

경사 하강법

경사 하강법

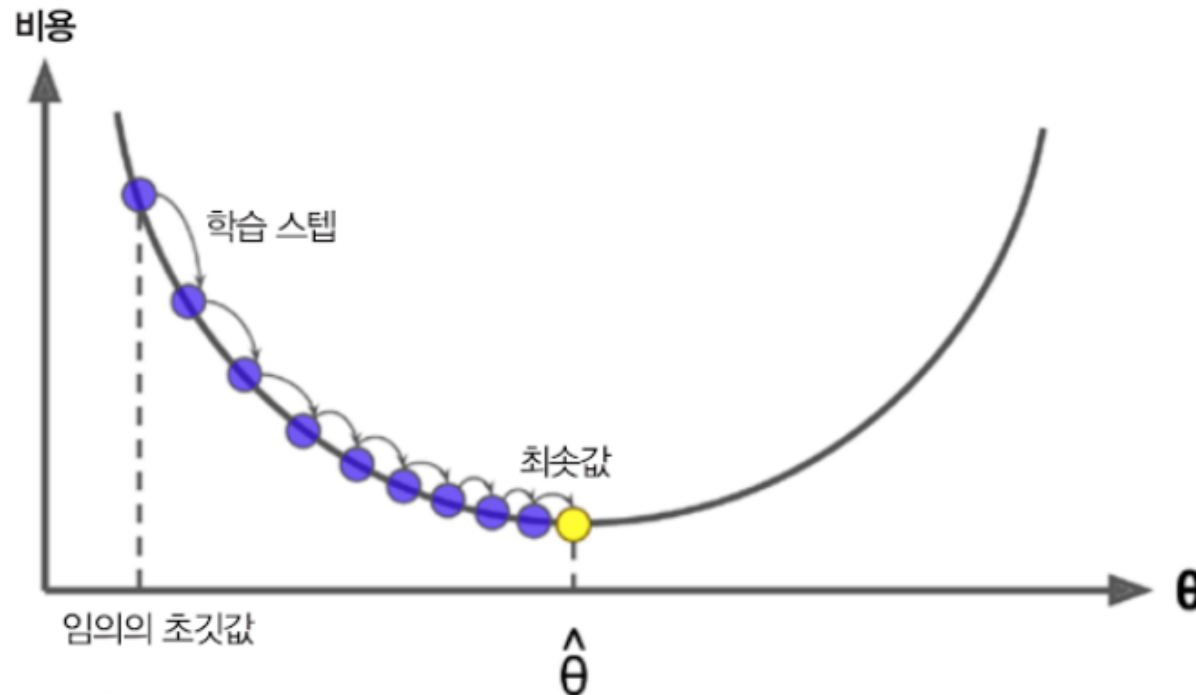
- 경사 하강법 (GD; Gradient Descent)
 - 주어진 문제의 비용 함수의 결과 값을 최소화 하도록 반복하여 매개변수를 조정해 가는 최적화(optimization) 기법 중 하나
 - “늦은 밤에 산 속에서 길을 잃었을 때 가장 좋은 방법은 매 위치에서 가장 가파른 **경사를 따라서 아래로 내려가는 것**”



