

# Differentiation

## Gradient Descent (경사 하강법)

소프트웨어 끈대 강의

노기섭 교수

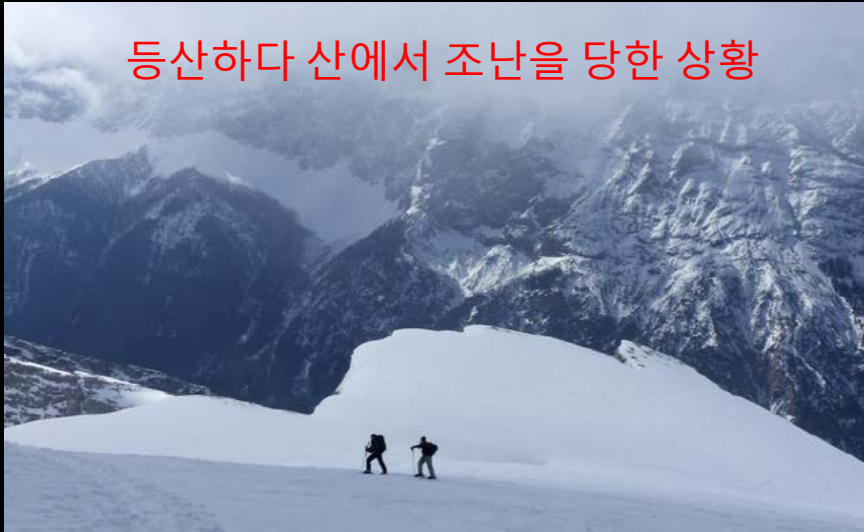
(kafa46@cju.ac.kr)

# Course Overview

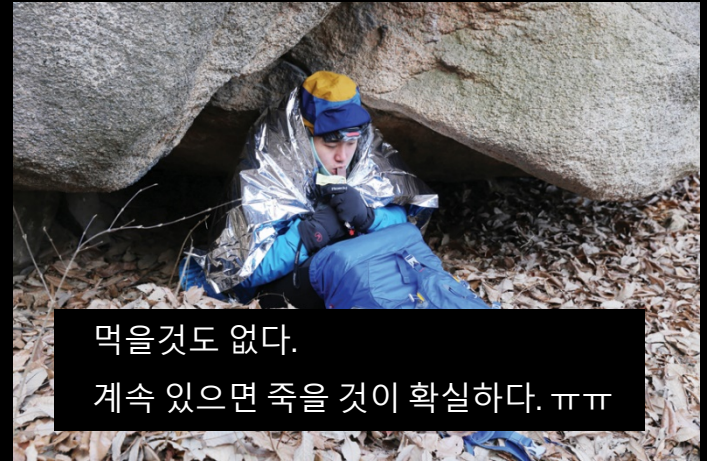
Topic	Contents
01. Orientation 오리엔테이션	Course introduction, motivations, final objectives 과정 소개, 동기부여, 최종 목표
02. Learning in deeplearning 딥러닝 학습	How does the deeplearning learns knowledge from data 어떻게 딥러닝은 데이터로부터 지식을 배우는가?
03. Principle of differentiation 미분의 원리	Basics of differentiation (concepts, notation, 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

# Gradient Descent 가 필요한 상황

등산하다 산에서 조난을 당한 상황



이미지 출처: [https://m.health.chosun.com/svc/news\\_view.html?contid=2022113002070](https://m.health.chosun.com/svc/news_view.html?contid=2022113002070)



먹을것도 없다.

계속 있으면 죽을 것이 확실하다.  $\pi\pi$

이미지 출처: <https://san.chosun.com/news/articleView.html?idxno=9945>



이미지 출처: <https://m.blog.naver.com/gy007/221927670633>

눈을 다쳐서 앞을 볼 수 없다.  $\pi\pi$



스마트폰은 완전히 박살났다.

여러분이라면  
어떻게  
하실건가요?

어떤 전략을  
선택할 건가요?

