Backpropagation을 위한
수학적 배경 지식

수업 목표

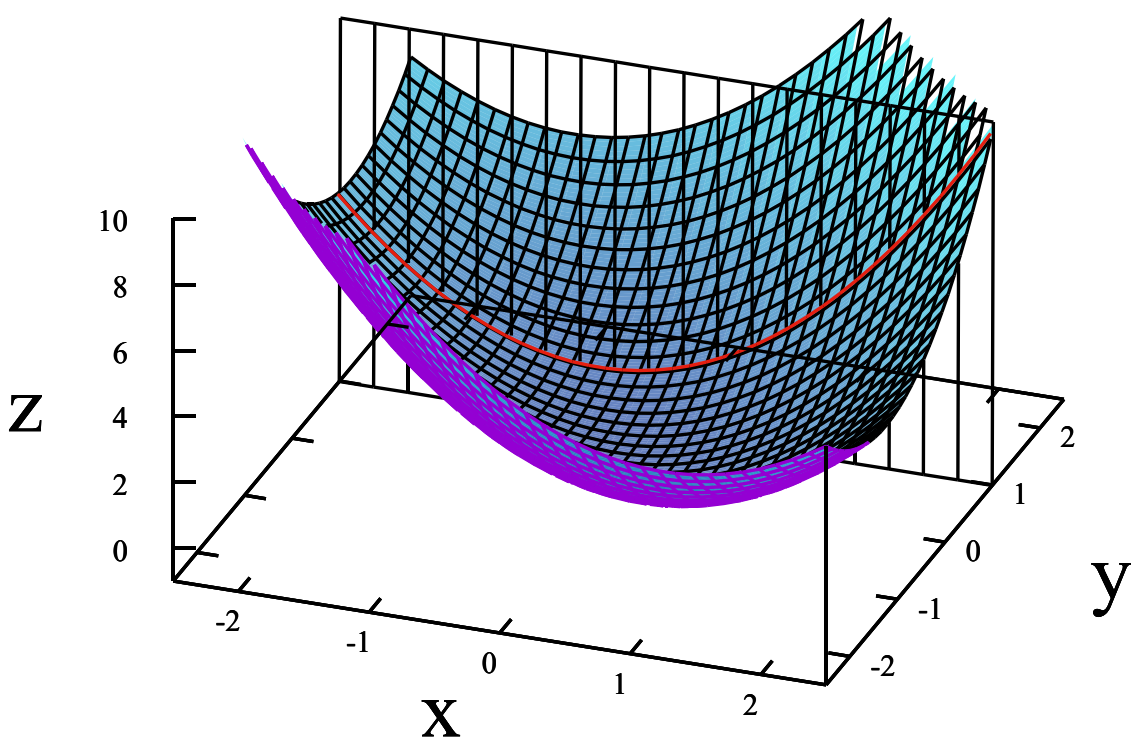
이번 수업의 핵심:

- 편미분의 개념
- Gradient descent를 통한 딥러닝 모델의 학습
- 합성 함수 미분을 위한 Chain rule 이해
- 편미분과 Chain rule을 활용한 Backpropagation 기본 개념

핵심 개념

- 편미분
- Gradient descent
- Chain rule
- Upstream gradient, Downstream gradient, Local gradient

다변수 함수의 편미분



다변수 함수 $f(x,y)$를 x로 **편미분**: $\frac{\partial f}{\partial x}$

- x 이외에 모든 변수를 상수 취급
  → 단변수 함수 $g(x) = f(x, y = a)$

- $g(x)$를 x에 대해서 미분한 함수

$$\frac{d}{dx}g(x) = \frac{\partial}{\partial x}f(x,y)$$

Gradient Descent (경사하강법)

다음의 선형회귀 모델과 문제를 생각해보자:

- 모델:

$$y = ax + b$$

- Loss function: $\mathcal{L}(a, b) = \sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2$

