II. 이차 미분 근사법

- 일변량 : 이차 도함수
  - Newton 방법 : 통계학에서 AA 사용 , Fisher 의 scoring 방법
    일반화 선형 모형에서의 최대 우도 추정값 얻는데 사용
  - Secant 방법

- 다변량 : 헷세 행렬 (분산 · 공분산 행렬과 유사!)
  - Newton 방법
  - 준 (Quasi) Newton 방법 : DFP, BFGS

이차 미분 근사법

- 일변량 : Newton 방법

◦ 이차 Taylor 전개

: $q(x) = f(x^{(k)}) + (x-x^{(k)}) f'(x^{(k)}) + \frac{(x-x^{(k)})^2}{2} f''(x^{(k)})$

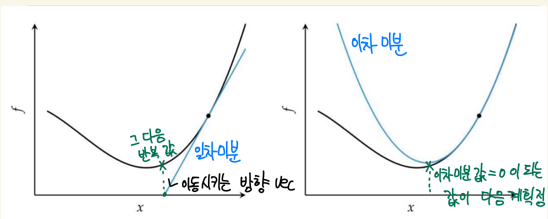
→ 최소값은 도함수의 기울기 = 0 미분해 보자!

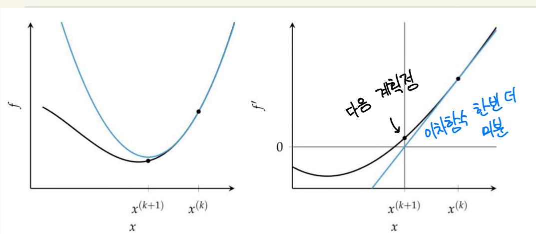
◦ 이차 Taylor 전개 도함수의 근

$$\frac{\partial q(x)}{\partial x} = f'(x^{(k)}) + (x-x^{(k)}) f''(x^{(k)}) = 0$$

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$$

이차 미분 값이 계산되어야 함.





→ 2차 미분 정보 이용해 기울기 = 0 인 값으로 이동

일변량 문제에서 Newton 법은 bisection Method와 유사.

↳ 1차 미분 함수에 다시 이분해 구간 줄여가며 계산

⇒ 반복수 적어짐. 이차 미분 정보량 많아졌기 때문!

- 일변량 : Secant 방법

• $f''(x^{(k)}) \approx \frac{f'(x^{(k)}) - f'(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}$: 이차 도함수를 수치 미분값으로 대체

• $x^{(k+1)} \leftarrow x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f'(x^{(k)}) - f'(x^{(k-1)})} \cdot f'(x^{(k)})$

↳ 나눠줘야 하니까 역수!

⇒ k번째 계획점 얻기 위해서 k, (k-1)번째 계획점 알아야 함

초기값 2개. x^{k}, $x^{(k-1)}$

- 다변량 : Newton 방법

- $g = \nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n} \right]$: 경사도

- $H = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{bmatrix}$: 헷세 행렬

→ $\frac{\text{1차 미분}}{\text{2차 미분}}$: 행렬 나누기 불가능 , 역행렬 구해야 함

2차 테일러 전개

- $q(x) = \underbrace{f(x^{(k)})} + (\underbrace{g^{(k)}}_{\text{상수}})^T (x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^T H^{(k)} (x - x^{(k)})$

$\nabla q(x^{(k)}) = g^{(k)} + H^{(k)}(x - x^{(k)}) = 0$ ← 2차 테일러 전개식 미분후 0이 되는 값

$x^{(k+1)} = x^{(k)} - \underbrace{(H^{(k)})^{-1}}_{\text{information 행렬}} g^{(k)}$

⇒ 고전 방법 中 Newton 법이 가장 빨리 최저점 찾아줌.

H 행렬 커지면 Newton 반복 어려움 :(

- Quasi - Newton 방법

- $x^{(k+1)} \leftarrow x^{(k)} - \underbrace{\alpha^{(k)} Q^{(k)}}_{\text{Hessian 행렬 역할.}} g^{(k)}$ — gradient discent와 유사한 형식

$\alpha^{(k)}$: 단계값 (or 학습률 고정)
$Q^{(k)}$: Hessian 행렬의 근사행렬

- DFP : Davidon - Fletcher - Powell
- BFGS : Broyden - Fletcher - Goldfrab - Shanno

⇒ DFP, BFGS 모두 Q 행렬 파악법