경사 하강법

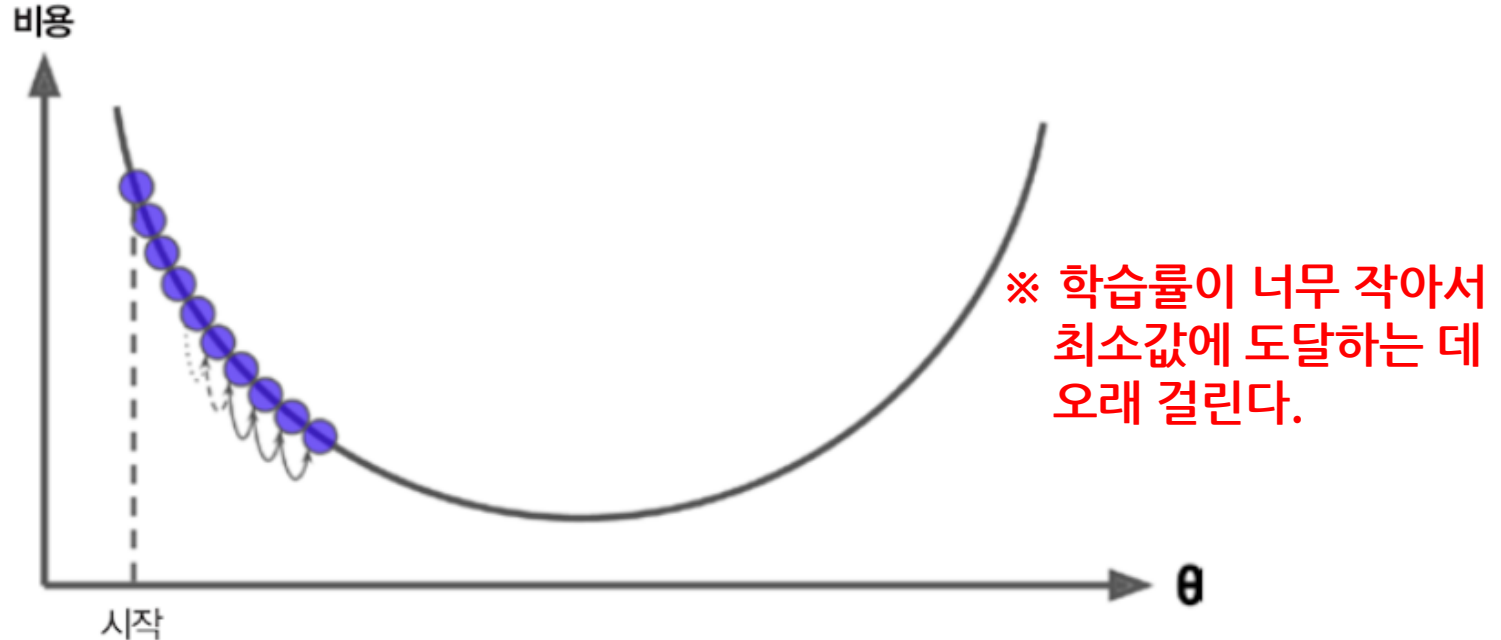
- 경사 하강법

- 변수 θ 에 대해서 비용 함수의 현재 경사(gradient)를 계산하여 비용 함수의 값이 감소되는 방향으로 진행한다.
- 결과적으로 경사가 0이 되면 비용 함수의 값은 최소값에 수렴하게 된다.



