 – 중앙 차분 오차 증명 (생략 가능)

2) 중앙 차분

By Tayler series,

$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(c)}{3!}(x-a)^3$, $c \in [a, a+h]$

ⓐ $x \leftarrow a+h$

→$f(a+h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \frac{f^{(3)}(c)}{3!}h^3$

ⓑ $x \leftarrow a-h$

→ $f(a-h) = f(a) - f'(a)h + \frac{f''(a)}{2!}h^2 - \frac{f^{(3)}(c)}{3!}h^3$

ⓐ-ⓑ → $f(a+h) - f(a-h) = 2f'(a)h + 2\frac{f^{(3)}(c)}{3!}h^3$, 양변 $2h$로 나눈 후 정리

$\frac{f(x+h)-f(x-h)}{2h} - f'(a) = \frac{f^{(3)}(c)}{3!}h^2$

$\frac{f(x+h)-f(x-h)}{2h}$: 중앙 차분으로 구한 수치 미분 값, $f'(a)$: 미분계수 값

$\left|\frac{f(x+h)-f(x-h)}{2h} - f'(a)\right|$ = 오차 = $\left|f^{(3)}(c)\right|\frac{h^2}{3!}$

∴ Error of forward scheme = $\left|f^{(3)}(c)\right|\frac{h^2}{3!}$

수치 미분 – 중앙 차분 오차 증명 (생략 가능)

- Error of forward scheme = $|f''(c)|\frac{h}{2!}$
- Error of forward scheme = $\left|f^{(3)}(c)\right|\frac{h^2}{3!}$

For sufficiently small h, h>0 → $h \gg h^2$

따라서 중앙 차분의 결과가 오차가 작다.


감사합니다.