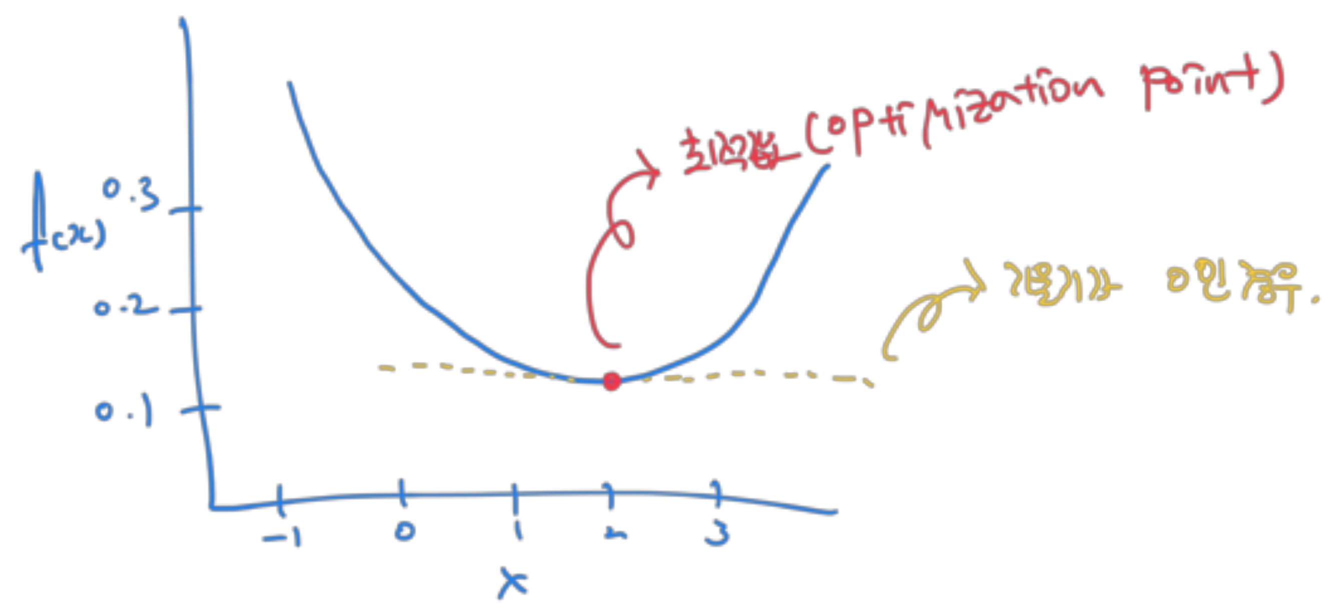
최적화

- 그리드 서치와 최적화
  - 그리드 서치는 목적함수에 대한 모든 변수를 시도해서 최적값을 찾는다
  - 실제로 많은 비용이 소모된다.



- 기울기가 0인 부분은 도함수 ($\frac{df}{dx}$) 가 0인 부분을 의미.

단일변수에 대해서...

$$\frac{df(x)}{dx} = 0$$

다변수에 대해서... (편미분)

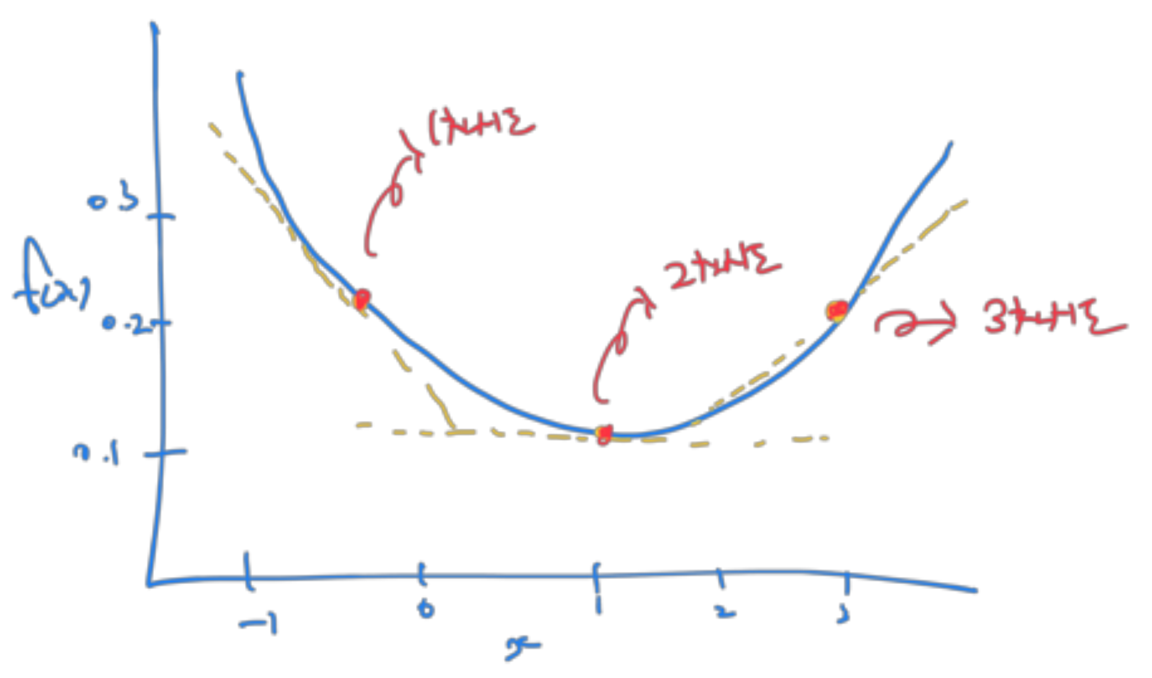
$$\frac{\partial f(x_1, x_2, x_3 \cdots x_n)}{\partial x_1} = 0$$

$$\frac{\partial f(x_1, x_2, x_3, \cdots x_n)}{\partial x_2} = 0$$

$$\vdots$$

$$\frac{\partial f(x_1, x_2, x_3 \cdots x_n)}{\partial x_n} = 0$$

- SGD (Steepest Gradient Descent)
  - step 사이즈를 조절하며 기울기로 간다.