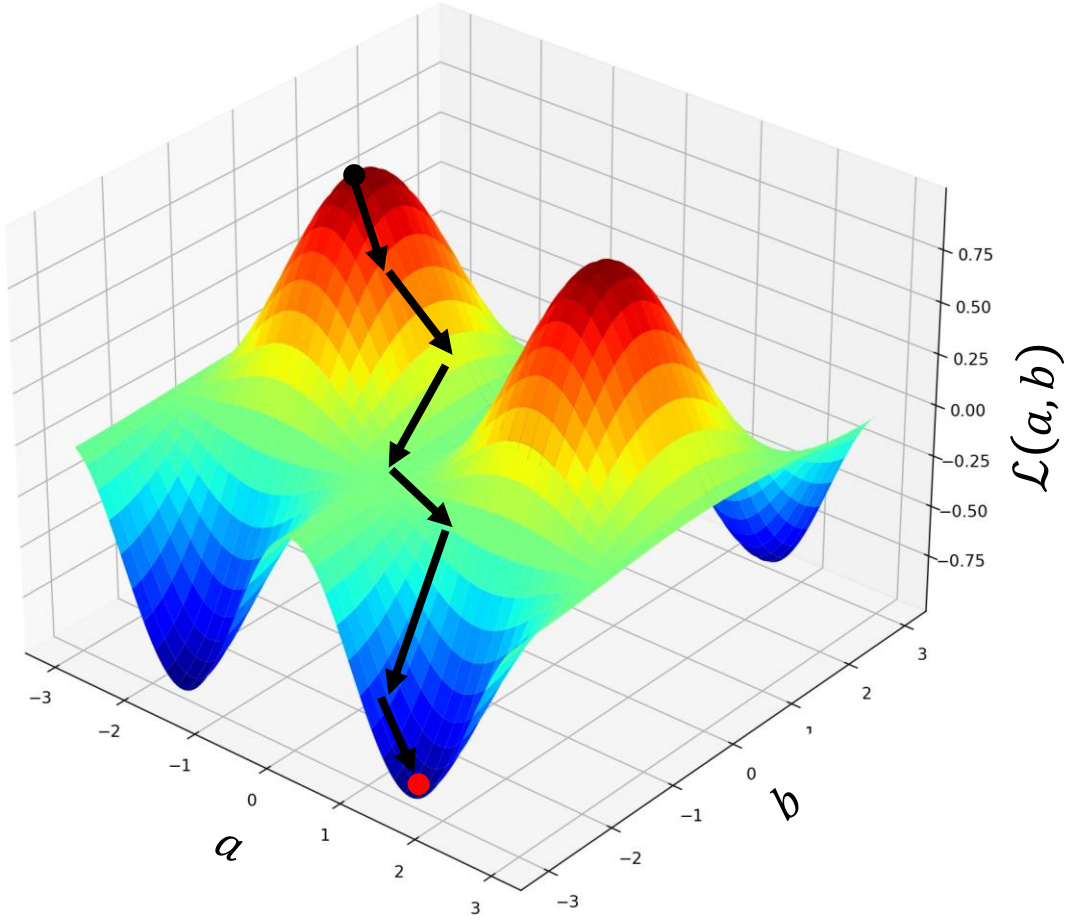
- Training Data: e.g., $a = 2, b = 1$

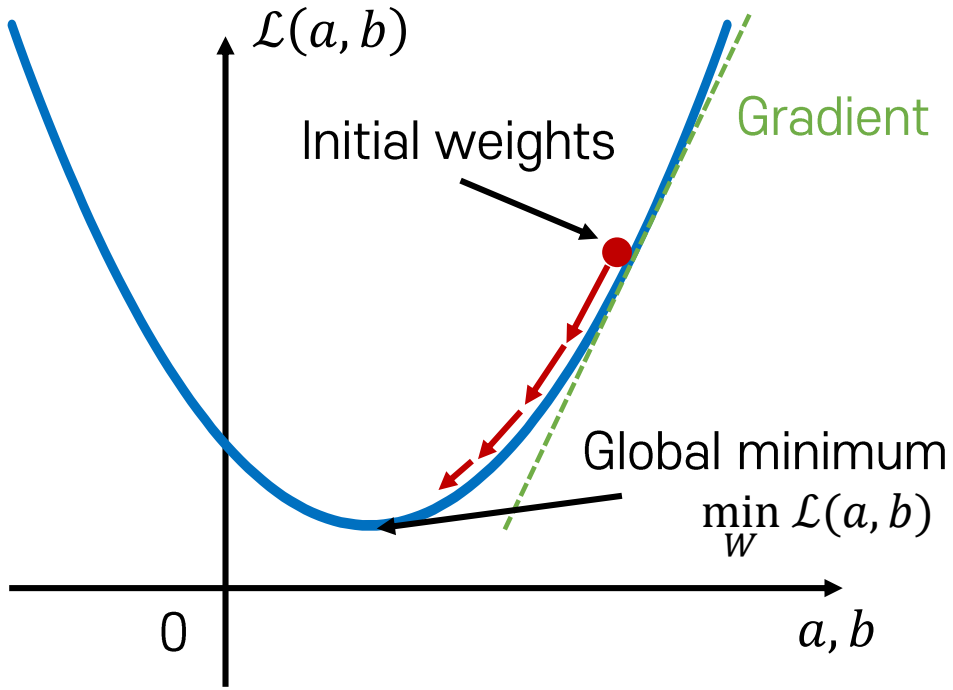
x_i	y_i
-2	-1
0	1
1	2

$ax_i + b$	$y_i - (ax_i + b)$	$\left(y_i - (ax_i + b)\right)^2$
-3	2	4
1	0	0
3	-1	1
		↓
$\mathcal{L}(a, b) = \sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2$		5

