

Part. 05

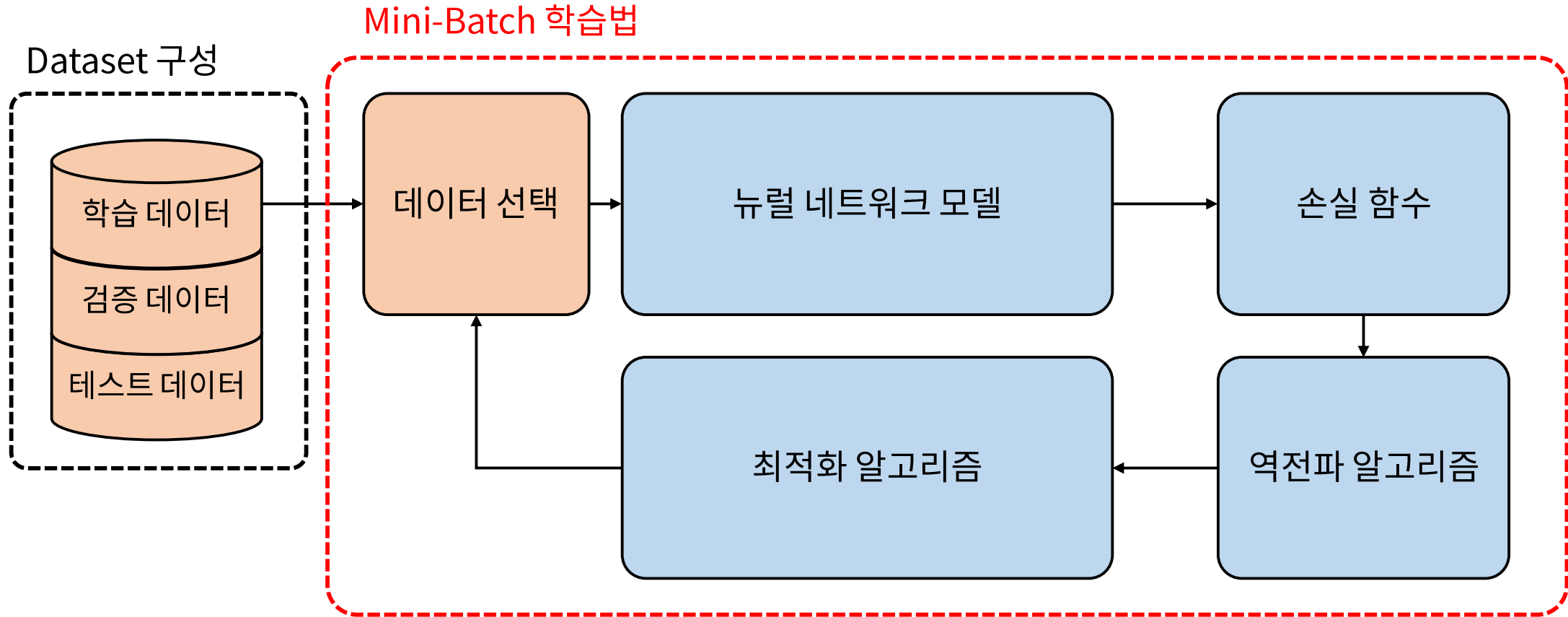
Training Neural Network

| Mini-Batch 학습법

FASTCAMPUS
ONLINE

