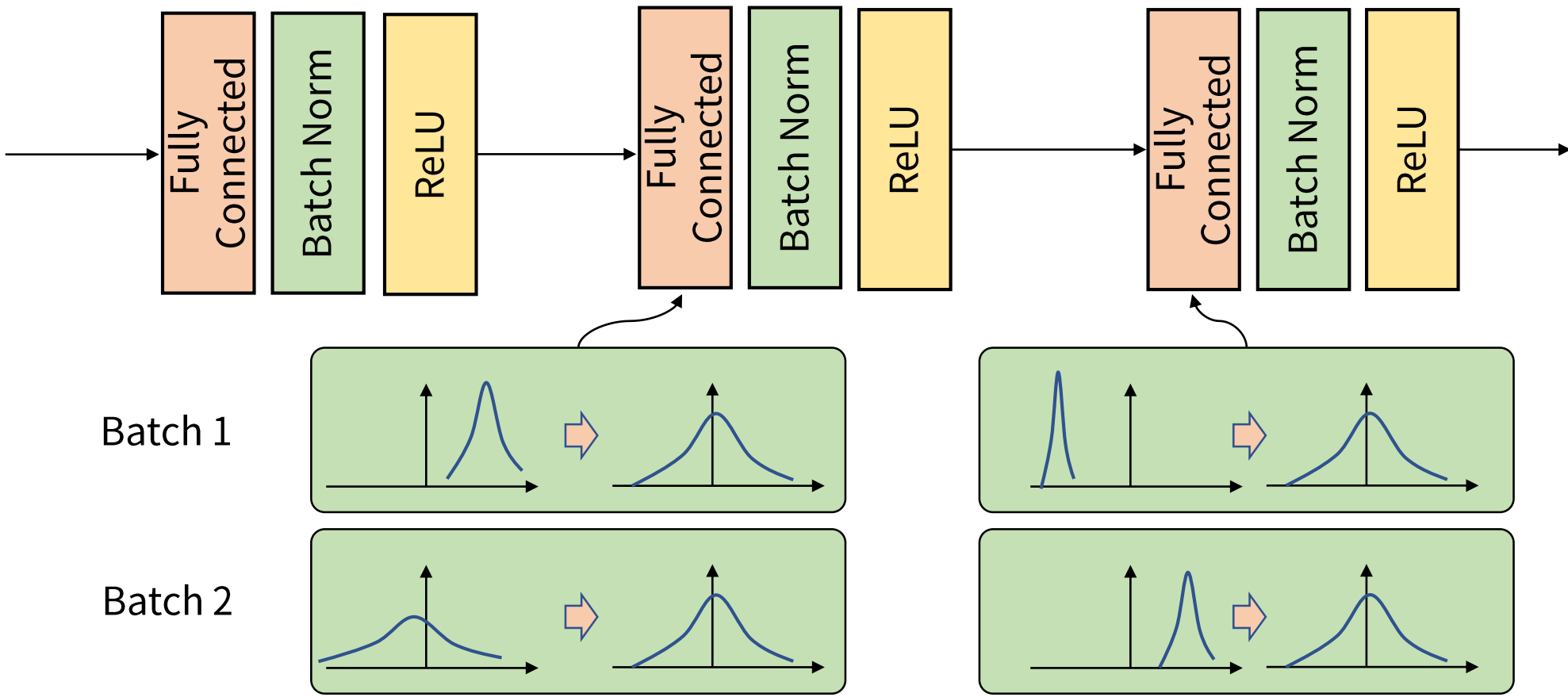


Chapter 08. 효과적이면서도 쉽게 쓸 수 있는 기법들

STEP2.

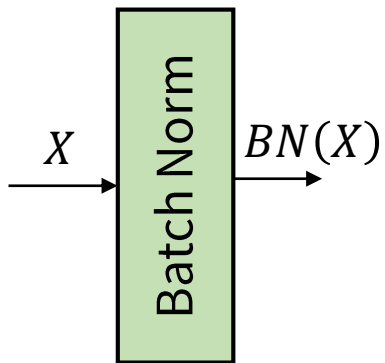
배치 정규화와 변형 기법들

배치 정규화 Batch Normalization



여러 번 언급했던 Batch Normalization을 조금 더 자세히 알아보자.

학습 단계 Training Phase



배치별로 계산

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

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

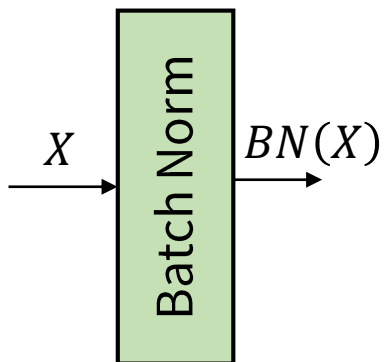
정규화

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.

추론 단계 Inference Phase



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

정규화

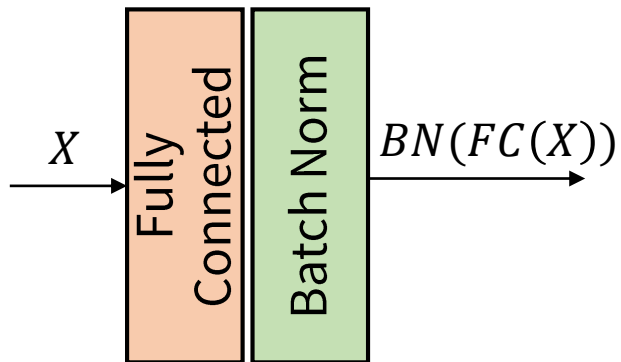
역전파로 학습된 값

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

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

전결합 계층과 결합



$$FC(X) = WX + b$$

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

$$BN(FC(X)) = \gamma \left(\frac{WX + b - \mu_{BN}}{\sigma_{BN}} \right) + \beta$$

$$= \frac{\gamma}{\sigma_{BN}} WX + \frac{b - \mu_{BN} + \sigma_{BN} \beta}{\sigma_{BN}}$$

역할이 겹치므로 b 를 사용하지 않는다.

- 전결합 계층의 편향이 BatchNorm의 편향과 역할이 겹치므로, 전결합 계층의 편향을 제거한다.

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.