

경사하강법 튜아보기

AI 스터디 라운지_LIVE 리뷰

2025.07.18(금)

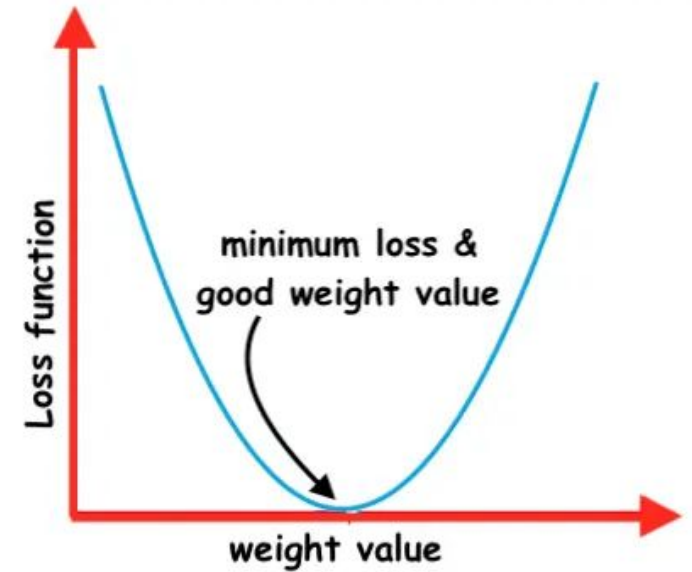
River코치

프리코스 2

- 1. 기초 수학 첫걸음 - 2. 경사하강법 ([link](#))
- 3. 딥러닝 핵심 기초 - 5. 최적화: Optimization ([link](#))

손실함수 (Loss Function) L

- 실제값(정답)과 모델의 예측값의 차이
- 예: MSE(Mean Squared Error), Cross Entropy
- 비용함수(Cost Function) 또는 목적함수(Objective Function)으로도 불림
- 우리가 원하는 것은 손실함수의 값을 최대한 줄이는 것
(즉, 손실을 최소화하는 모델 파라미터 찾기)
- 모델 학습(training): 반복적으로 모델 파라미터(가중치 등)를 조정해서 손실 함수가 점점 작아지도록 만드는 과정



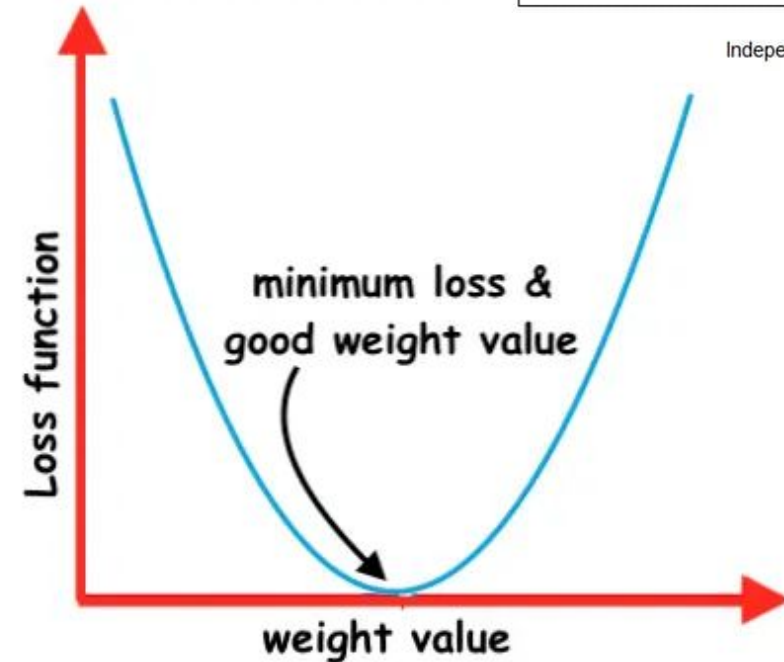
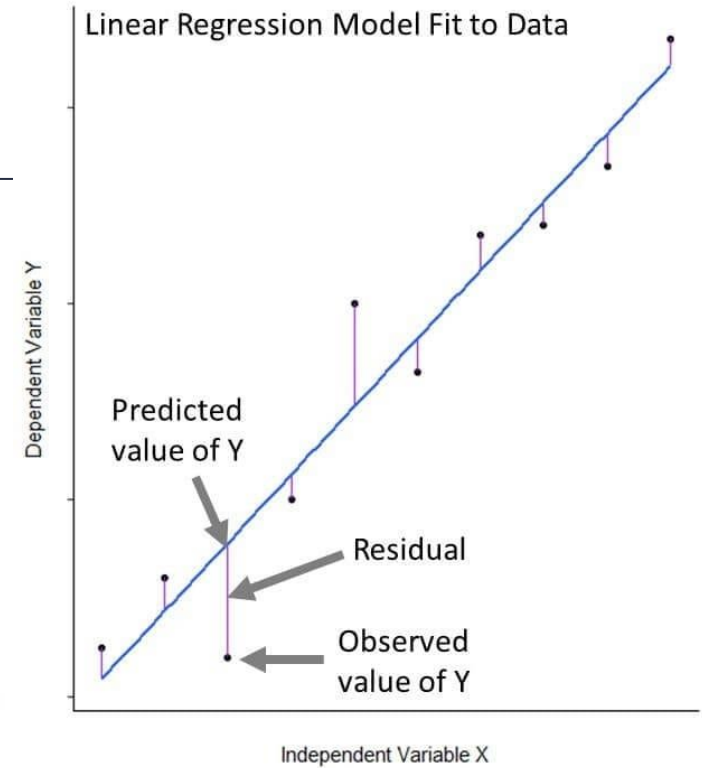
$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