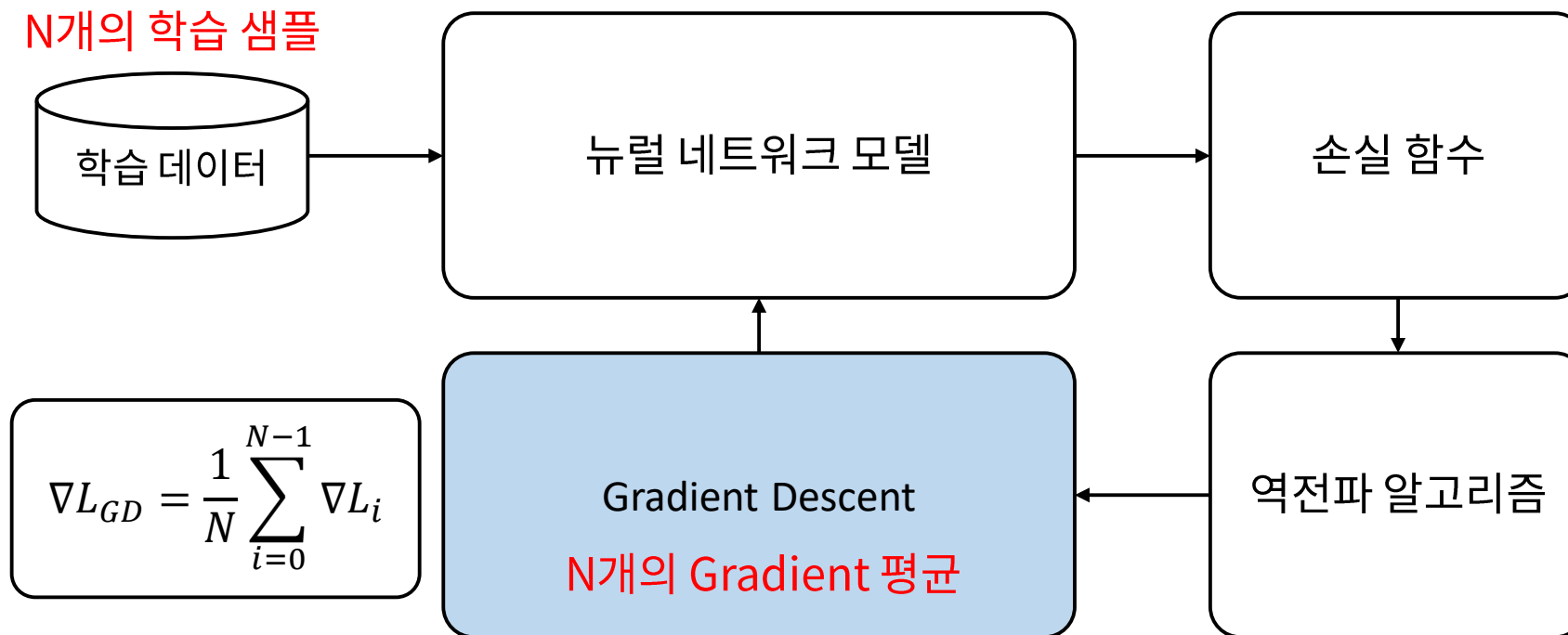
강사. 신제용

I 뉴럴 네트워크의 학습



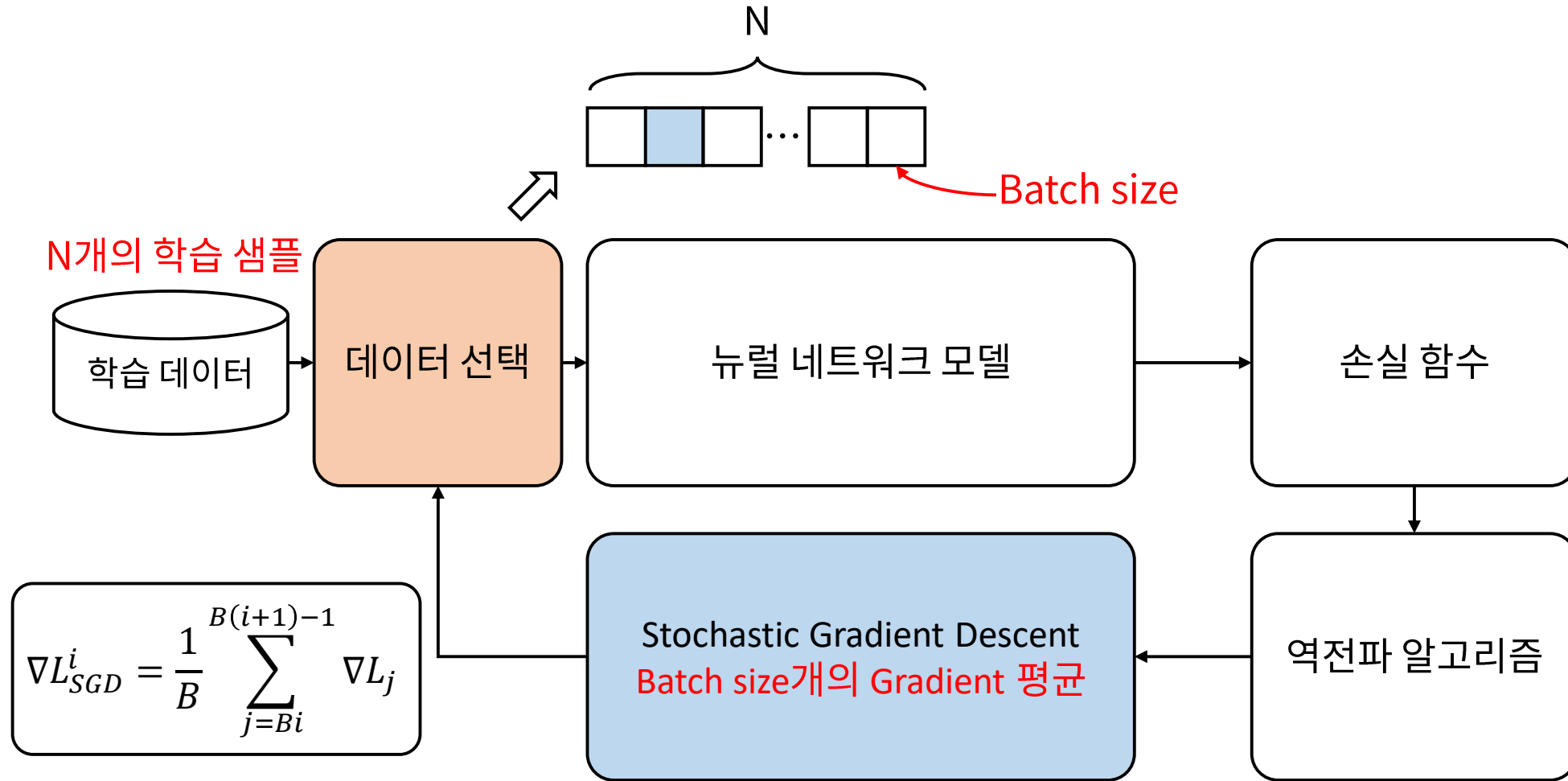
뉴럴 네트워크의 학습 과정

I Vanilla Gradient Descent