Gradient Descent (경사하강법)

a, b를 랜덤한 초기값으로 정해주고, 아래 수식을 통해 조금씩 갱신해서,
최종적으로 Loss function $\mathcal{L}(a, b)$가 최소값이 되는 최적의 a, b 값을 도출하는 알고리즘





$$a := a - \alpha \frac{\partial \mathcal{L}(a,b)}{\partial a}$$

$$b := b - \alpha \frac{\partial \mathcal{L}(a,b)}{\partial b}$$

Gradient Descent (경사하강법)

1. 먼저 a, b를 임의의 값으로 초기화: $a = 2, b = 1$

2. Loss function에 대한 a, b의 편미분 함수:

$$\mathcal{L}(a,b) = \sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2, \qquad \frac{\partial \mathcal{L}(a,b)}{\partial a} = \sum_{i=1}^{n} -2x_i\left(y_i - (ax_i + b)\right)$$

x_i	y_i	$ax_i + b$	$y_i - (ax_i + b)$	$\left(y_i - (ax_i + b)\right)^2$	$-2x_i\left(y - (ax_i + b)\right)$
-2	-1	-3	2	4	8
0	1	1	0	0	0
1	2	3	-1	1	2

$\mathcal{L}(a,b)$	$\frac{\partial \mathcal{L}(a,b)}{\partial a}$
5	10

Gradient Descent (경사하강법)

1. 먼저 a, b를 임의의 값으로 초기화: $a = 2, b = 1$

2. Loss function에 대한 a, b의 편미분 함수:

$$\mathcal{L}(a,b) = \sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2, \qquad \frac{\partial \mathcal{L}(a,b)}{\partial b} = \sum_{i=1}^{n} -2\left(y_i - (ax_i + b)\right)$$

x_i	y_i	$ax_i + b$	$y_i - (ax_i + b)$	$\left(y_i - (ax_i + b)\right)^2$	$-2\left(y - (ax_i + b)\right)$
-2	-1	-3	2	4	-4
0	1	1	0	0	0
1	2	3	-1	1	2

$\mathcal{L}(a,b)$	$\frac{\partial \mathcal{L}(a,b)}{\partial b}$
5	-2

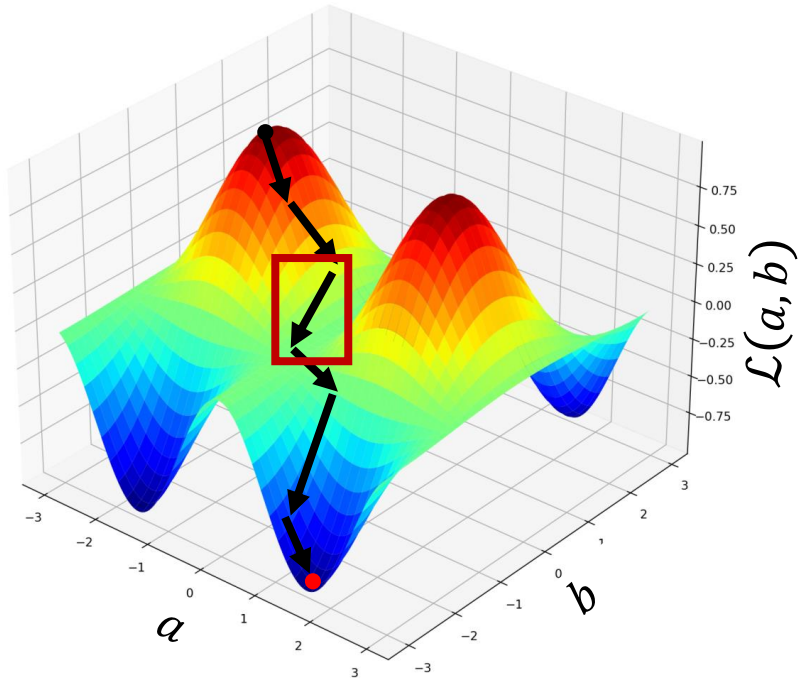
Gradient Descent (경사하강법)

3. 편미분 값과 사전에 정한 Learning rate $\alpha = 0.1$를 사용하여 a, b를 갱신

$$a := a - \alpha \frac{\partial \mathcal{L}(a,b)}{\partial a} = 2 - 0.1 \times 10 = 1.0$$

$$b := b - \alpha \frac{\partial \mathcal{L}(a,b)}{\partial b} = 1 - 0.1 \times (-2) = 1.2$$

- 한번 이동을 step이라 표현 →



Gradient Descent (경사하강법)