손실함수 (Loss Function) L

- (예시) Linear Regression

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

Parameter w 에 대한 함수로
표현됨:
(예) $\hat{y} = w_1x + w_0$

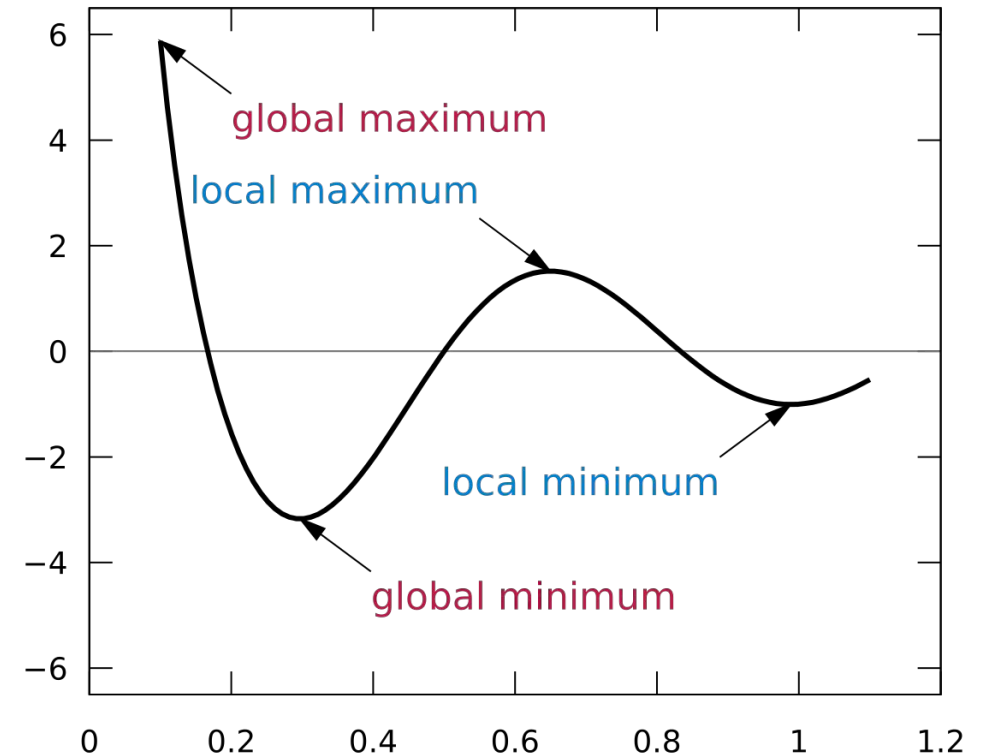


최솟값 찾기의 어려움

“미분해서 0이 되는 점을 찾으면 되지 않을까?”