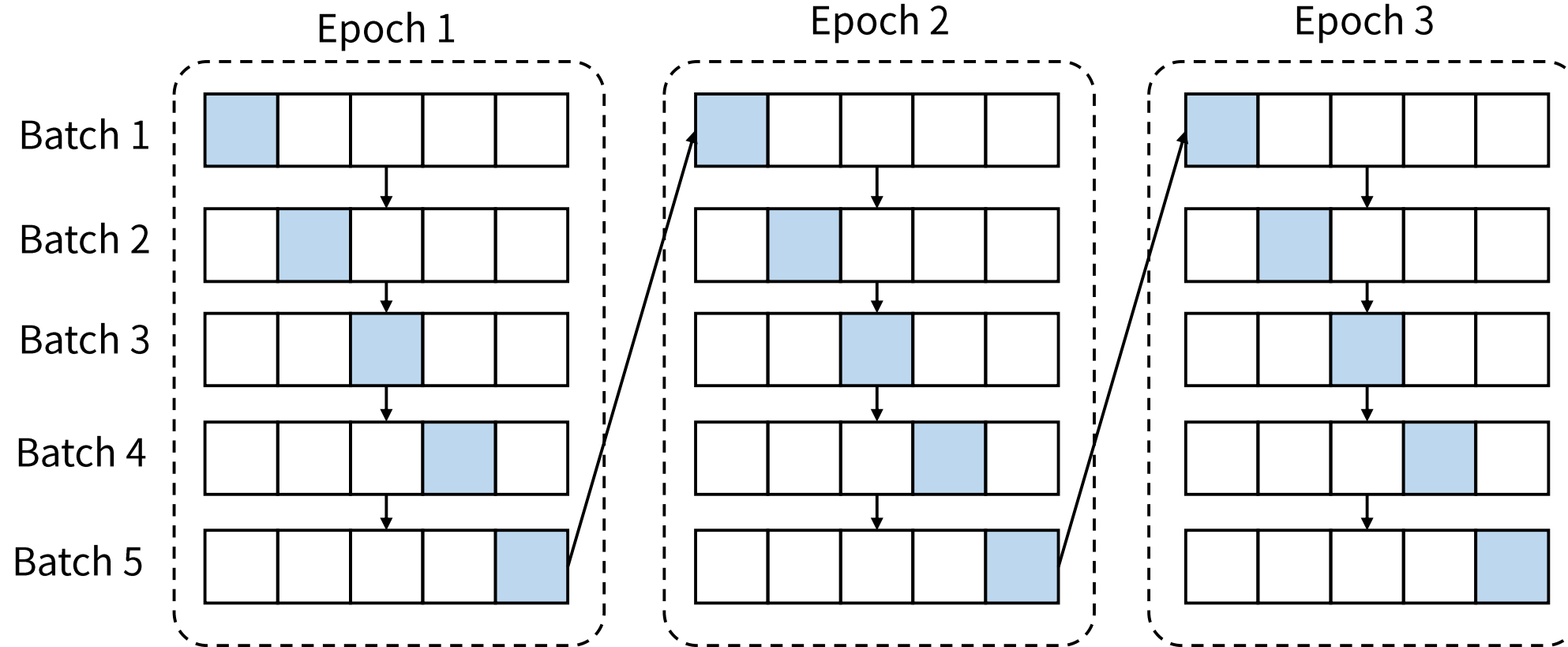
Vanilla Gradient Descent (GD)의 경우, Gradient를 한번 업데이트 하기 위해 모든 학습 데이터를 사용한다.