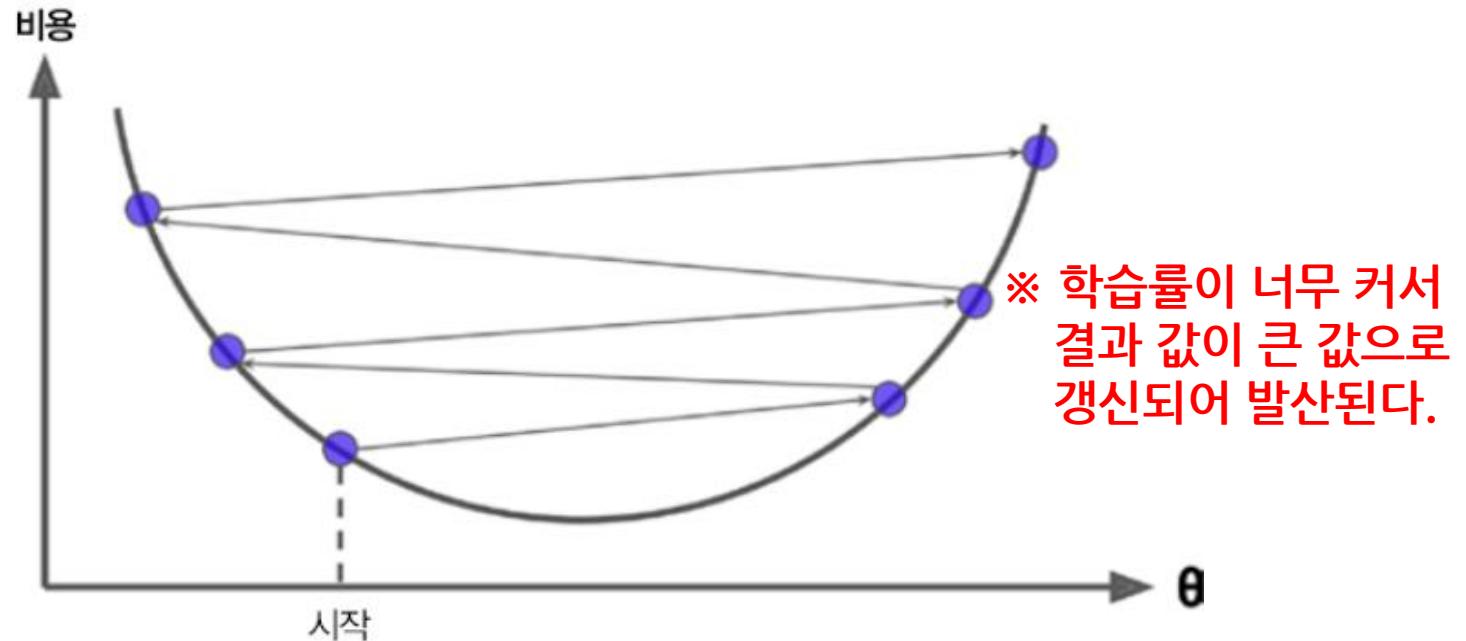
경사 하강법

- 경사 하강법
 - 학습 스텝은 매 계산마다 적용되는 이동 거리에 해당하며, **학습률(learning rate)**이라고 한다.
 - 이 학습률이 너무 작은 경우에는 결과 값이 최소값에 수렴하기까지 계산 반복이 많이 발생한다.



경사 하강법

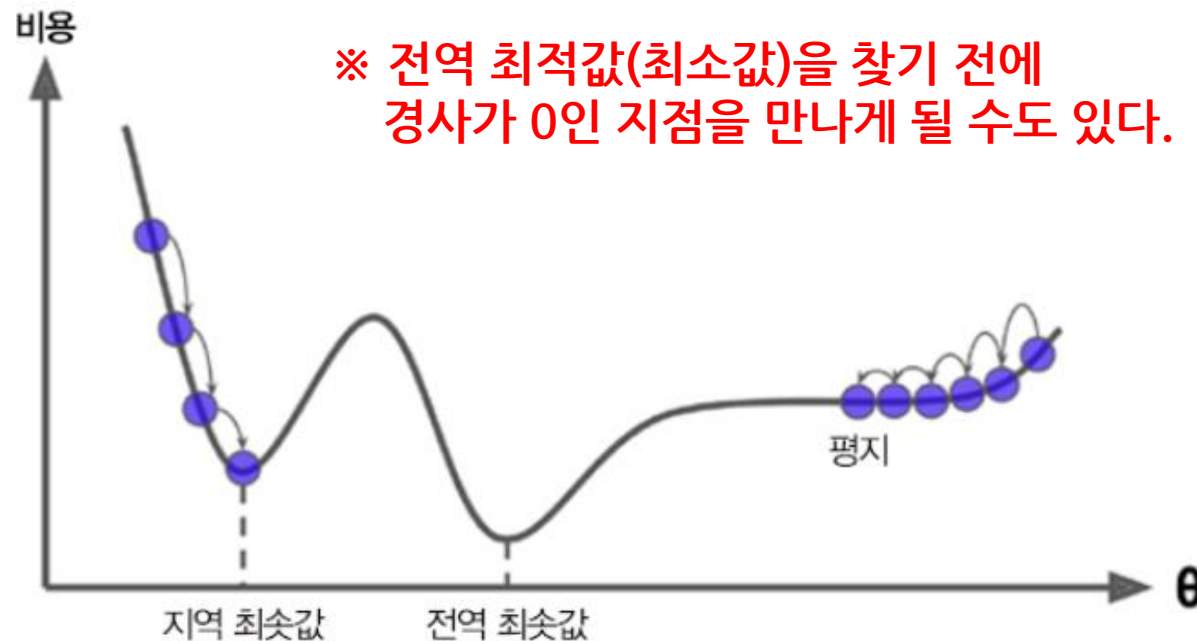
- 경사 하강법
 - 학습 스텝은 매 계산마다 적용되는 이동 거리에 해당하며, **학습률(learning rate)**이라고 한다.
 - 반대로, 학습률이 너무 큰 경우에는 아예 반대쪽으로 건너 뛰어서 잘못된 곳으로 가게 될 수도 있다.