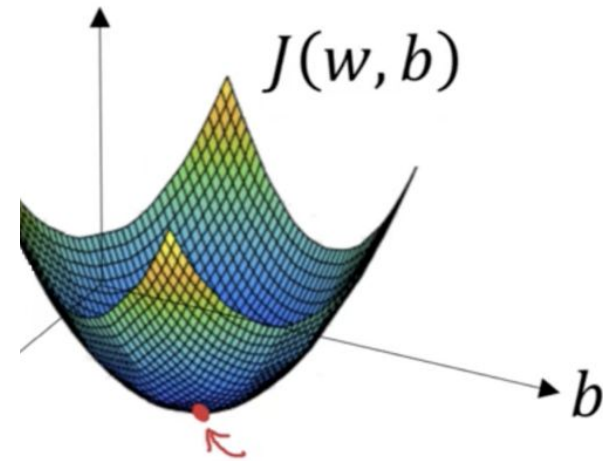
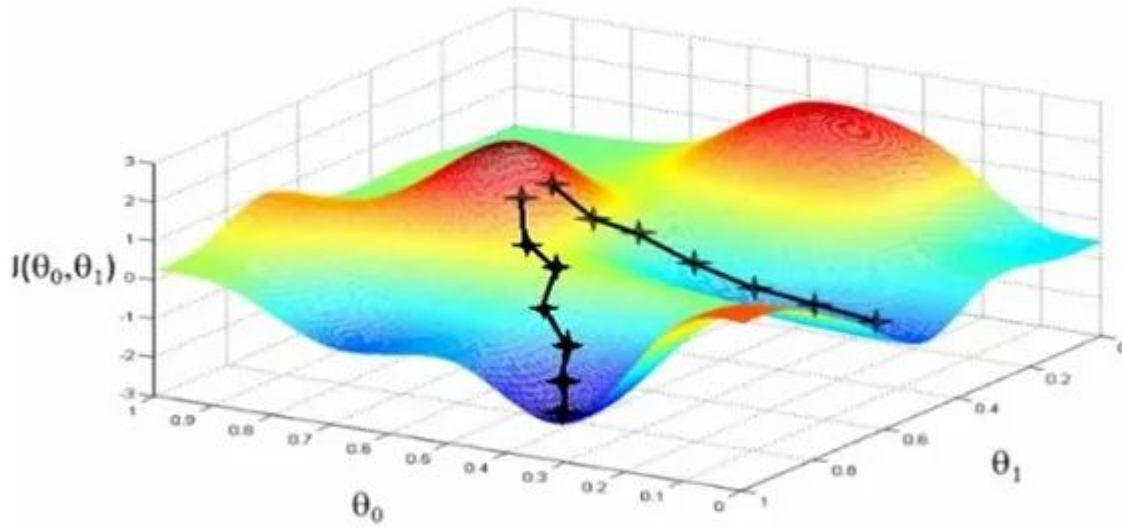
- 데이터가 큰 경우: 미분 연산량이 매우 큼
- 모델이 복잡한 경우: 복잡한 비선형 함수

$$\theta = (X^T X)^{-1} X^T y$$



• History

- 19세기 프랑스의 수학자 오귀스탱 루이 코시(Augustin-Louis Cauchy)가 제안(1847)
- 천체의 움직임을 계산하기 위해 사용

