

Chapter 05. 합성곱 신경망(CNN) 작동 원리

STEP2. 배치 정규화

일반 경사 하강법 Vanilla Gradient Descent

N개의 학습 샘플



뉴럴 네트워크 모델

손실 함수

$$\nabla L_{GD} = \frac{1}{N} \sum_{i=0}^{N-1} \nabla L_i$$

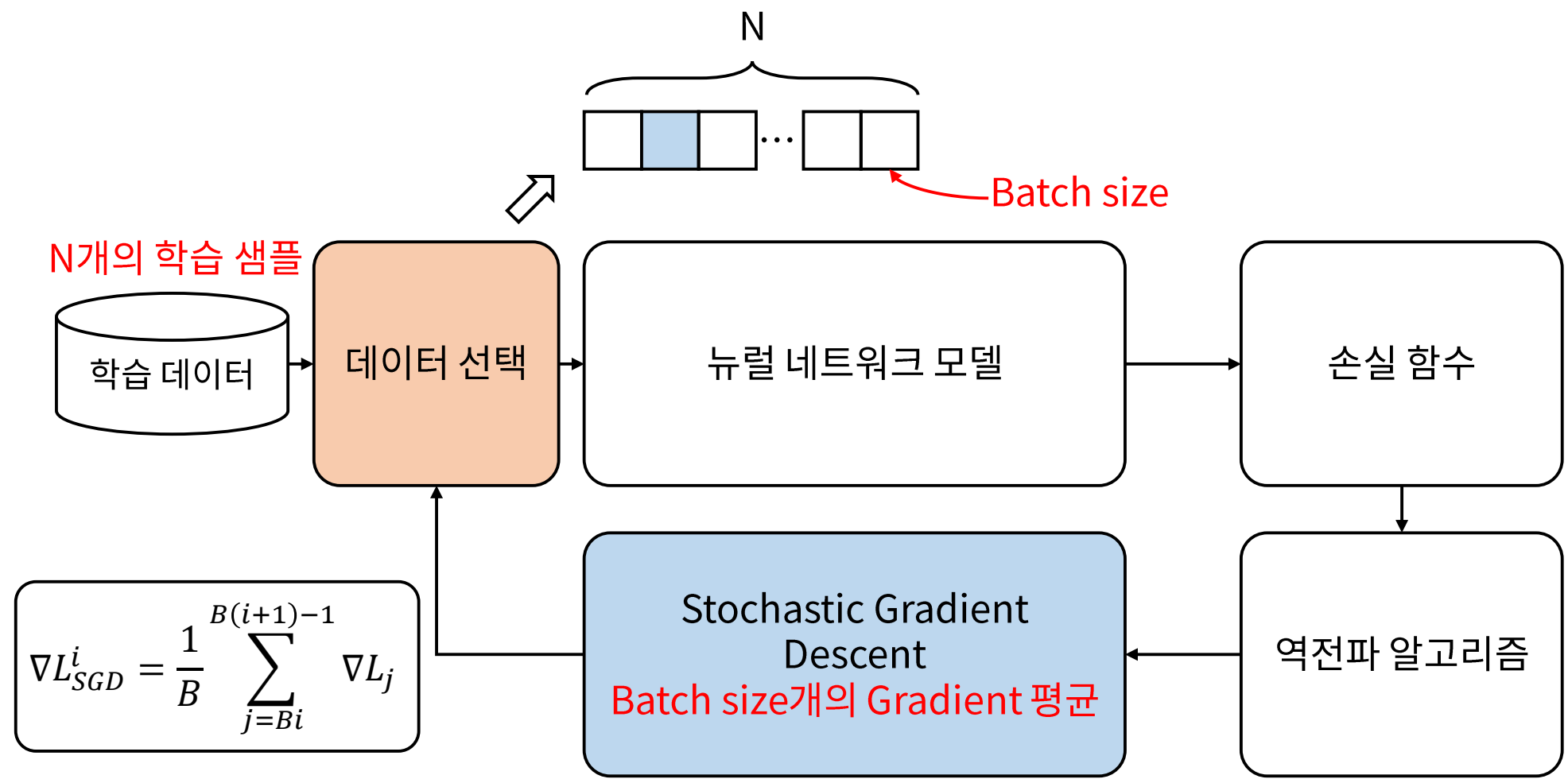
Gradient Descent

N개 Gradient의 평균

역전파 알고리즘

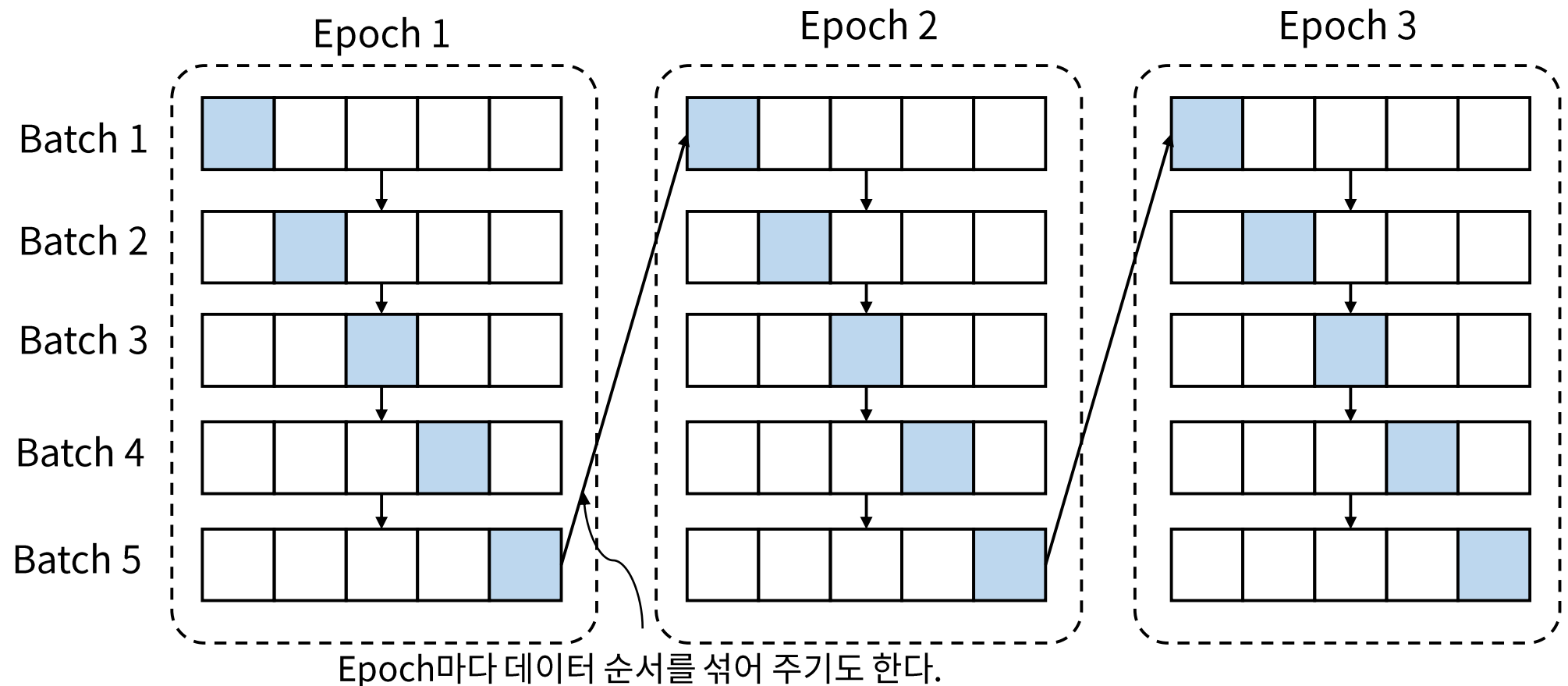
일반 경사 하강법의 경우, Gradient를 한번 업데이트 하기 위해 모든 학습 데이터를 사용한다.

