Part. 03

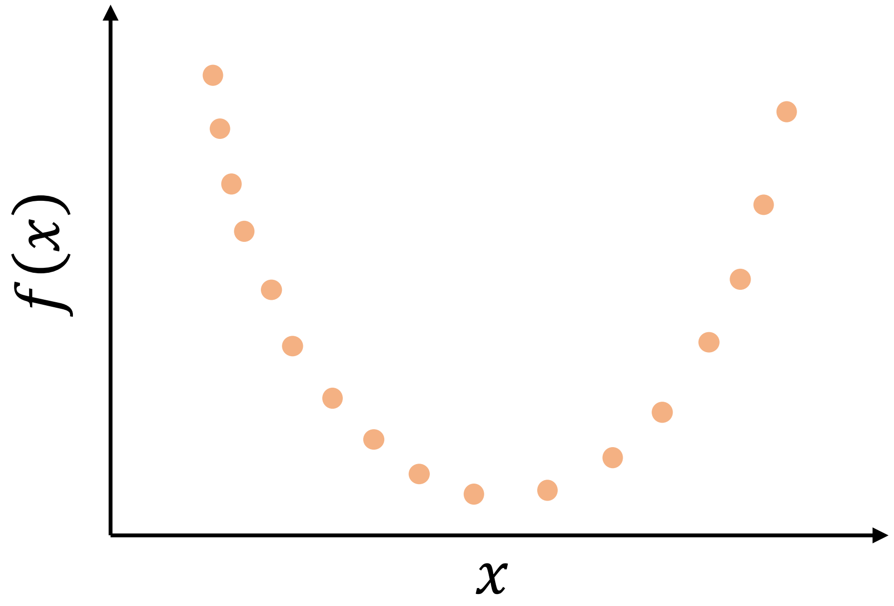
Optimization Algorithms

| Gradient Descent

FASTCAMPUS
ONLINE

강사. 신제용

I 무차별 대입법

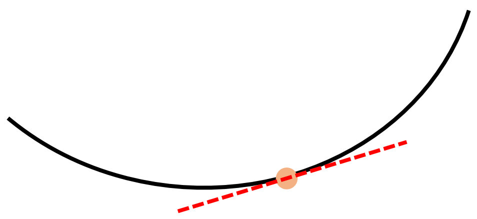


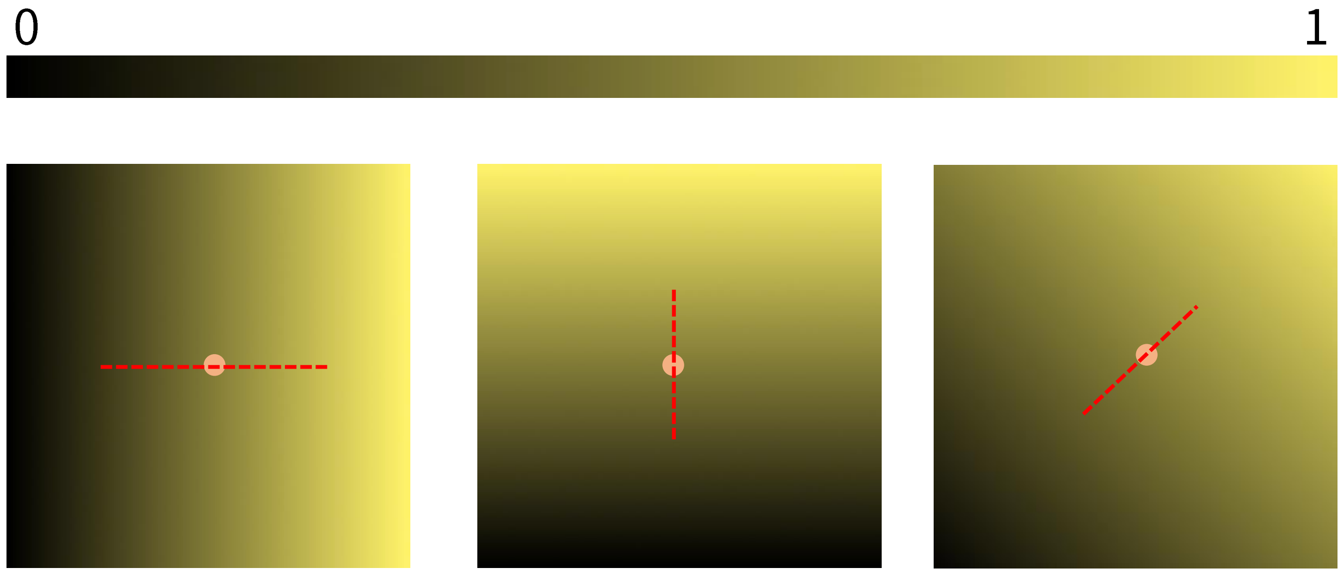
무차별 대입법 (Brute-force)