I Stochastic Gradient Descent



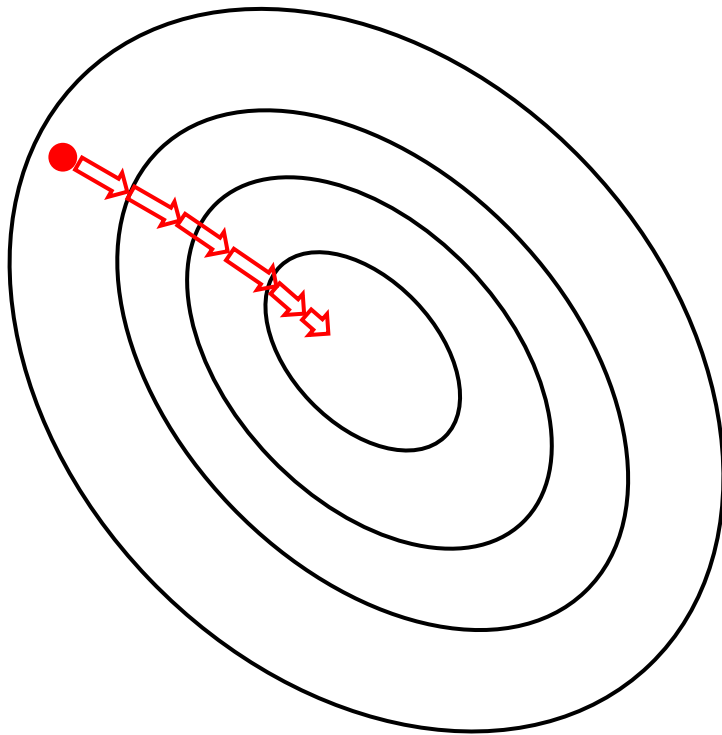
Stochastic GD (SGD)의 경우, Gradient를 한번 업데이트 하기 위해 일부의 데이터만을 사용한다.