확률적 경사 하강법 Stochastic Gradient Descent



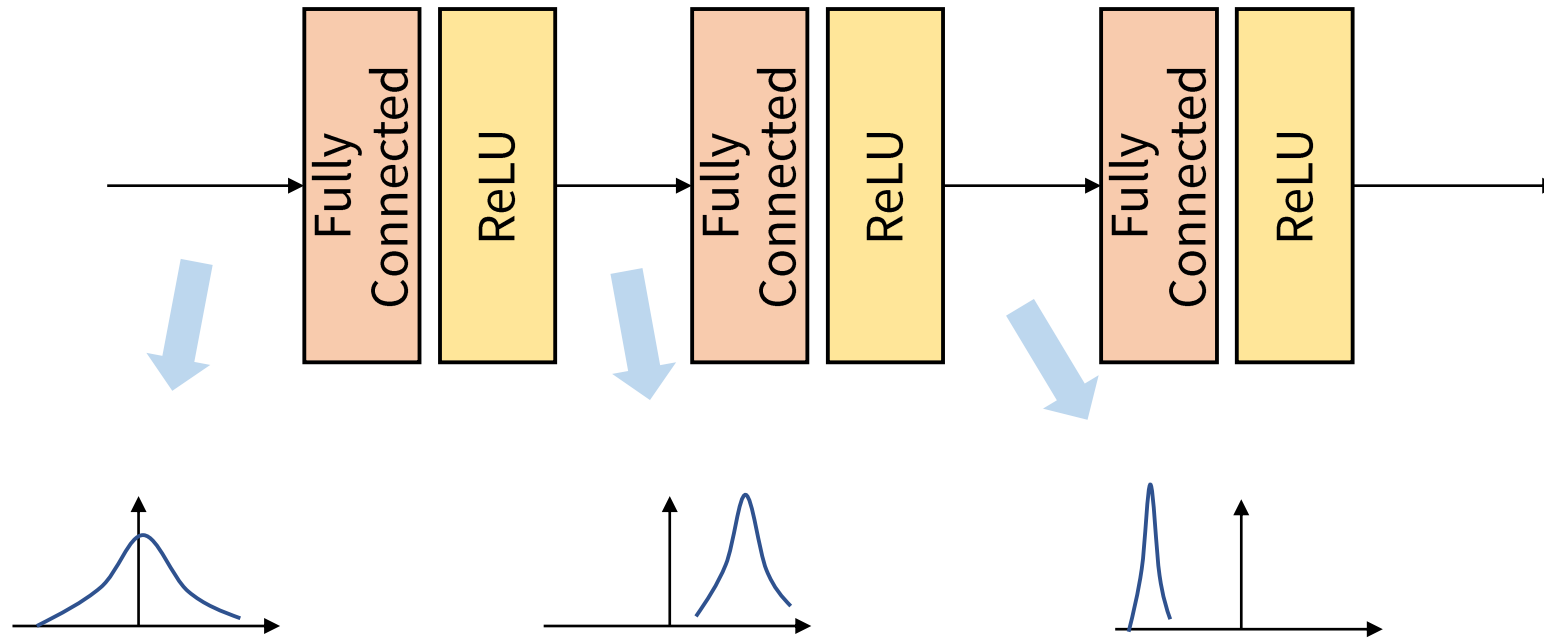
Stochastic GD (SGD)의 경우, Gradient를 한번 업데이트 하기 위해 일부의 데이터만을 사용한다.

미니 배치 학습법



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

Internal Covariate Shift



학습 과정에서 계층별로 입력의 데이터 분포가 달라지는 현상을 **Internal Covariate Shift**라고 한다.