- 가능한 모든 수를 대입해 보는 방법
- 가장 단순한 방법으로 함수를 알 수 있음
- 다음과 같은 문제로 최적화에 이용할 수 없다.
  - x^*(최적값)이 존재하는 범위를 알아야 함
  - x^*를 정확히 찾기 위해 무한히 촘촘하게 조사해야 함
  - $f(x)$의 계산 복잡도가 매우 높음

적게 대입해 보고 답을 찾을 수는 없을까?

미분과 기울기



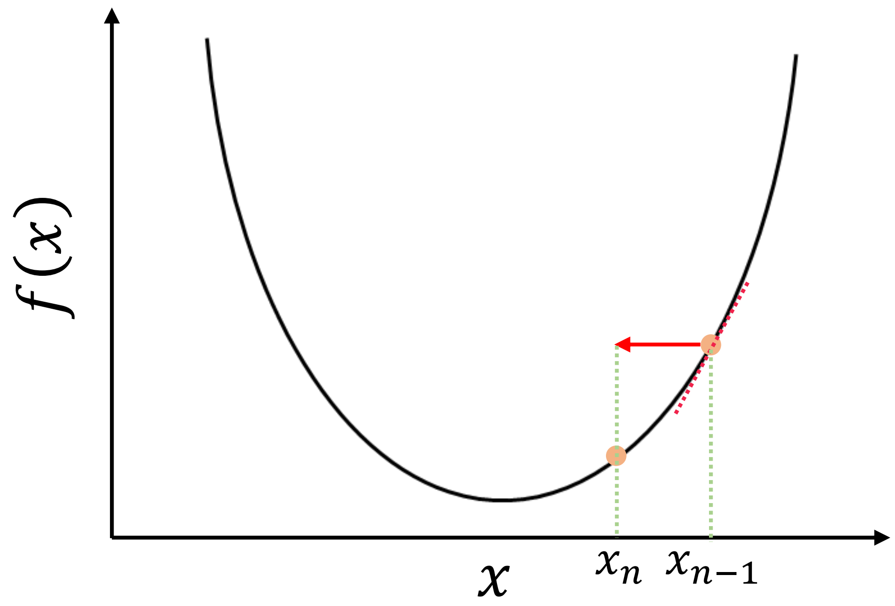


$$\frac{df(x)}{dx} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

$$\nabla f(\boldsymbol{x}) = \left[\frac{df(x_0)}{dx_0}, \frac{df(x_1)}{dx_1}, \dots, \frac{df(x_{N-1})}{dx_{N-1}}\right]^T$$

기울기(Gradient)는 스칼라를 벡터로 미분한 것이며, 벡터의 각 요소로 미분하면 된다.

I 경사 하강법



1-D의 경우

$$x_n = x_{n-1} - \alpha \frac{df(x_{n-1})}{dx}$$

N-D 의 경우

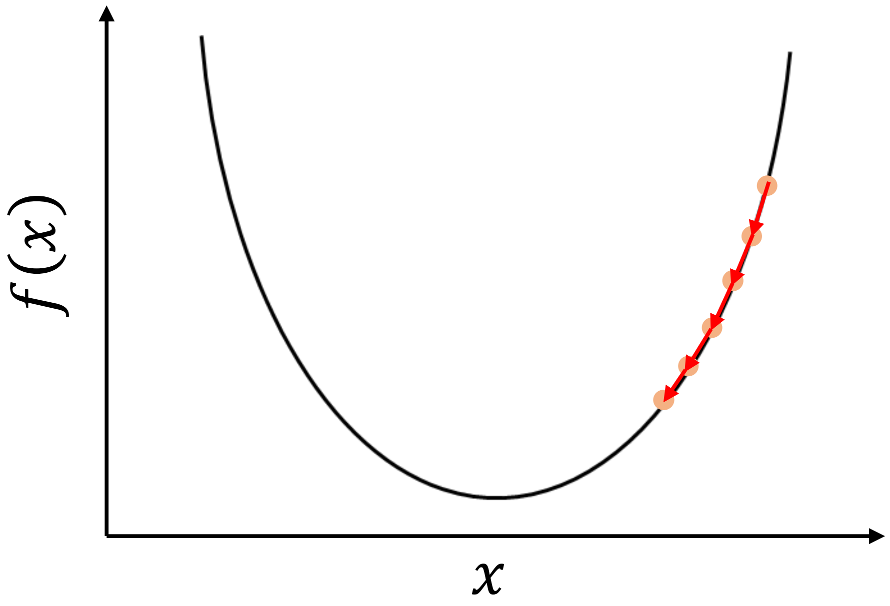
$$\boldsymbol{x}_n = \boldsymbol{x}_{n-1} - \alpha \nabla f(\boldsymbol{x}_{n-1})$$