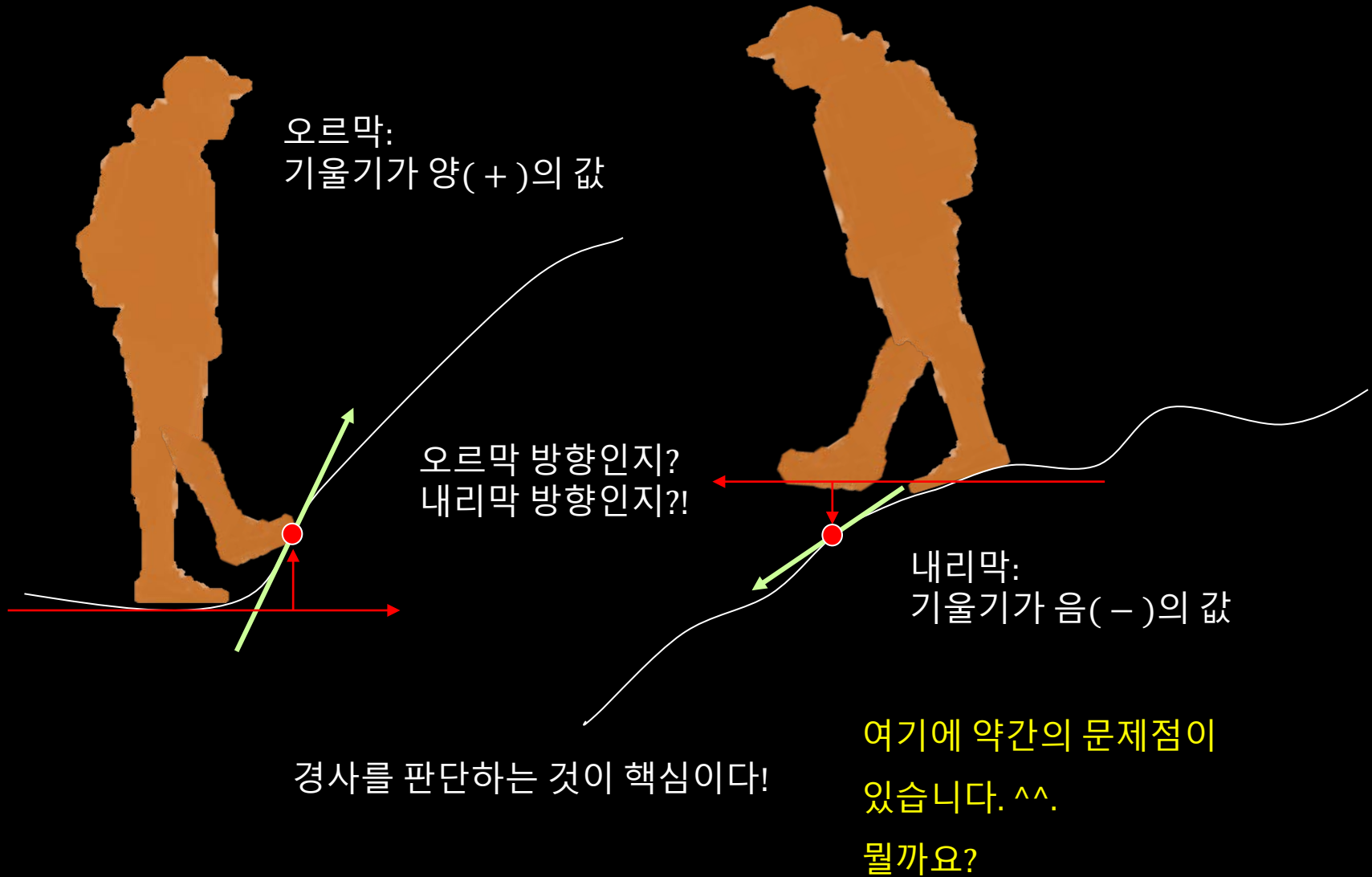
# 조난자의 전략과 딥러닝의 비교

시각장애인이 낮은 지역으로 이동하려는 상황과 정확히 같다.