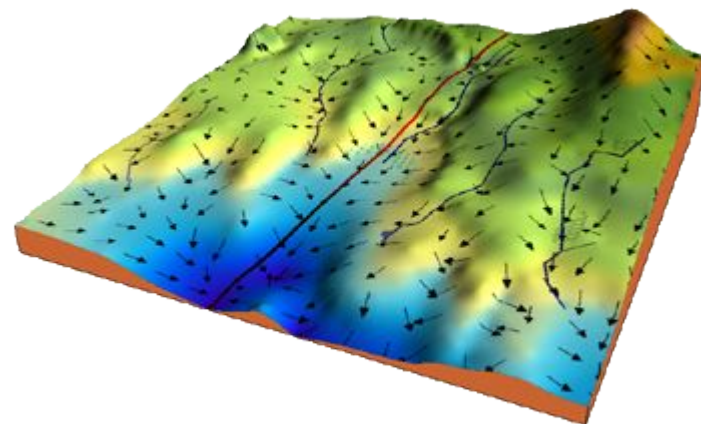


경사하강법 비유

- 등산 조난자가 집으로 돌아가려면



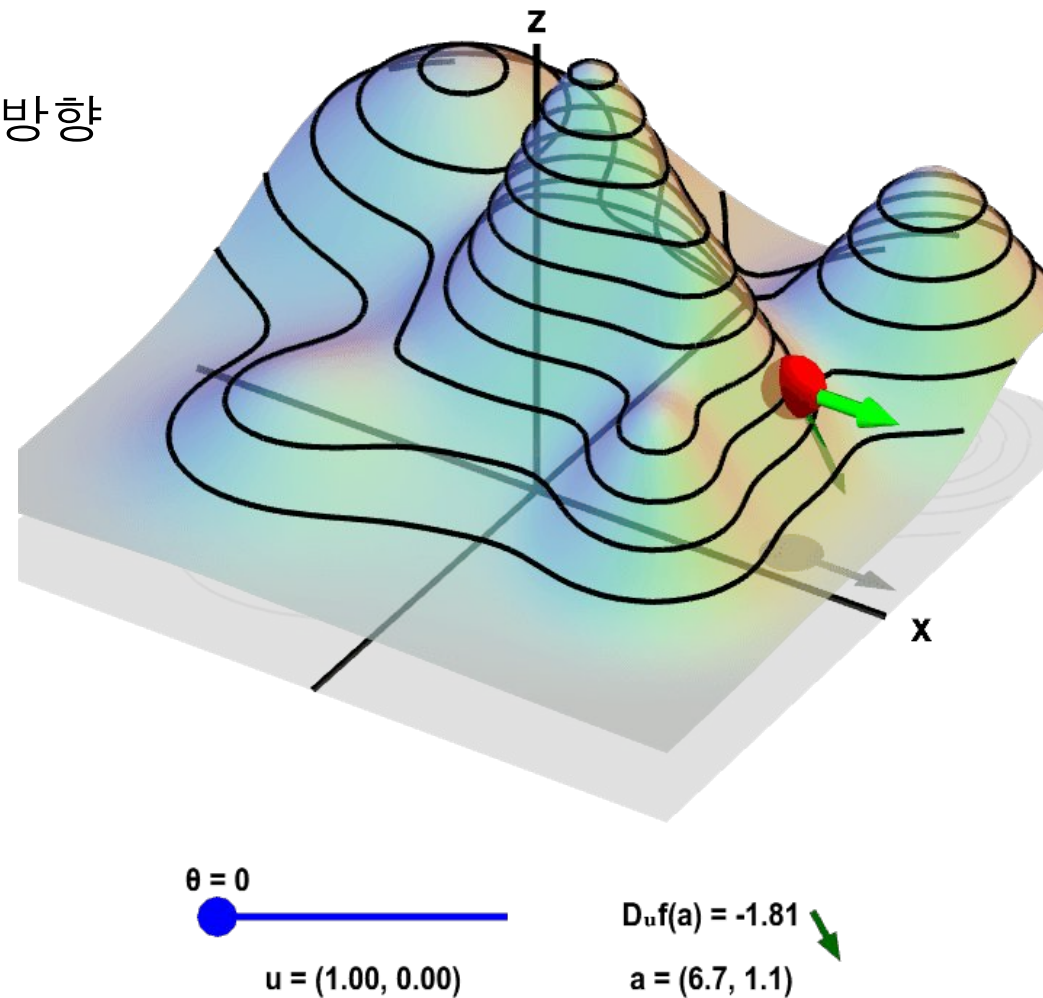
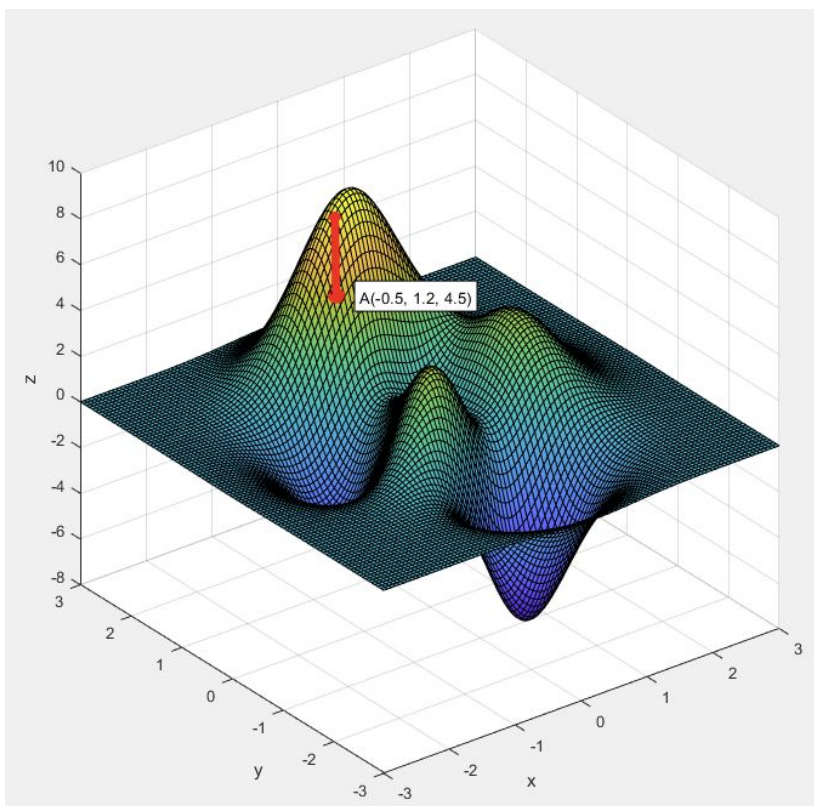
등산 조난자의 접근법



