12. 확률적인 방법들

(Stochastic Methods)

- Saddle point

  : 경사도가 '0'에 가까운 계획점

갇혀버리는 문제!

Global min 있는데도 불구,
Local min 찾음

↓

MCMC 방법 적용

모의 실험

- Noisy Descent

  : 난수 이용한 Gradient Descent

- Simulated Annealing

  : 경사도 이용하지 않는 임의 탐색 방법

- Noisy Descent

- $x^{(k+1)} = x^{(k)} + \alpha g^{(k)} + \varepsilon^{(k)}$
  난수를 통해 계획점에 파동 주기

  → $\varepsilon_\tau \sim N(0, \sigma^2)$

  → $\sigma^{(k)} = \frac{1}{k}$: k가 커지면 ε_τ는 0에 가까운 값!

⇒ Saddle point는 gradient $(g^{(k)})$가 0에 가까운 모형

Saddle point 에서 계획점 갱신 작게됨 ($\because \alpha < 1$)

→ 최적점에 다가가기 어렵다.

- Simulated Annealing

- 완전 임의 탐색법 : gradient 이용X . 난수 이용해 최적점 갱신
  MCMC 이용(o)
- 반복에 따라 탐색 공간의 범위를 '온도' 이용해 조절
- 계획점 이동을 위해 특정 확률 분포의 난수 이용
- $\Delta y = f(x') - f(x)$, t: 온도 → Δy가 음수여야 이동
  이동된 점

이동확률 $\begin{cases} 1 & (\text{if } \Delta y \le 0) \\ \min\left(e^{-\frac{\Delta y}{t}}, 1\right) & (\text{if } \Delta y > 0) \end{cases}$

계산 불안정.

- 온도 조절

  1) logarithmic annealing schedule

  : $t^{(k)} = \frac{t^{(1)} \cdot \ln(2)}{\ln(k+1)}$, $t^{(1)}$은 t의 초기값

  → 계산 더뎌짐 :(

  2) exponetial annealing schedule

  : $t^{(k)} = \gamma t^{(k)}$ ← $\gamma \approx 1$이면 줄어드는 변화량 적어짐

  3) fast annealing : 빠르게 이동시킴!

  : $t^{(k)} = \frac{t^{(0)}}{k}$

⇒ 복잡한 목적함수에도 쉽게 찾아줌