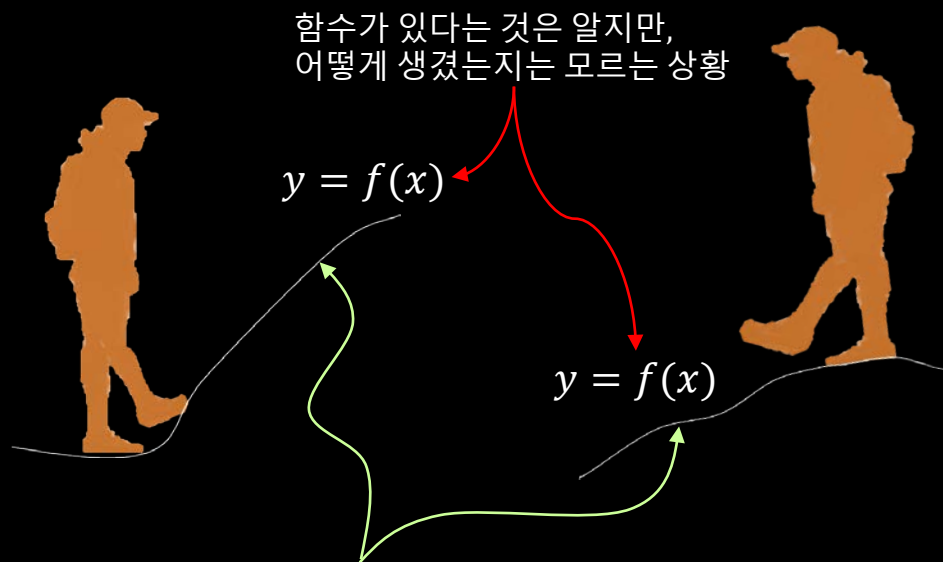
등산 조난자의 전략	딥러닝의 전략
그나마 낮은 지역으로 가야 생존 확률이 높아진다.	손실함수의 값이 최소가 되는 지점으로 가야 예측오차를 줄인다.

순번	등산 조난자의 알고리즘	딥러닝의 알고리즘
1	앞을 볼 수 없으니... $\pi\pi$ 한발짝 내딛어 오르막/내리막을 판단한다.	데이터의 정확한 분포를 모른다. $\pi$ 손실함수를 통해 오차의 증감을 판단한다.
2	내리막 방향으로 한발짝 움직인다. 만약 평지라면 목표에 도착한 것으로 판단하고 움직이는 것을 멈춘다.	손실값이 줄어드는 방향으로 움직인다. 만약, 손실값의 변화가 없다면, 최소지점으로 판단하고 멈춘다.
3	움직일때 위험할 수 있으니 한발짝 내딛는 거리는 적절히 조절한다.	추가 변수를 도입하여 움직이는 양을 적절히 조절한다 (learning rate, 학습률)
4	1~3번 과정을 반복한다.	1~3번 과정을 반복한다.

# 등산 조난자의 접근법



# 다변수 함수의 경우 고려사항



교수님~ 그런데요...

다변수 함수면 어떻게 해요?

우리는 3차원 공간에 있잖아요?

$x$  축 뿐만 아니라  $y$  축도 고려해야죠!!

이런 상황 때문에  
편미분에 배운 겁니다.^^

경사를 표현하는 함수가  
일변수 함수라면 아무 문제 없다.



$f(x)$ 에 대한 도함수  
(기울기함수, derivative)를 구한다.