I Mini-Batch 학습법

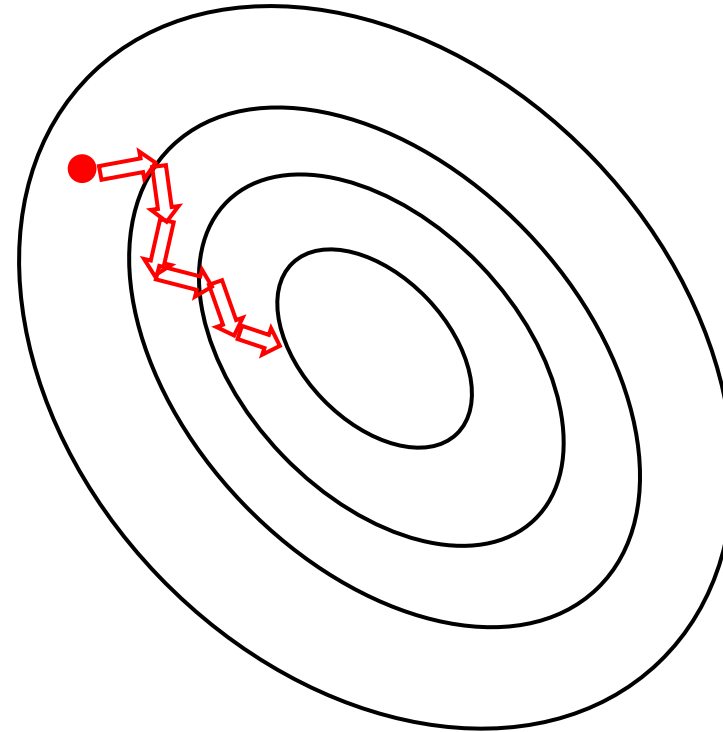


학습 데이터 전체를 한번 학습하는 것을 Epoch, 한 번 Gradient를 구하는 단위를 Batch라고 한다.

I GD vs. SGD



Vanilla Gradient Descent



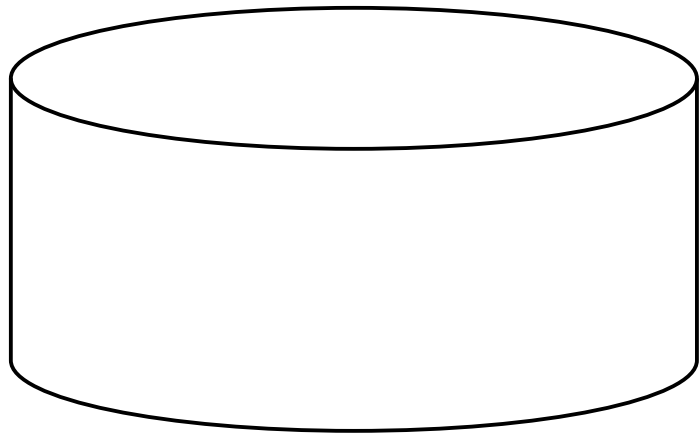
Stochastic Gradient Descent

확률적 기울기(Stochastic Gradient)보다 실제 기울기가 더 좋은 결과가 나타난다.

그럼에도 SGD를 사용하는 이유는?

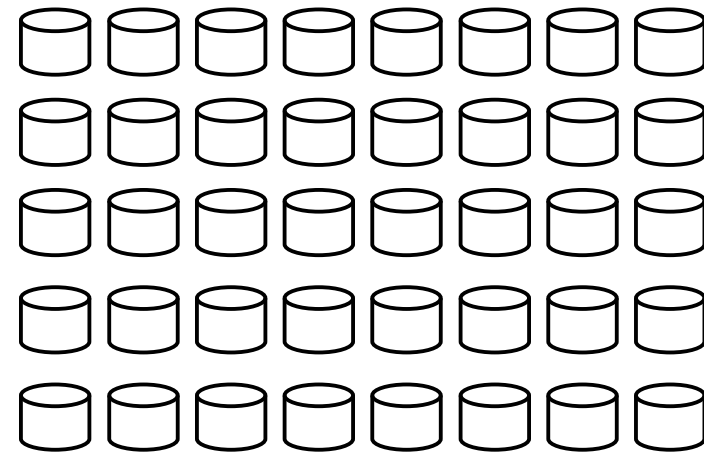
I GD vs. SGD

Training Data > 10k



>>

Batch Size : 8, 16, 32, 64, 128 ...



Gradient를 한 번 Update하는 데 걸리는 시간이 수백, 수천 배 이상 차이가 난다!
SGD에서 Gradient의 질이 떨어지는 부분은 심화 최적화 알고리즘으로 커버 가능.