Step 사이즈가 크면, 최적해를 찾기가 어려워 진다.

- 2차 함수에서는 스텝 사이즈에 따라서 계곡 현상 및 진동현상이 발생할 수 있다.
- 진동현상을 없앨수 있는 방법은 헤세안 행렬을 이용하거나 모멘텀 방법을 이용한다.

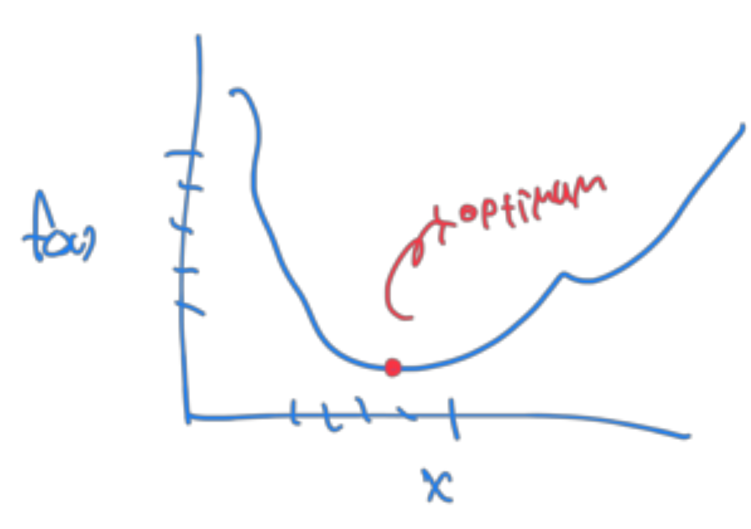
- 2차 도함수를 이용한 Newton 방법
  - 목적함수가 2차 함수라는 가정하에 한번에 최저점을 찾는 방법
  - 그레디언트 벡터에 헤세안 행렬의 역행렬을 곱해서 방향과 거리가 변형된 그레디언트 값을 이용한다.

$$x_{n+1} = x_n - [H f(x)]^{-1} \Delta f(x_n)$$

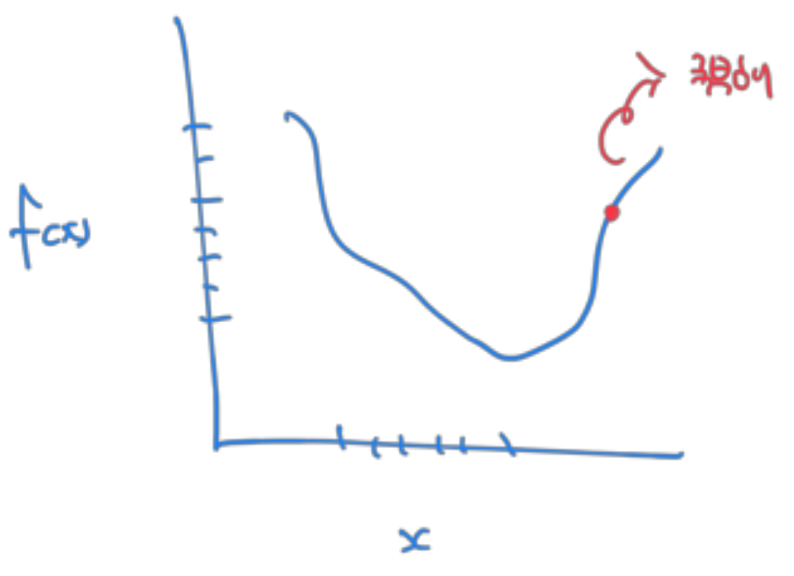
  - 스텝사이즈가 필요없고 목적함수가 2차 함수와 비슷한 모양이면 한번에 수렴한다는 장점이 있지만, 1차 도함수 (그라디언트 벡터) 뿐만 아니라 2차 도함수 (헤세안 행렬)도 함께 필요하다.

- Quasi-Newton 방법
  - Newton 방법이 헤세안 행렬을 함께 구해줘야 하는 문제가 있다
  - 사람이 구한 헤세안 행렬을 사용하는 대신에 현재 시도하고 있는 x_n 주변의 몇몇 점에서 함수의 값을 구하고 이를 이용하여 2차 도함수의 근사값을 수치적으로 계산하여 이용한다. (BFGS - Broyden - Fletcher - Goldfarb - Shanno)

- 전역최적화 문제 (Global optimization problem)



=> 잘 찾아지는 경우



=> 국부해에 빠지는 경우.

- Convex 문제
  - 목적함수의 2차 도함수의 값이 항상 0 이상이 되는 영역에서만 정의된 최적화 문제

$$\frac{\partial^2 f}{\partial x^2} \geq 0$$

  - Convex 문제에서는 유일한 전역최저점이 존재한다.