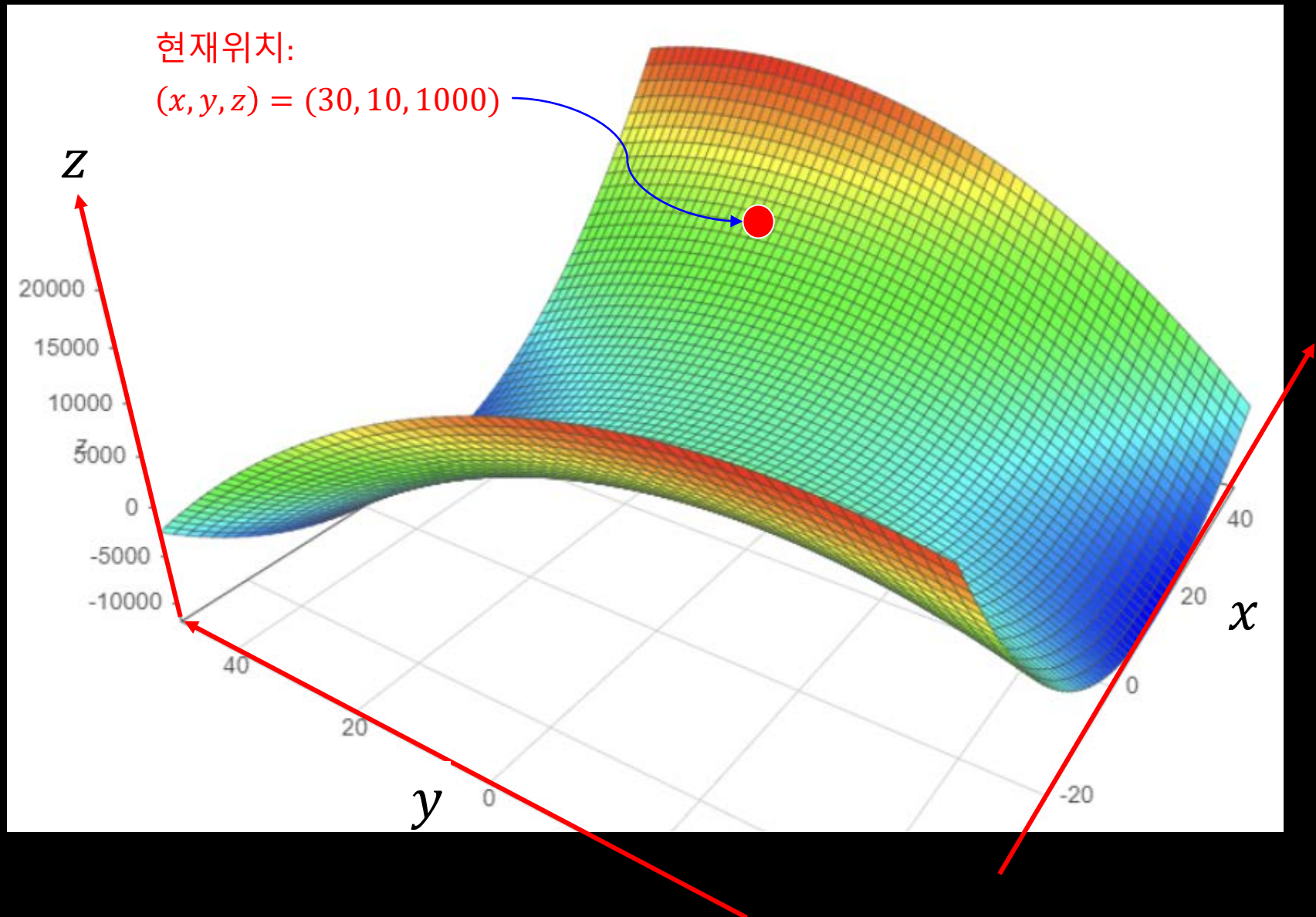


$$f'(x) = \frac{d}{dx}f(x)$$

해당 지점의 값(예:  $k$  지점)을 대입해서  
양수인지 음수인지 판단하면 간단하게 해결