경사 하강법

- 경사 하강법

- 경사 하강법은 전역 최적값(global optimum)을 찾지 못하고 지역 최적값(local optimum)을 구하게 될 수도 있다는 문제가 있다.



경사 하강법

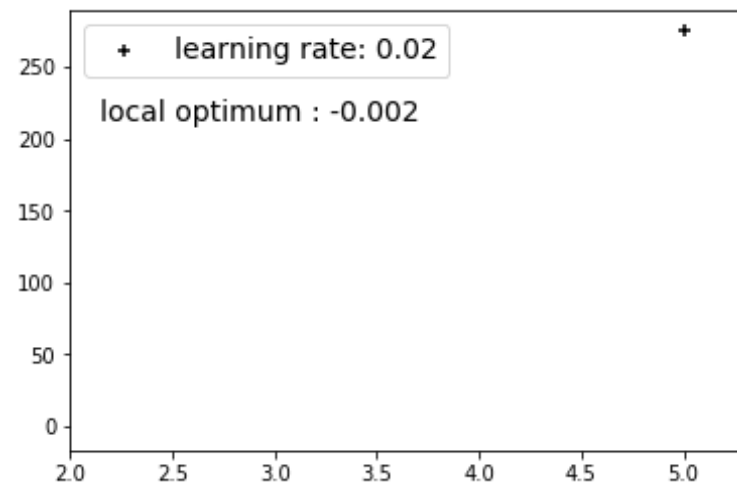
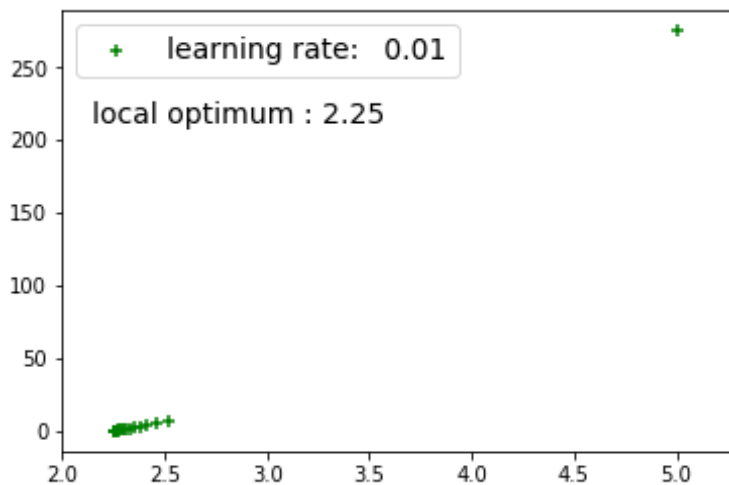
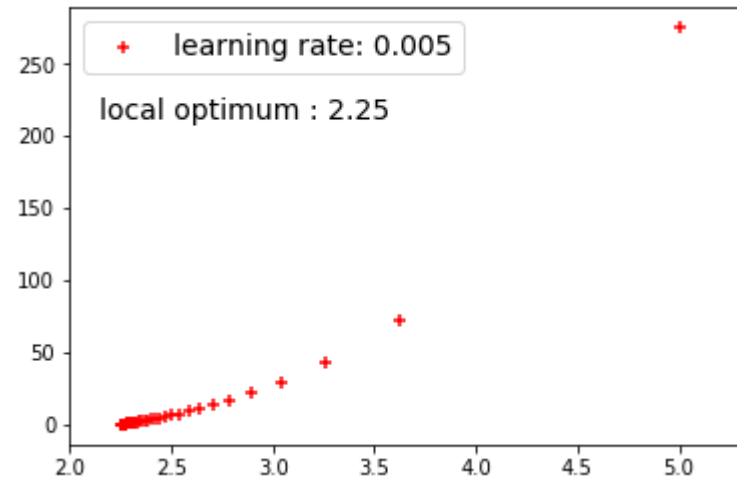
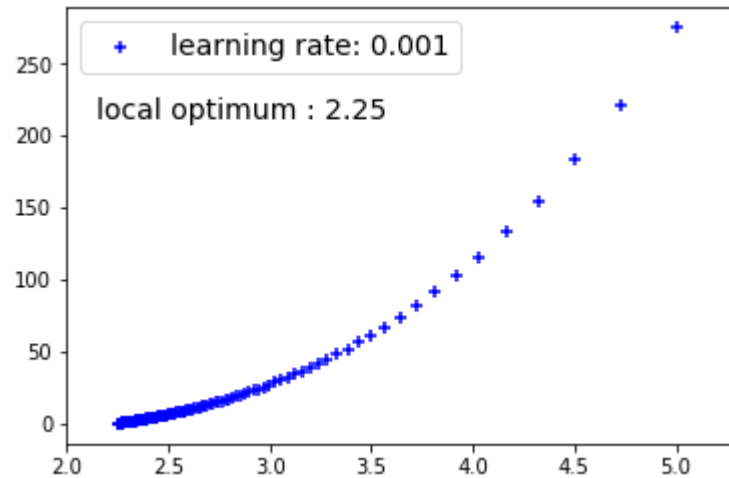
- 경사 하강법의 구현 예

```
1 x_old = 0
2 x_new = 5
3 learning_rate = 0.01
4 precision = 0.000001
5
6 def cost_function(x):
7     return 4 * x**3 - 9 * x**2
8
9 while abs(x_new - x_old) > precision:
10     x_old = x_new
11     x_new = x_old - learning_rate * cost_function(x_old)
12
13 print("지역 최적값 =", x_new)
```

