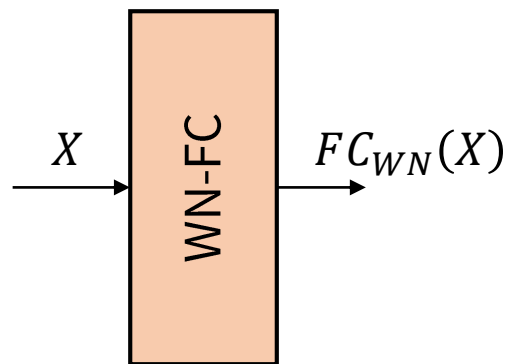


배치 정규화의 한계

$$\left\{ \begin{array}{l} \mu_{\text{batch}} = \frac{1}{B} \sum_i x_i \\ \sigma_{\text{batch}}^2 = \frac{1}{B} \sum_i (x_i - \mu_{\text{batch}})^2 \end{array} \right.$$

- Mini-batch에 의해 크게 영향을 받는다.
 - 배치의 크기가 너무 작으면 잘 동작하지 않는다.
 - 메모리의 한계로 인해 RNN이나 크기가 큰 CNN에 적용하기 어렵다
 - 배치의 크기가 너무 커도 잘 동작하지 않는다.
 - 병렬화 연산 효율이 떨어진다.

가중치 정규화 Weight Normalization



$$FC(x)_i = w_i^T x + b_i$$

$$FC_{WN}(x)_i = w_i^T x + b_i$$

$$= \frac{g_i}{\|v_i\|} v_i^T x + b_i$$

Re-parametrization

$$w = \frac{g}{\|v\|} v$$

- FC 계층의 w 는 ‘방향’과 ‘크기’를 같이 학습하지만, 이를 분리하여 g 와 v 로 나누어 학습
- 학습 시 자유도가 개선되어 최적화가 더 쉽게 이루어짐
- 학습 시 CNN에서 배치 정규화 대비 연산량이 매우 감소 (Feature에 적용 vs. Weight에 적용)

계층 정규화 Layer Normalization

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

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

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

배치 정규화



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

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

$$\begin{cases} \mu_{\text{layer}} = \frac{1}{N} \sum_j \mathbf{x}_{:,j}^T \\ \sigma_{\text{layer}}^2 = \frac{1}{N} \sum_j (\mathbf{x}_{:,j}^T - \mu_{\text{layer}})^2 \end{cases}$$

계층 정규화

- 배치 정규화와 달리, 각 샘플 내에서 **계층에 대해 정규화 수행** (모든 데이터가 서로 independent하다는 가정)
- 배치의 크기에 영향을 받지 않으며, **RNN에서도 잘 동작함**