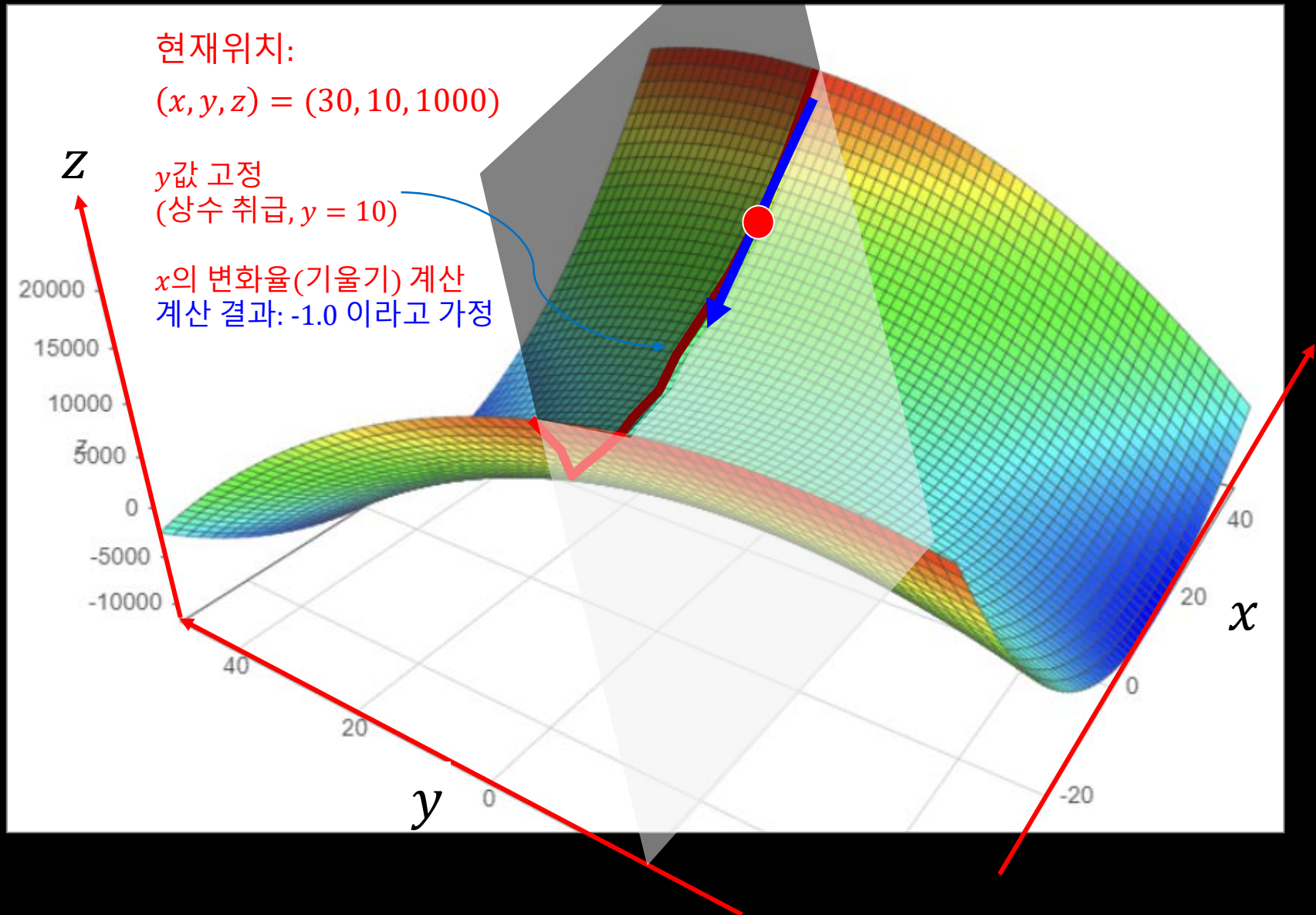
$$f'(k) > 0 \quad \text{또는} \quad f'(k) < 0 \quad \text{또는} \quad f'(k) = 0$$