- 갱신 이전의 값: $a = 2, b = 1$

x_i	y_i
-2	-1
0	1
1	2

$ax_i + b$	$y_i - (ax_i + b)$	$\left(y_i - (ax_i + b)\right)^2$
-3	2	4
1	0	0
3	-1	1
		↓
$\mathcal{L}(a,b) = \sum_{i=1}^{n}\left(y_i - (ax_i + b)\right)^2$		5

- 갱신 이후의 값: $a = 1.0, b = 1.2$

x_i	y_i
-2	-1
0	1
1	2

$ax_i + b$	$y_i - (ax_i + b)$	$\left(y_i - (ax_i + b)\right)^2$
-0.8	-0.2	0.04
1.2	-0.2	0.04
2.2	-0.2	0.04
		↓
$\mathcal{L}(a,b) = \sum_{i=1}^{n}\left(y_i - (ax_i + b)\right)^2$		0.12

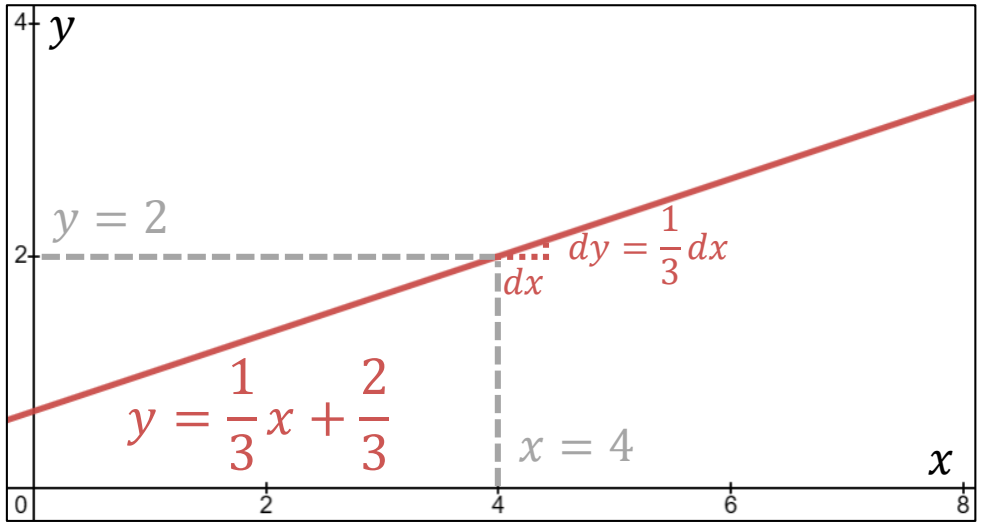
Chain Rule (연쇄 법칙): 합성 함수의 미분

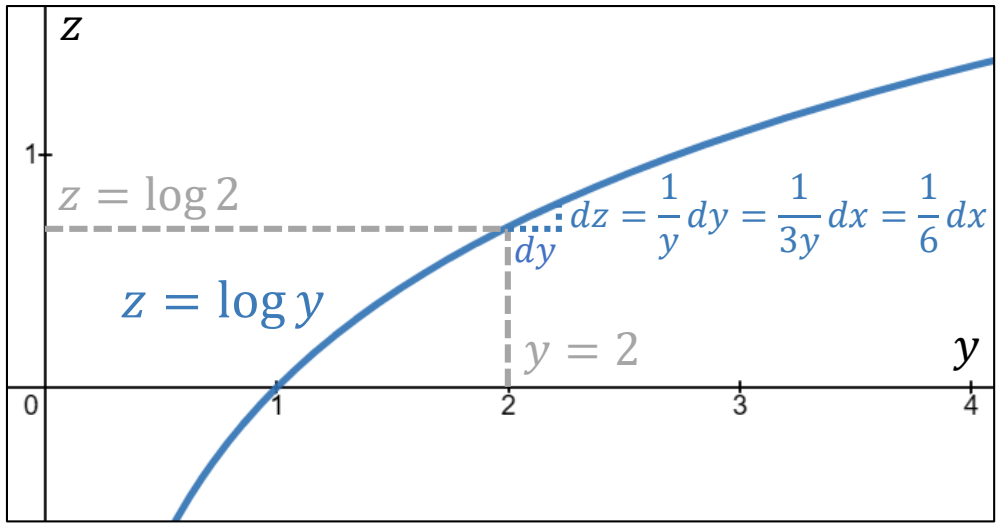
- 합성 함수 $z = h(x) = g\big(f(x)\big)$에 대해서 x에 대한 미분 $\frac{d}{dx}h(x)$은 다음과 같음:

$$\frac{dh}{dx} = \frac{dh}{df}\frac{df}{dx} = g'\big(f(x)\big)f'(x)$$

→ 합성 함수의 미분은 각각의 함수 미분의 곱

- 만약 $y = f(x) = \frac{1}{3}x + \frac{2}{3}, \frac{dy}{dx} = \frac{1}{3}$ 그리고 $z = g(y) = \log y, \frac{dz}{dy} = \frac{1}{y}$ 이라면:





