

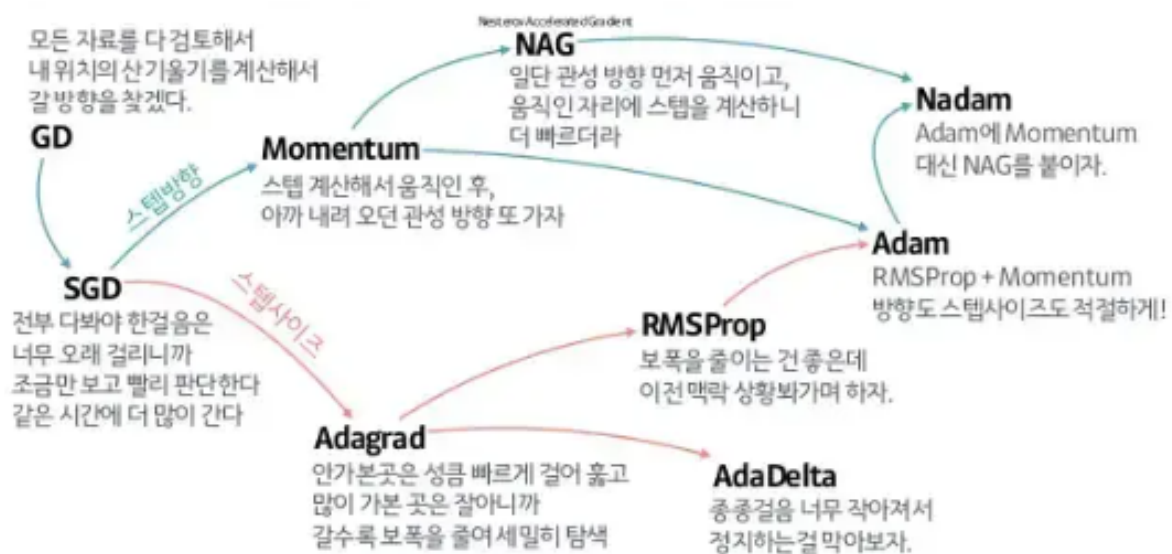
Momentum

≡ Type

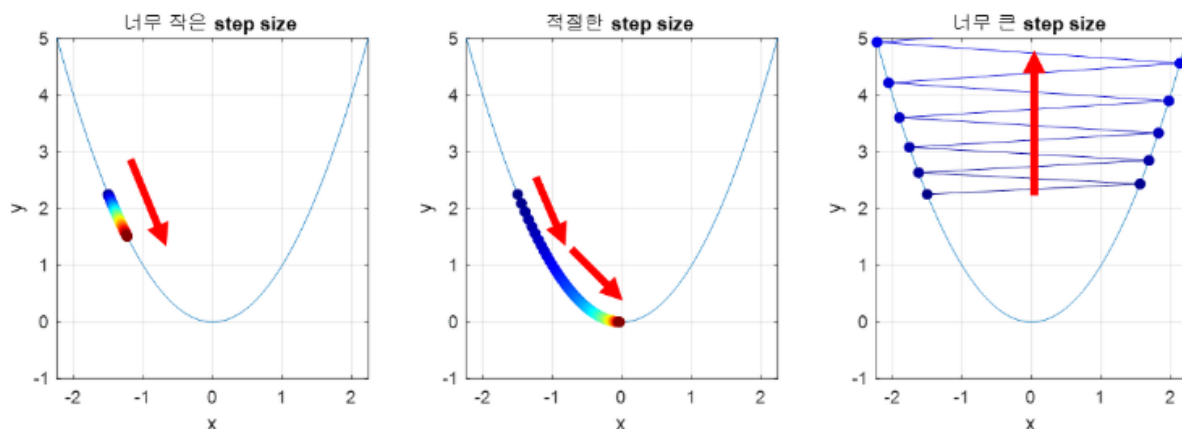
Optimization

Outline

SGD의 높은 편차를 줄이고 수렴을 부드럽게 하기 위해 고안



What is GD (SGD)?



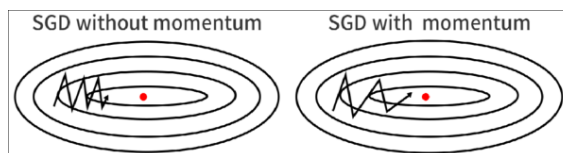
Limits

- Local Minimum
 - Oscillation
-

What is Momentum?

- 경사 하강법으로 이동할 때 관성을 부여하는 최적화 기법
- SGD의 높은 편차를 줄이고 수렴을 부드럽게 하기 위해

Math



- v_t : t 번째 time step에서 x 의 이동 벡터
- γ : 관성계수(momentum term) ≈ 0.9
- η : 학습률(Learning rate)

$$v_t = \gamma v_{t-1} + \eta \nabla f(x_{t-1})$$

- Momentum은 이전에 이동했던 방향을 기억하면서 이전 기울기의 크기를 고려하여 어느 정도 추가로 이동
- 이전 gradient들의 영향력을 매 업데이트마다 배 씩 감소
- momentum term 는 보통 0.9 정도로 정함

Limits

- 학습률 설정
- 관성계수 설정