# 먼저, 간단하게 3차원 공간 - 현재 위치 확인



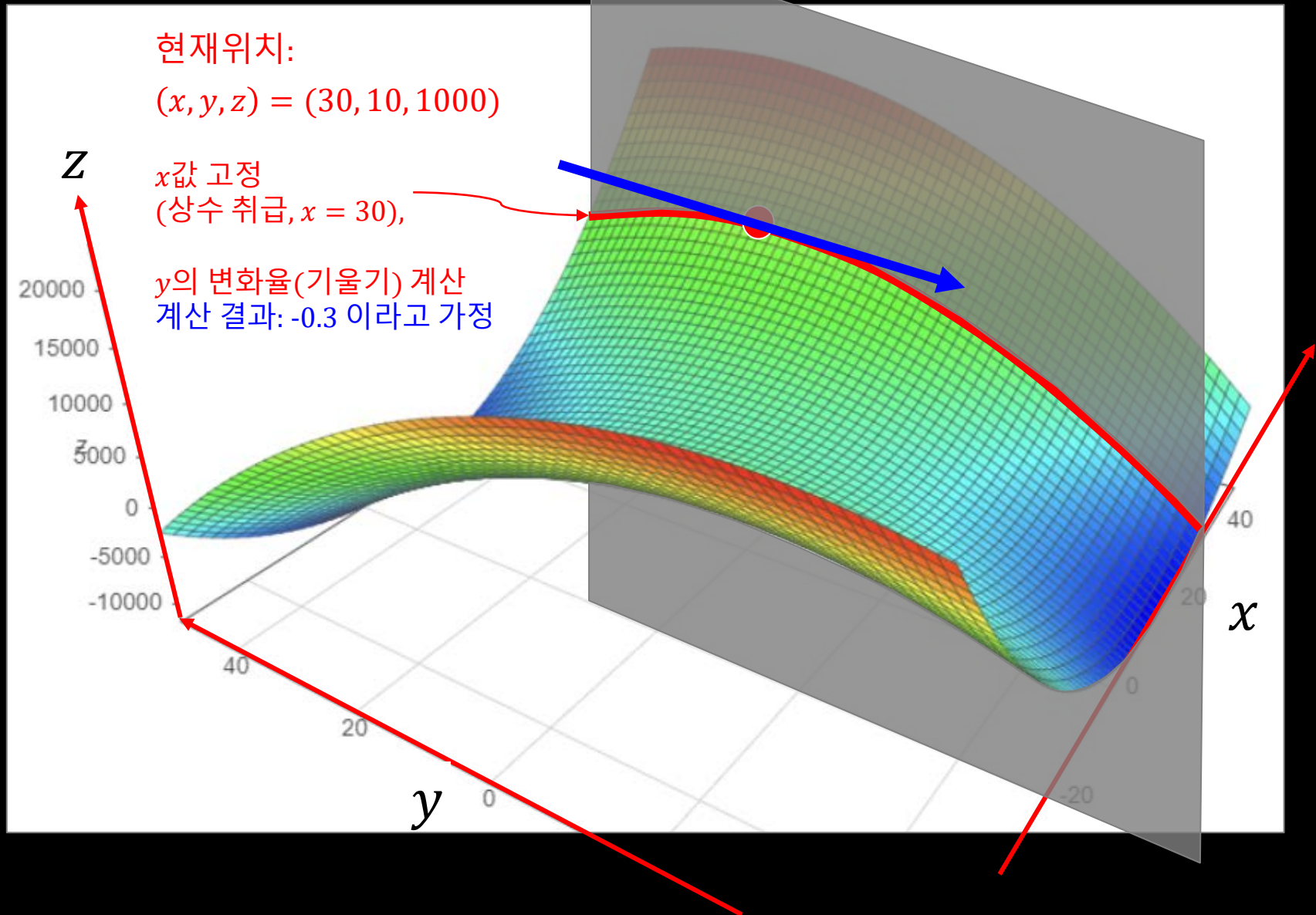


# 먼저, 간단하게 3차원 공간 - $x$ 에 대하여 편미분

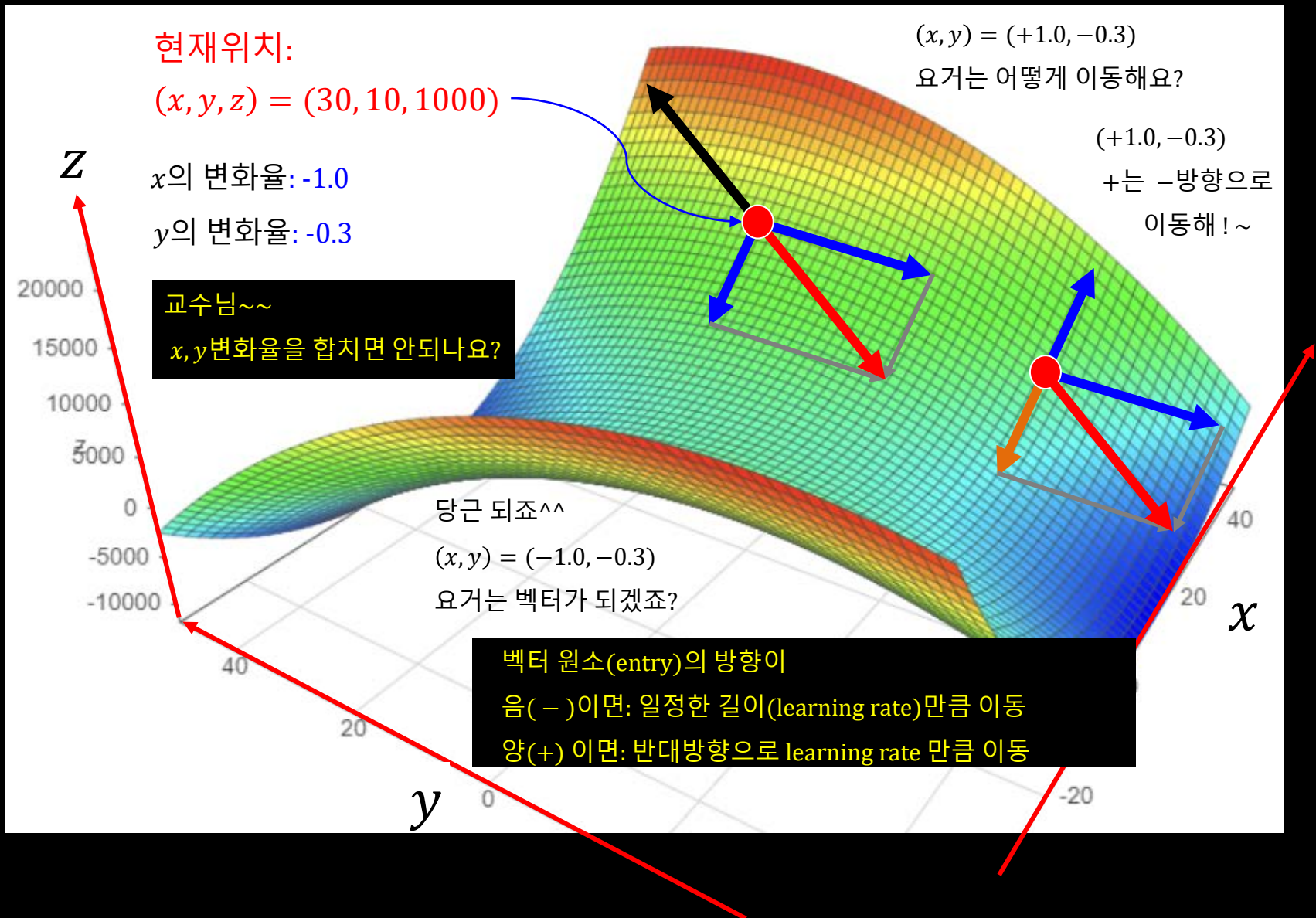




# 먼저, 간단하게 3차원 공간 - $y$ 에 대하여 편미분



# 먼저, 간단하게 3차원 공간 - 현재 위치 확인



# 마지막 남은 문제.... $\pi\pi\pi$

교수님~ 그런데요...

2변수 함수도 (결괏값  $z$  포함:  $x, y, z$ )

이렇게 복잡하네요...

딥러닝 변수는 파라미터....

파라미터는 찾아야 하는 변수....

근데, 파라미터는 겁나게 많다면서요....

786차원은 어떻게 찾나요?

그러면 1,024차원은 어떻게 하죠?  $\pi\pi\pi$

우리는 망한거 아닌가요?

그건 framework (예: pytorch)가  
알아서 계산해 줍니다.  
(수많은 동일 작업의 반복입니다.)

일은 컴퓨터에게 맡겨도  
원리는 반드시 알아야 합니다.  
(이것에 동의 못하는 분은  
딥러닝 엔지니어로서의  
자격이 없습니다.)

컴퓨터가 알려주는  
결괏값이 어떻게 나왔는지는  
꼭!  
꼭!!!!!!!!!!!!!!  
꼭!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
이해하고 있어야 합니다.