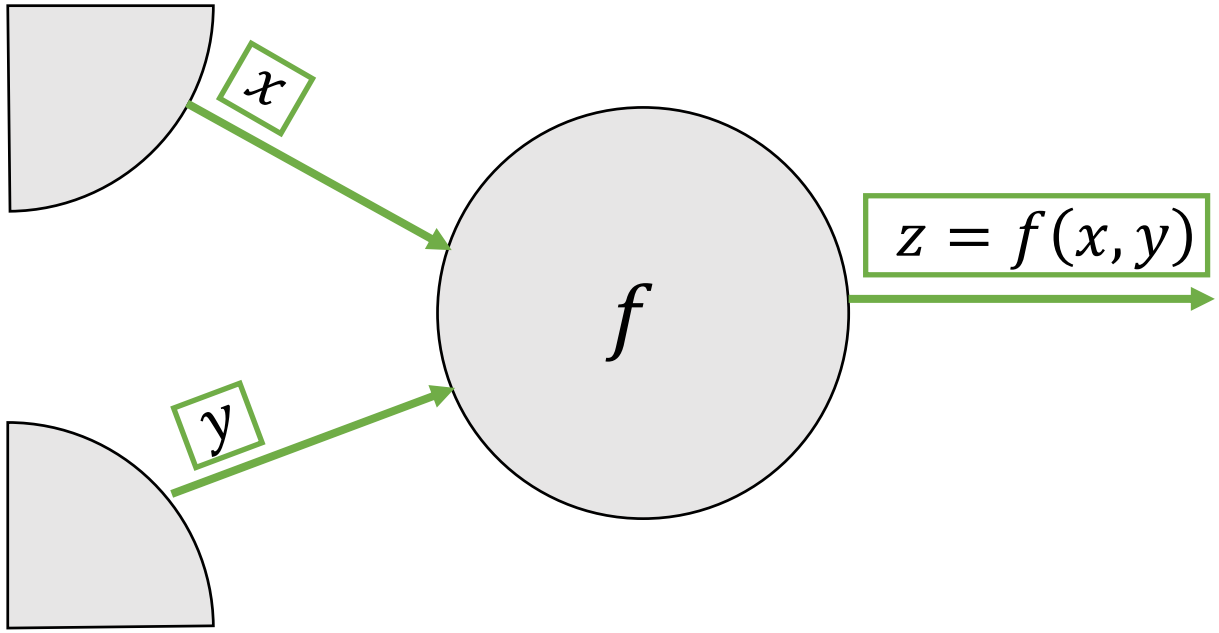
$$\frac{dz}{dx} = \frac{1}{3} \times \frac{1}{y}$$

다변수 함수의 Chain Rule

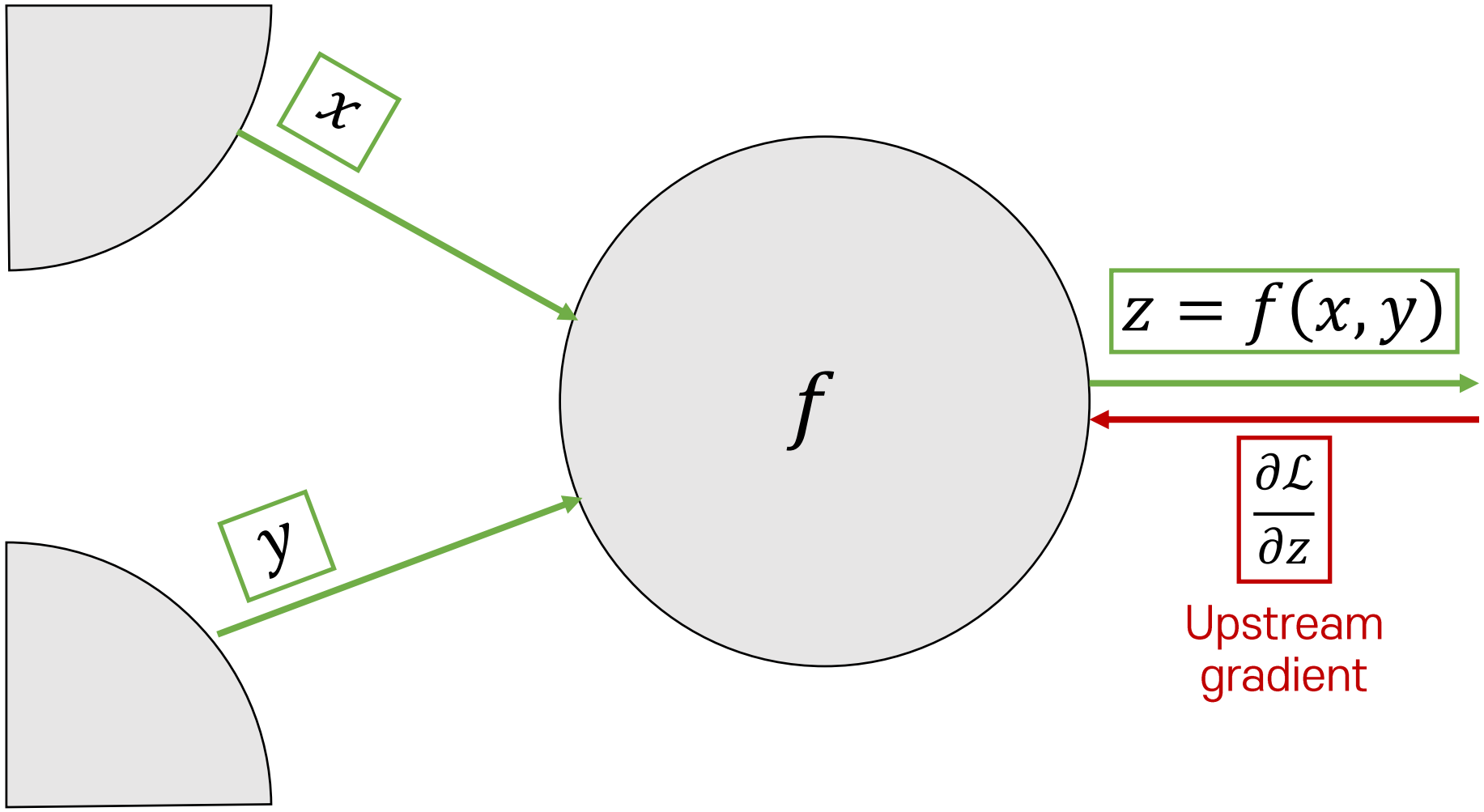
- 최종 Loss $\mathcal{L}$이 $z = f(x, y)$에 종속적인 상황 가정 → $\mathcal{L}(z) = \mathcal{L}\big(f(x, y)\big)$