α: 학습률 (Learning rate)

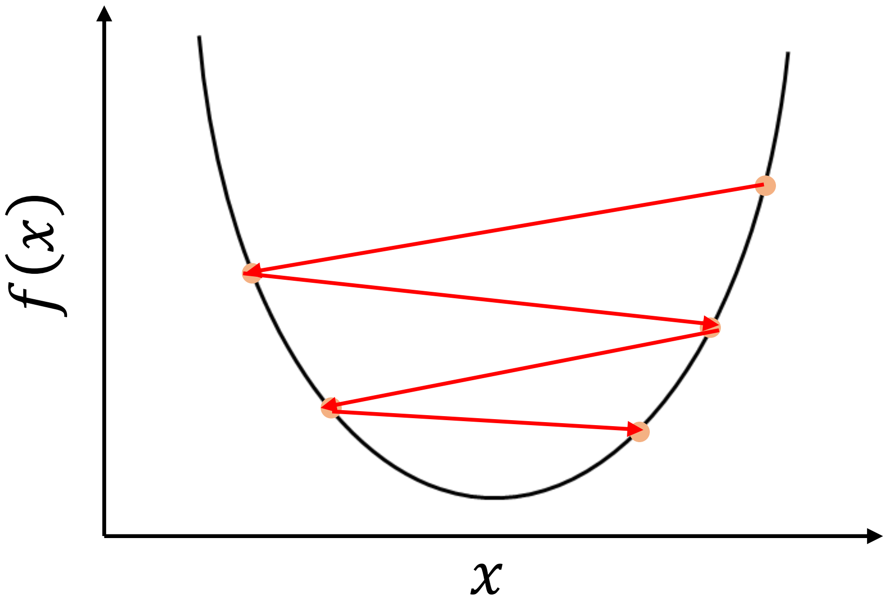
경사 하강법 (Gradient descent)의 한 스텝

경사 하강법은 $f(x)$의 값이 변하지 않을 때 까지 스텝을 반복한다.