지역 최적값 = 2.25

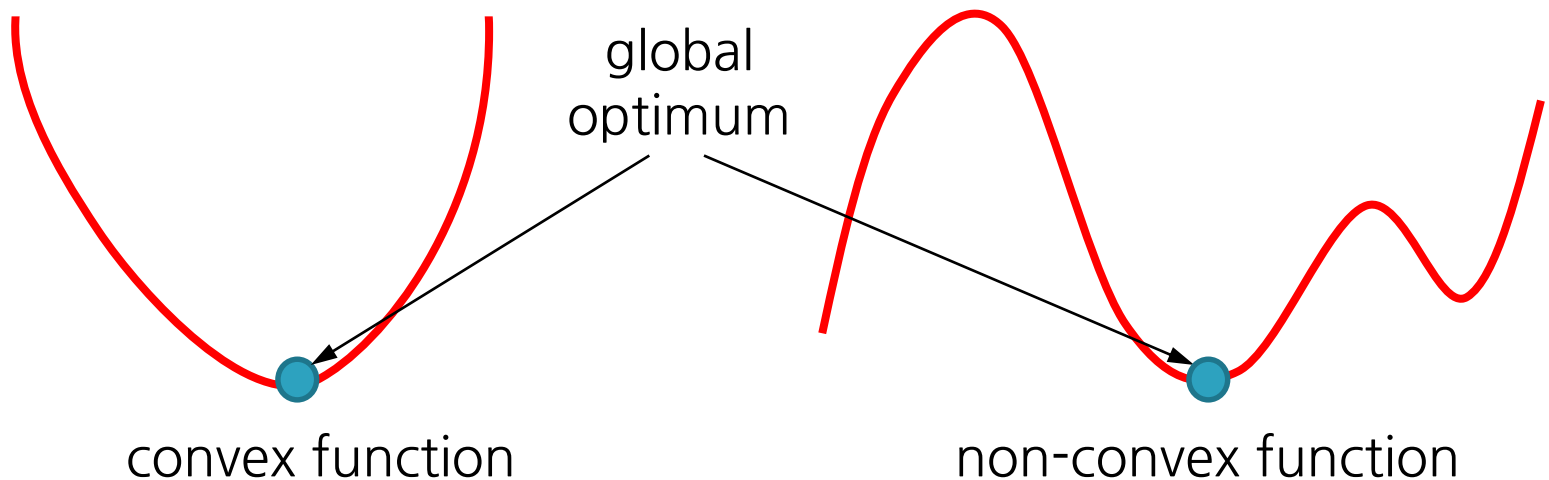
경사 하강법

- 경사 하강법의 구현 예



경사 하강법

- 회귀 분석에서의 경사 하강법
 - 선형 회귀 분석에서의 비용 함수 RSS는 2차 함수이면서 볼록 함수(convex function)이므로, 별도의 지역 최소값이 없고 1개의 전역 최소값만 존재한다.
 - 따라서 선형 회귀에서는 경사 하강법을 이용하여 RSS의 전역 최소값을 구할 수 있다.