- Loss 값을 낮추기 위해 x, y에 대해 Gradient Descent를 사용 → $\frac{\partial \mathcal{L}}{\partial x}, \frac{\partial \mathcal{L}}{\partial y}$ 필요

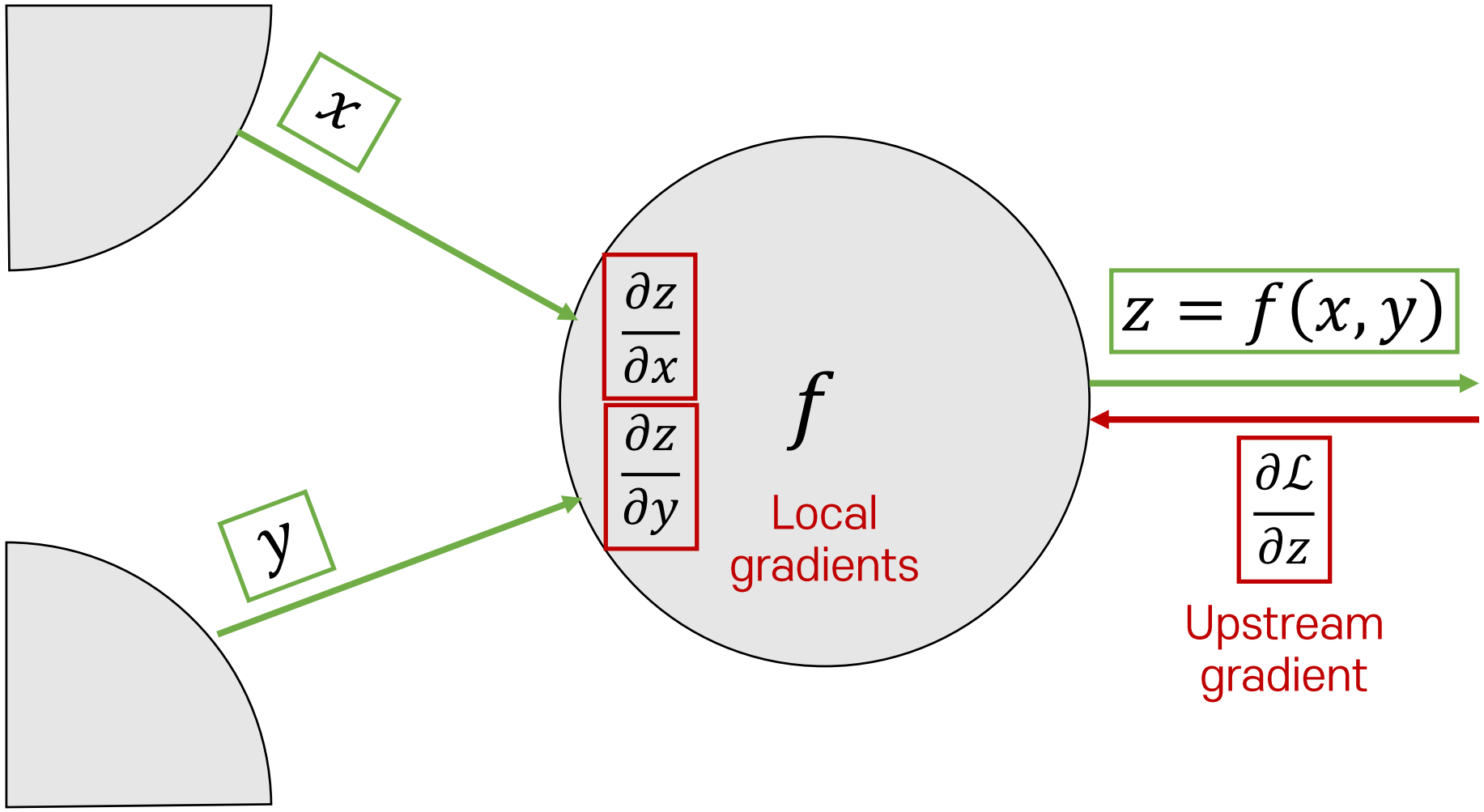
$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial \mathcal{L}}{\partial z}\frac{\partial z}{\partial x}, \qquad \frac{\partial \mathcal{L}}{\partial y} = \frac{\partial \mathcal{L}}{\partial z}\frac{\partial z}{\partial y}$$