I 학습률의 선택



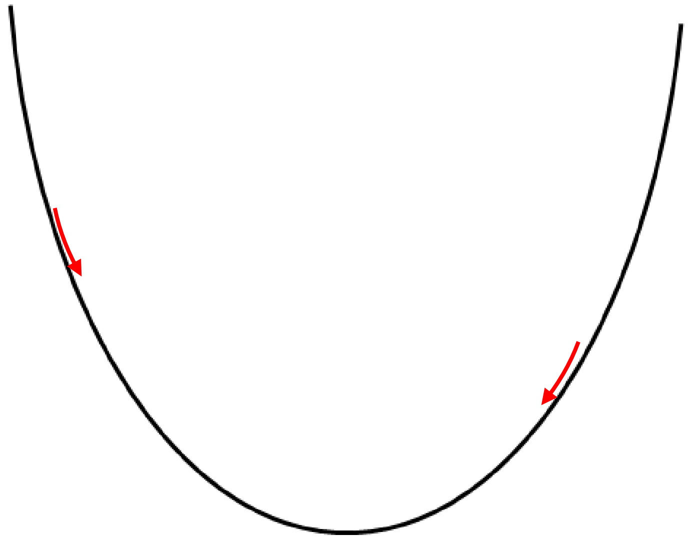
α가 너무 작은 경우



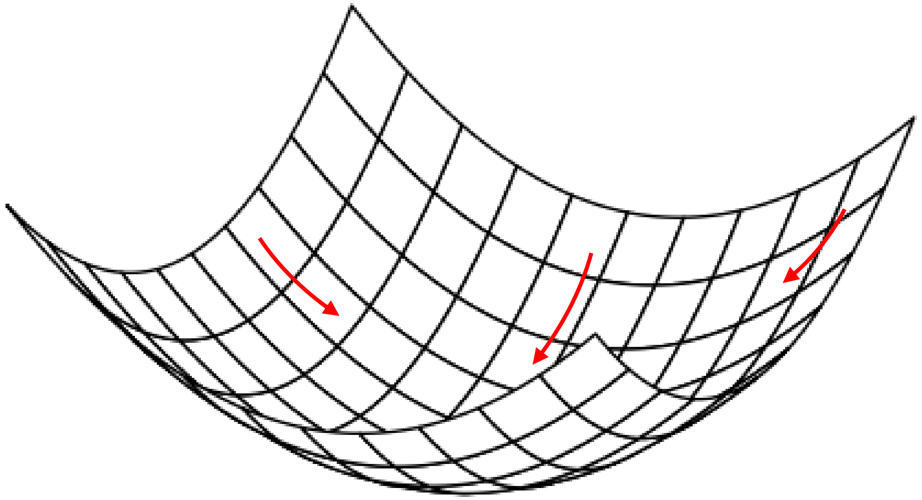
α가 너무 큰 경우

적절한 학습률(Learning rate)을 선택하는 것은 매우 중요하다.

| 볼록 함수



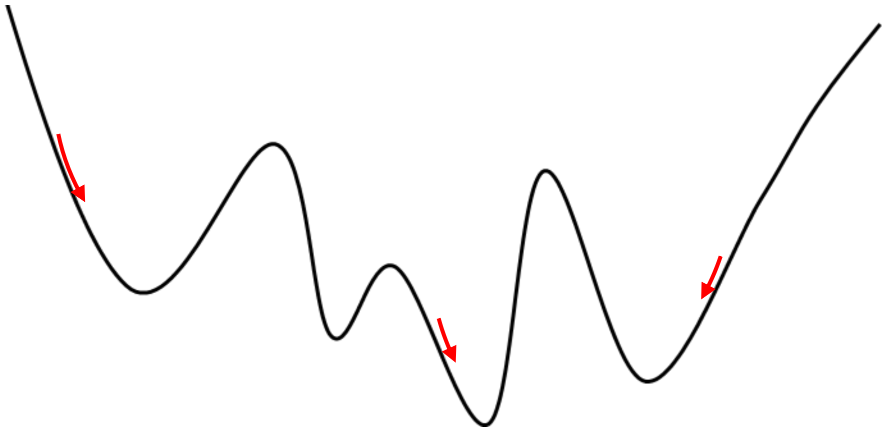
1-D Convex function



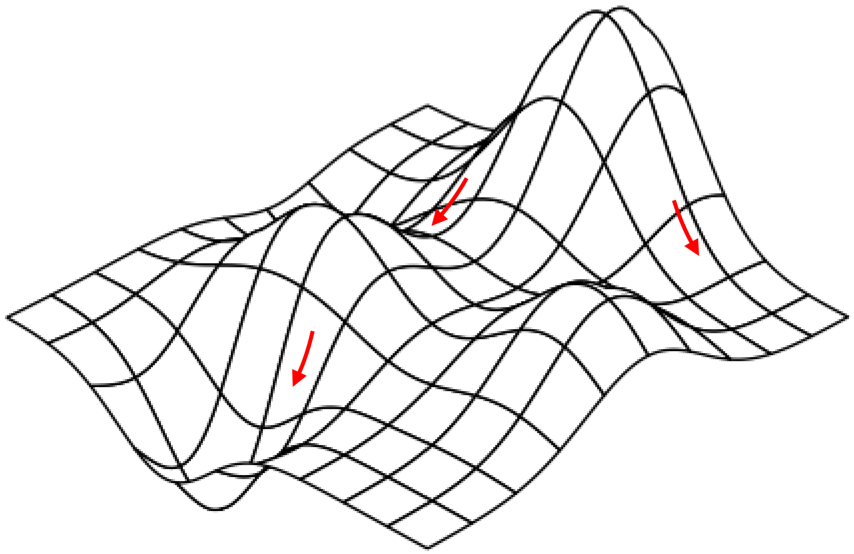
2-D Convex function

볼록 함수(Convex function)는 어디서 시작하더라도 경사 하강법으로 최적 값에 도달할 수 있다.

| 비볼록 함수



1-D Non-convex function



2-D Non-convex function

비볼록 함수(Non-convex function)는 시작 위치에 따라 다른 최적 값을 찾는다.

즉, Local minimum에 빠질 위험이 있다.