

Part. 09

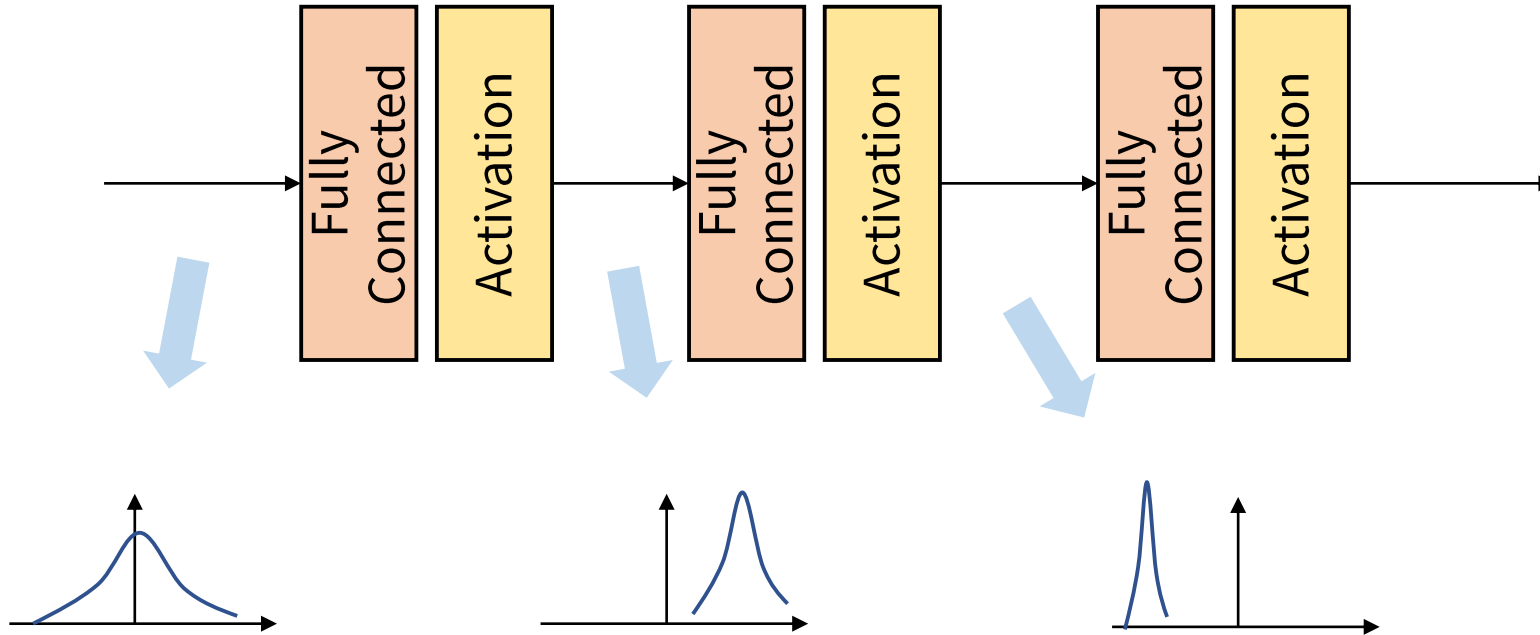
Modern Neural Networks

# | Batch Normalization

FASTCAMPUS  
ONLINE

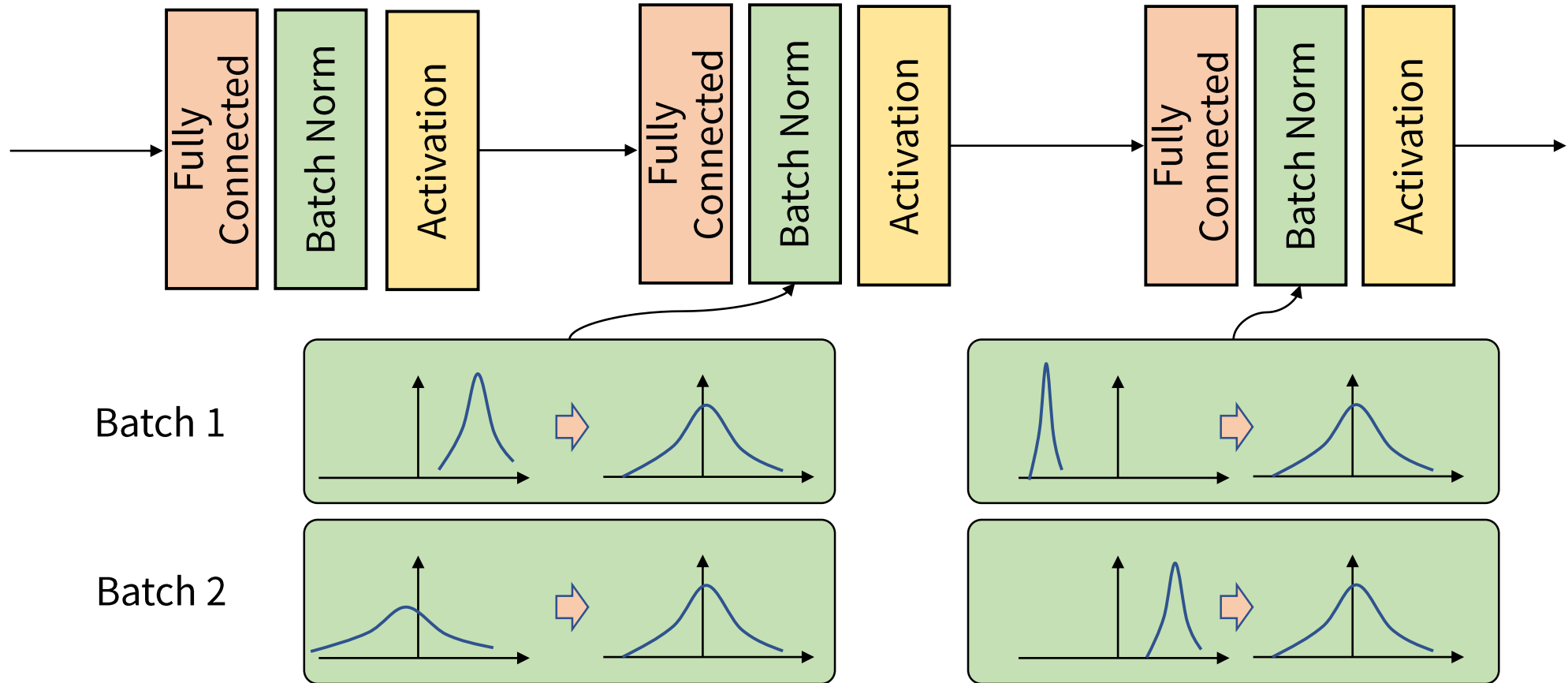
강사. 신제용

# I Internal Covariate Shift



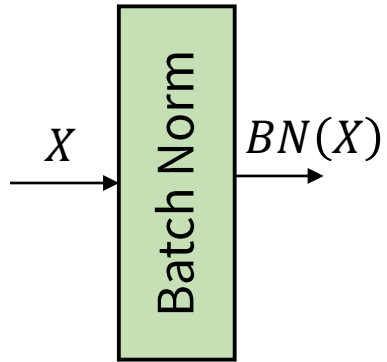
학습 과정에서 계층별로 입력의 데이터 분포가 달라지는 현상을 **Internal Covariate Shift**라고 한다.

# I 배치 정규화 (Batch Normalization)



학습 과정에서 각 배치별로 평균과 분산을 이용해 정규화하는 계층을 배치 정규화 계층이라 한다.

## I 배치 정규화의 학습 단계



배치별로 계산

$$\begin{cases} \mu_{\text{batch}} = \frac{1}{B} \sum_i x_i \\ \sigma_{\text{batch}}^2 = \frac{1}{B} \sum_i (x_i - \mu_{\text{batch}})^2 \end{cases}$$

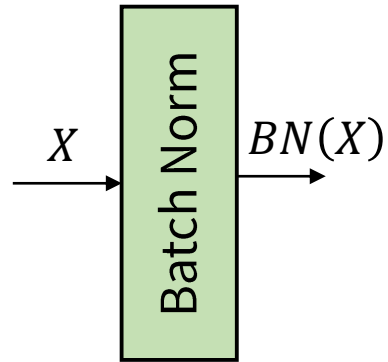
$$BN(X) = \gamma \left( \frac{X - \mu_{\text{batch}}}{\sigma_{\text{batch}}} \right) + \beta$$

정규화

추가 스케일링, 편향  
역전파 알고리즘으로 학습

- 정규화로 인해, 모든 계층의 Feature가 동일한 Scale이 되어 학습률 결정에 유리하다.
- 추가적인 Scale, Bias를 학습하여 Activation에 적합한 분포로 변환할 수 있다.

## I 배치 정규화의 추론 단계



$$BN(X) = \underbrace{\gamma}_{\text{역전파로 학습된 값}} \left( \underbrace{\frac{X - \mu_{BN}}{\sigma_{BN}}}_{\text{정규화}} \right) + \underbrace{\beta}_{\text{역전파로 학습된 값}}$$

$$\begin{cases} \mu_{BN} = \frac{1}{N} \sum_i \mu_{\text{batch}}^i \\ \sigma_{BN} = \frac{1}{N} \sum_i \sigma_{\text{batch}}^i \end{cases}$$