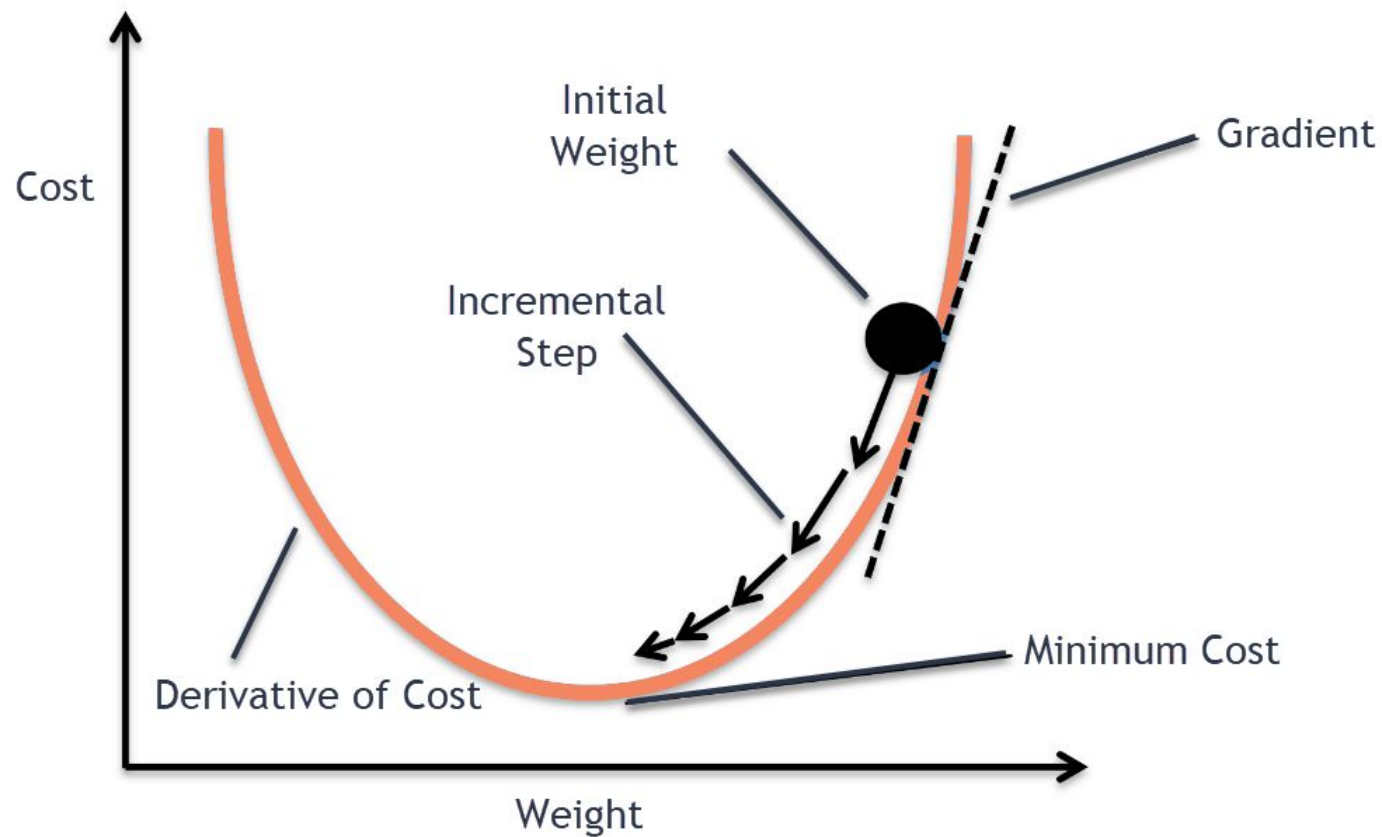
경사(Gradient)