# One more thing



# Stochastic Gradient Descent

## Stochastic Gradient Descent (확률적 경사 하강법)

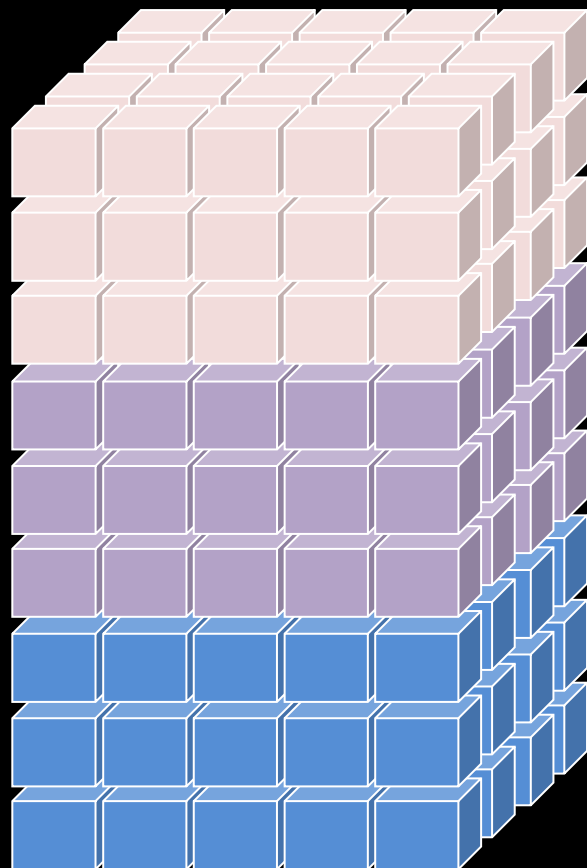
이전 참고 강의

확률 이론

[Probability]\_07-4. Stochastic Gradient Descent에  
왜 "Stochastic"라는 단어가 붙은 건가?  
미니배치에서 Stochastic 의미는?

유튜브 링크: <https://youtu.be/DEQhCJ0nav4>

# SGD converge



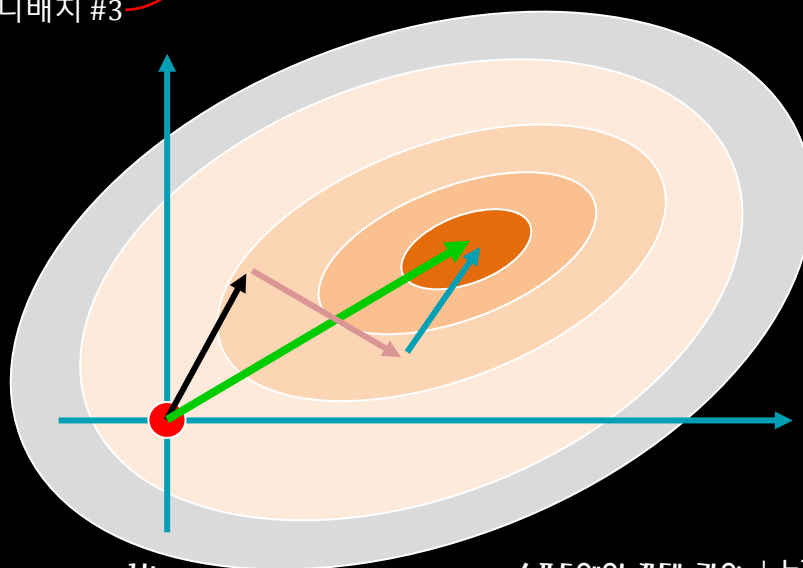
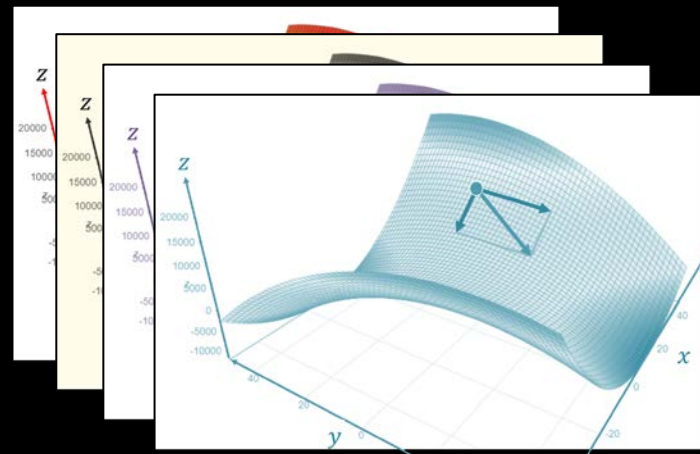
전체 데이터

미니배치 #1

미니배치 #2

미니배치 #3

$$L(\theta) = \sum_{i=1}^n (\hat{y}_i - y_i)^2$$







수고하셨습니다 ..^^..