학습 과정에서 이동 평균을 계산

- 추론 과정에서는 평균과 분산을 이동 평균(또는 지수 평균)하여 고정
- 추론 단계에서는 정규화와 추가 Scale, Bias를 결합하여 단일 곱, 더하기 연산으로 줄일 수 있음

## I 합성곱 계층의 배치 정규화

$$X \in \mathbb{R}^{B \times N}$$

$$BN(\mathbf{X}) = \gamma \left( \frac{\mathbf{X} - \boldsymbol{\mu}_{\text{batch}}}{\boldsymbol{\sigma}_{\text{batch}}} \right) + \boldsymbol{\beta}$$

$$\begin{cases} \boldsymbol{\mu}_{\text{batch}} = \frac{1}{B} \sum_i \mathbf{x}_i \\ \boldsymbol{\sigma}_{\text{batch}}^2 = \frac{1}{B} \sum_i (\mathbf{x}_i - \boldsymbol{\mu}_{\text{batch}})^2 \end{cases}$$

전결합 계층 - 각 뉴런 별로 정규화

$$\mathbf{X} \in \mathbb{R}^{B \times H \times W \times C}$$

$$BN(\mathbf{X}) = \gamma \left( \frac{\mathbf{X} - \boldsymbol{\mu}_{\text{batch}}}{\boldsymbol{\sigma}_{\text{batch}}} \right) + \boldsymbol{\beta}$$

$$\begin{cases} \boldsymbol{\mu}_{\text{batch}} = \frac{1}{BHW} \sum_i \sum_j \sum_k \mathbf{x}_{i,j,k} \\ \boldsymbol{\sigma}_{\text{batch}}^2 = \frac{1}{BHW} \sum_i \sum_j \sum_k (\mathbf{x}_{i,j,k} - \boldsymbol{\mu}_{\text{batch}})^2 \end{cases}$$

합성곱 계층 - 각 채널 별로 정규화

합성곱 계층에 적용할 경우, 채널 별로 정규화한다. 즉, 배치, 높이, 너비에 대해 평균과 분산을 계산한다.