- Gradient

- 함수값이 가장 급격하게 변하는(증가하는) 방향



경사하강법

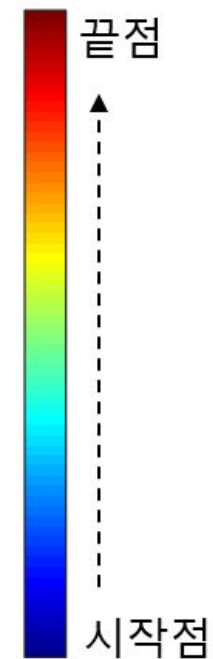
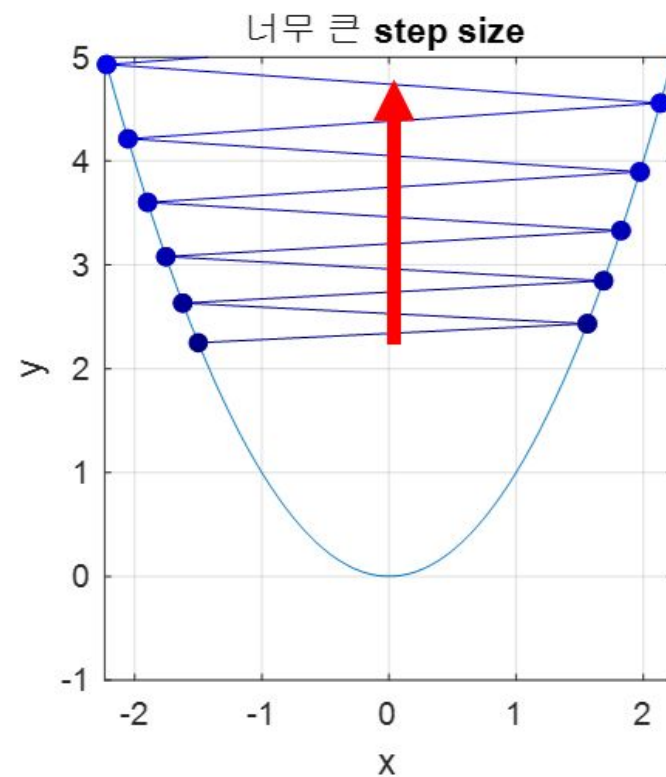
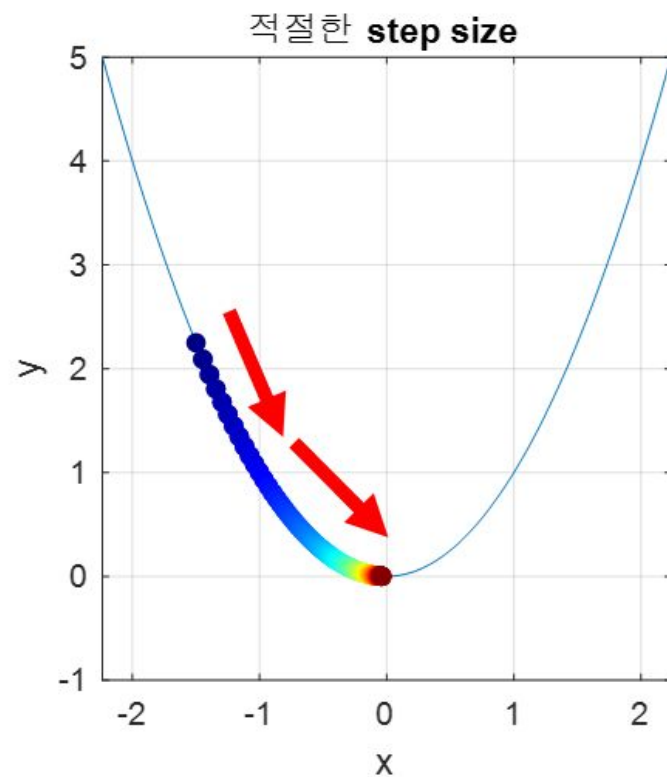
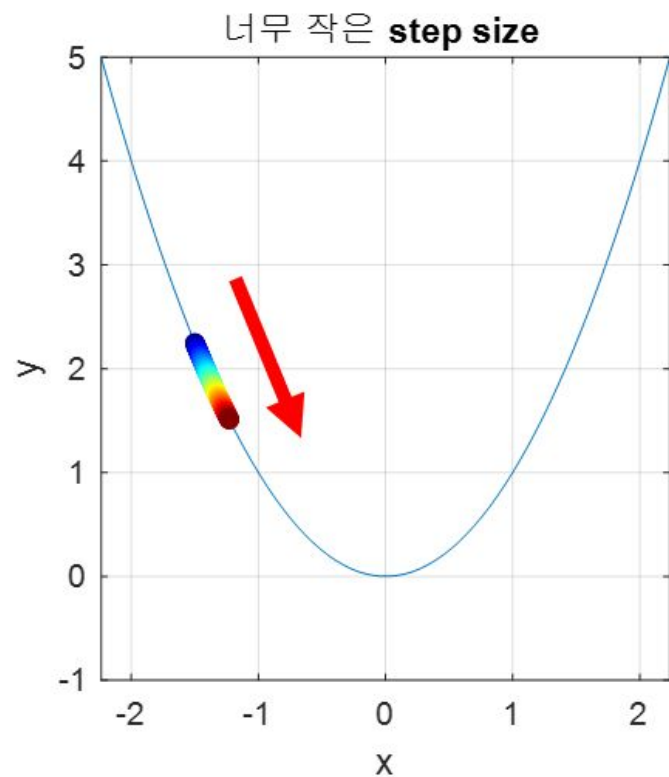


