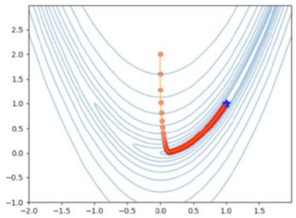
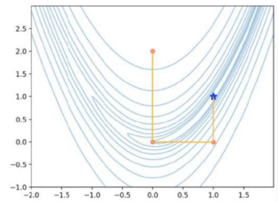
STEP29. 뉴턴 방법으로 푸는 최적화 (수동 계산)

- 경사하강법 (일반적으로 수렴이 느림)



- 뉴턴 방법



- 뉴턴 방법으로 최적화하면 더 적은 단계로 최적의 결과를 얻을 가능성이 높음
- 경사하강법은 계곡에서 서서히 최솟값에 접근해감
- **뉴턴 방법은 계곡을 뛰어넘어 단번에 목적지에 도착**함
- 경사하강법은 5만번, 뉴턴 방법은 6회 갱신만에 도달함
- **갱신 횟수는 초깃값이나 학습률 등의 설정에 따라 크게 좌우됨**
- 일반적으로 초깃값이 **정답에 충분히 가까우면 뉴턴 방법이 더 빨리 수렴함**

뉴턴 방법을 활용한 최적화 이론

뉴턴 방법의 최적화 원리

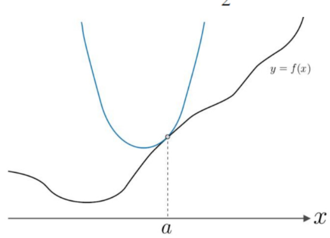
- y = f(x) 라는 함수의 최솟값을 구하는 문제
- 뉴턴 방법으로 최적화하려면 테일러 급수에 따라 y = f(x) 를 변환

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \frac{1}{3!}f'''(a)(x-a)^3 + \cdots$$

- 테일러 급수에 따라 어떤 점 a를 기점으로 f를 x의 다항식으로 나타낼 수 있음
- 증가하는 걸 어느 시점에서 중단하면 f(x)를 근사적으로 나타낼 수 있음

- **2차 미분에서 중단 (2차까지 테일러 급수로 근사)**

$$f(x) \simeq f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2$$



- 근사한 2차 함수는 a에서 y = f(x)에 접하는 곡선임

근사한 2차 함수의 최솟값 구하기

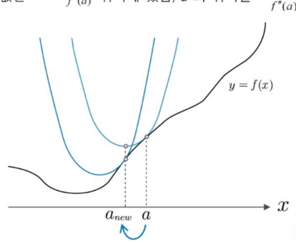
- 2차 함수의 최솟값은 해석적으로 구할 수 있음
- 2차 함수의 미분 결과가 0인 위치를 확인하면 됨

$$\frac{d}{dx}\left(f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2\right) = 0$$

$$f'(a) + f''(a)(x-a) = 0$$

$$x = a - \frac{f'(a)}{f''(a)}$$

- 근사한 2차 함수의 최솟값은 $x = a - \frac{f'(a)}{f''(a)}$ 위치에 있음, a 의 위치를 - $\frac{f'(a)}{f''(a)}$ 만큼 갱신하면 됨



- 갱신된 a 의 위치에서 같은 작업을 반복함

뉴턴방법 vs 경사하강법

경사하강법

$$x \leftarrow x - \alpha f'(x)$$

- 경사하강법
  - α 계수를 사람이 수동으로 결정
  - α 의 값만큼 1차미분 진행하여 x 값을 갱신

뉴턴방법

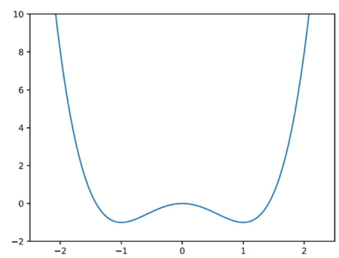
$$x \leftarrow x - \frac{f'(x)}{f''(x)}$$

- 뉴턴 방법
  - 2차 미분을 이용하여 경사 하강법에서 말하는 α를 자동으로 조정

뉴턴 방법 정리

- 경사하강법은 1차 미분만의 정보를 사용함
- 뉴턴 방법 최적화는 2차 미분의 정보도 이용함
- 물리 세계로 비유하면 경사하강법은 속도 정보만 사용하고, 뉴턴은 속도와 가속도 정보 사용
- 뉴턴 방법이 추가된 2차 미분으로 효율적으로 탐색을 기대할수 있음
- 결과적으로 목적지에 더 빨리 도달할 확률이 커짐

뉴턴 방법을 이용한 구체적인 문제 풀이



- y = x^4 + -2x^2 수식의 최적화

- 오목한 부분이 두 곳이며, 최솟값은 x가 각각 -1과 1인 위치
- 초깃값을 x = 2로 설정한 후 최솟값 중 하나인 x = 1에 도달하는지 검증

뉴턴 방법을 활용한 최적화 구현

1. 뉴턴방법 구현

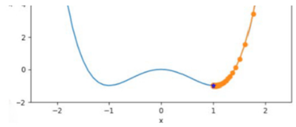
```python
import numpy as np
from dezero import Variable

def f(x):
    y = x ** 4 - 2 * x ** 2
    return y

def gx2(x):
    return 12 * x ** 2 - 4

x = Variable(np.array(2.0))
iters = 10

for i in range(iters):
    print(1, x)

    y = f(x)
    x.cleargrad()
    y.backward()

    x.data -= x.grad / gx2(x.data)
```

- DeZero는 2차 미분을 자동으로 구하지 못하므로 수동으로 2차 미분을 구함
- 차 미분은 역전파로 구하고 2차 미분은 수동으로 코딩해 구함
- 뉴턴 방법의 갱신 수식에 따라 x를 갱신함
- 문제의 답인 최솟값은 1임
- 뉴턴 방법은 7회의 갱신 만으로 최솟값에 도달함
- 경사하강법은 학습률이 0.01로 진행시 절대 오차가 0.001 이하로 좁혀지기까지 124번 갱신

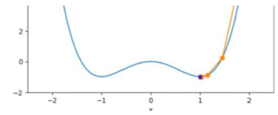
경사하강법의 갱신 경로와 뉴턴 방법을 활용한 최적화 기법의 갱신 경로

경사하강법

뉴턴방법



2
0
−2
−2
−1
0
1
2
x



2
0
−2
−2
−1
0
1
2