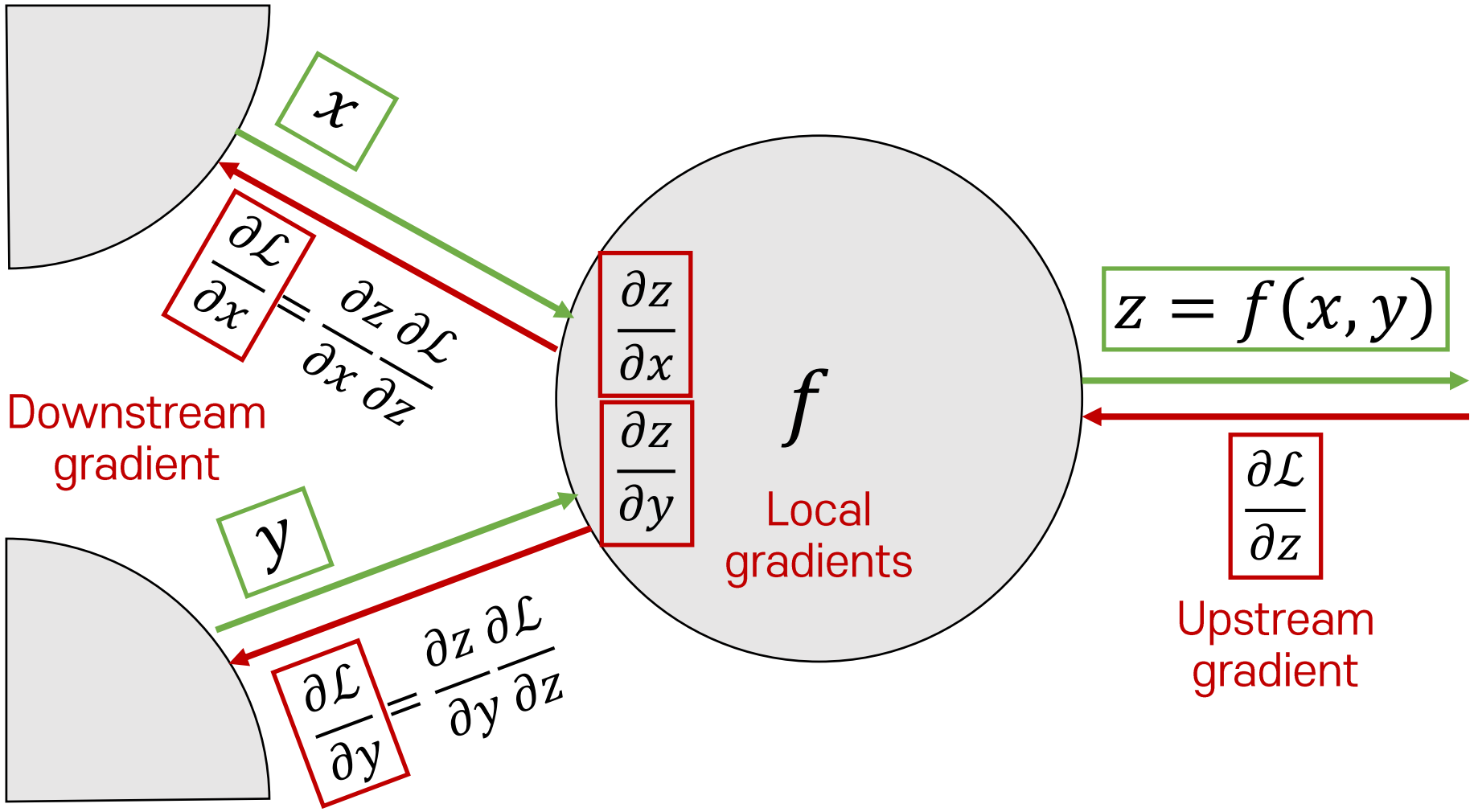
편미분과 Chain Rule



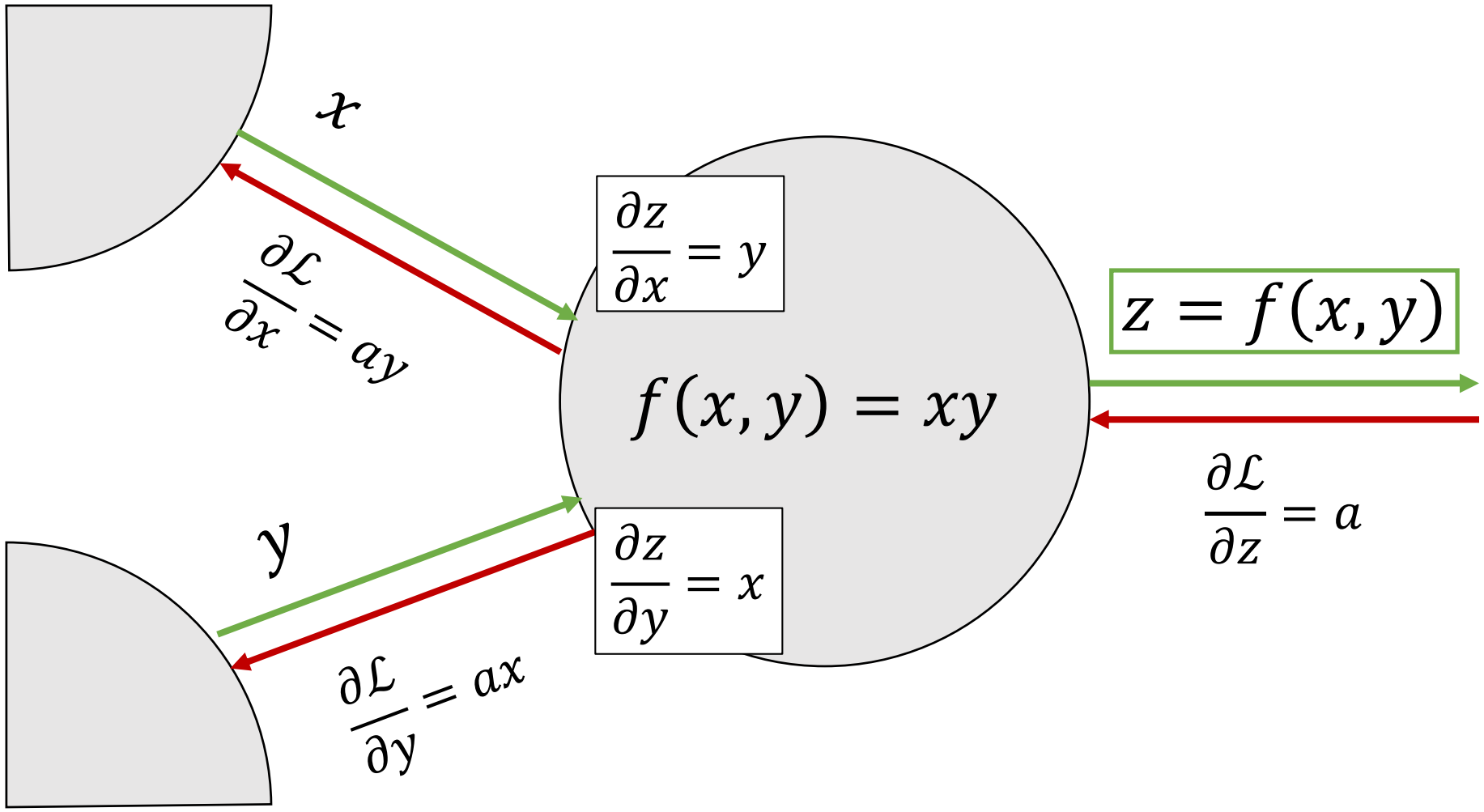
편미분과 Chain Rule



편미분과 Chain Rule



편미분과 Chain Rule



요약

- 최적화 과정에서의 Gradient descent의 필요성
- 편미분과 Chain rule의 이해
- 편미분과 Chain rule을 통한 Backpropagation의 수학적 배경