$$W_{t+1} \leftarrow W_t - \eta g_t$$

Learning rate

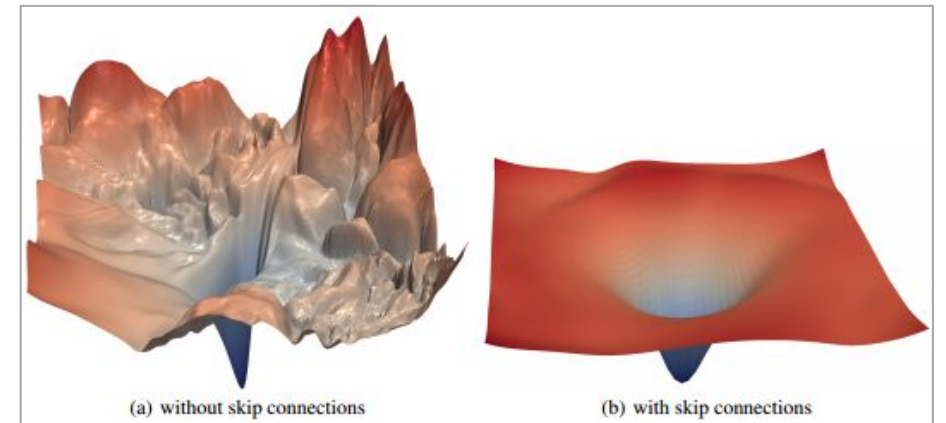
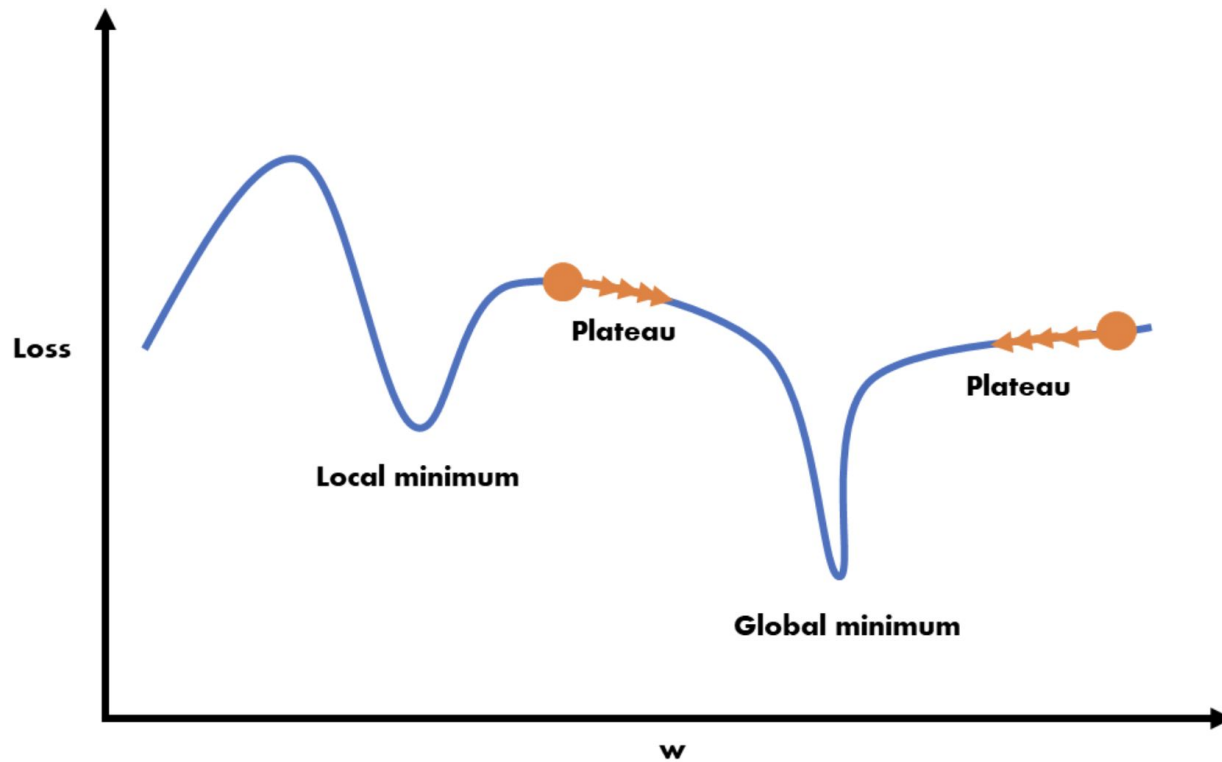
Gradient

학습률에 따른 수렴 여부



경사하강법의 한계점

- Local Minimum에 빠질 수 있음
- Plateau 현상
- 초깃값과 학습률에 영향을 받음



Gradient Descent Methods

모든 자료를 다 검토해서
내 위치의 산기울기를 계산해서
갈 방향을 찾겠다.

GD

전부 다봐야 한걸음은
너무 오래 걸리니까
일부만 보고 빨리 판단한다
짧은 시간에 더 많이 간다

SGD

Momentum

스텝 계산해서 움직인 후,
아까 내려 오던 관성 방향 또 가자

Nesterov Accelerated Gradient

NAG

일단 관성 방향 먼저 움직이고,
움직인 자리에 스텝을 계산하니
더 빠르더라

Nadam

Adam에 Momentum
대신 NAG를 붙이자.

Adam

RMSProp + Momentum
방향도 스텝사이즈도 적절하게!

RMSProp

보폭을 줄이는 건 좋은데
이전 맥락 상황봐가며 하자.

Adagrad

안가본곳은 성큼 빠르게 걸어 훑고
많이 가본 곳은 잘아니까
갈수록 보폭을 줄여 세밀히 탐색

AdaDelta

종종걸음 너무 작아져서
정지하는걸 막아보자.

경사는 하강

실력은 상승

End of Document
Thank You.