경사 하강법

- 경사 하강법의 유형

- 경사 하강법은 학습 데이터를 어떻게 할당하는가에 따라서 크게 3종류로 구분할 수 있다.

- ① 배치 경사 하강법 (batch gradient descent)

- 경사를 1회 계산하기 위해서 전체 학습 데이터를 사용

- ② 확률적 경사 하강법 (SGD; stochastic gradient descent)

- 경사를 1회 계산하기 위해서 1개의 학습 데이터를 사용

- ③ 미니배치 경사 하강법 (mini-batch gradient descent)

- 경사를 1회 계산하기 위해서 일부 학습 데이터를 사용

경사 하강법

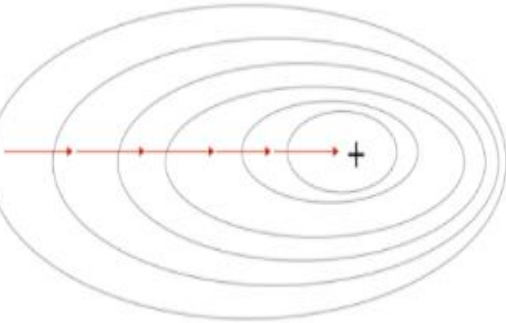
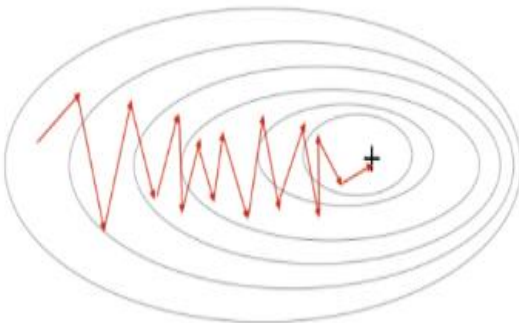
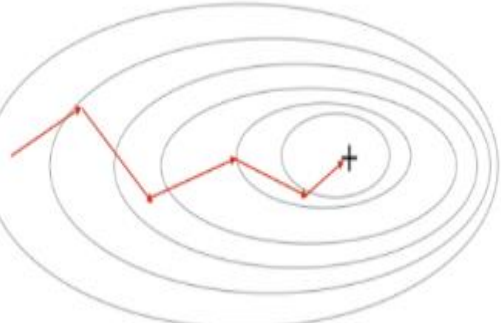
- 배치 (Batch)
 - 1회의 경사 업데이트에 사용되는 데이터 집합
 - 이 때 사용되는 데이터 집합의 개수를 배치 크기라고 한다.
예) 전체 데이터가 100개 있을 때 배치 크기가 20이면 :
1개의 배치마다 20개의 데이터가 있다.
배치는 총 5개가 있고, 경사는 5회 업데이트된다.
- 에폭 (Epoch)
 - 전체 데이터들을 한 번 사용한 횟수, 즉 학습의 반복 횟수
예) 데이터 100개, 배치가 5개일 때 에폭이 1000이면 :
학습이 총 1000회 수행된다.
경사는 총 5000회 업데이트된다.

경사 하강법

- 경사 하강법의 유형 다시 표현하기
 - 경사 하강법은 배치 크기와 적용에 따라서 크게 3종류로 구분할 수 있다.
- ① 배치 경사 하강법 (batch gradient descent)
 - 1개의 배치에 전체 학습 데이터가 모두 들어간다.
- ② 확률적 경사 하강법 (SGD; stochastic gradient descent)
 - 1개의 배치에 임의의 학습 데이터 1개만 들어간다.
- ③ 미니배치 경사 하강법 (mini-batch gradient descent)
 - 1개의 배치에 임의의 학습 데이터 여러 개가 들어간다.

경사 하강법

- 경사 하강법의 비교

배치 경사 하강법	확률적 경사 하강법	미니배치 경사 하강법
		
<p>일관된 방향을 향해 지속적으로 근접한다.</p> <p>데이터가 많으면 수행 시간이 오래 걸린다.</p>	<p>데이터가 많아도 수행 시간이 빠르다.</p> <p>무작위 선택에 의해서 노이즈가 많을 수 있고 데이터 편중이 있을 수 있다.</p>	<p>앞의 두 가지 방법의 절충안에 해당한다.</p> <p>전체 배치보다는 효율적이고, 확률적 경사 하강법보다는 노이즈가 적다.</p>