Deep Learning Basic

Jaewon Kim, Dankook Univ.

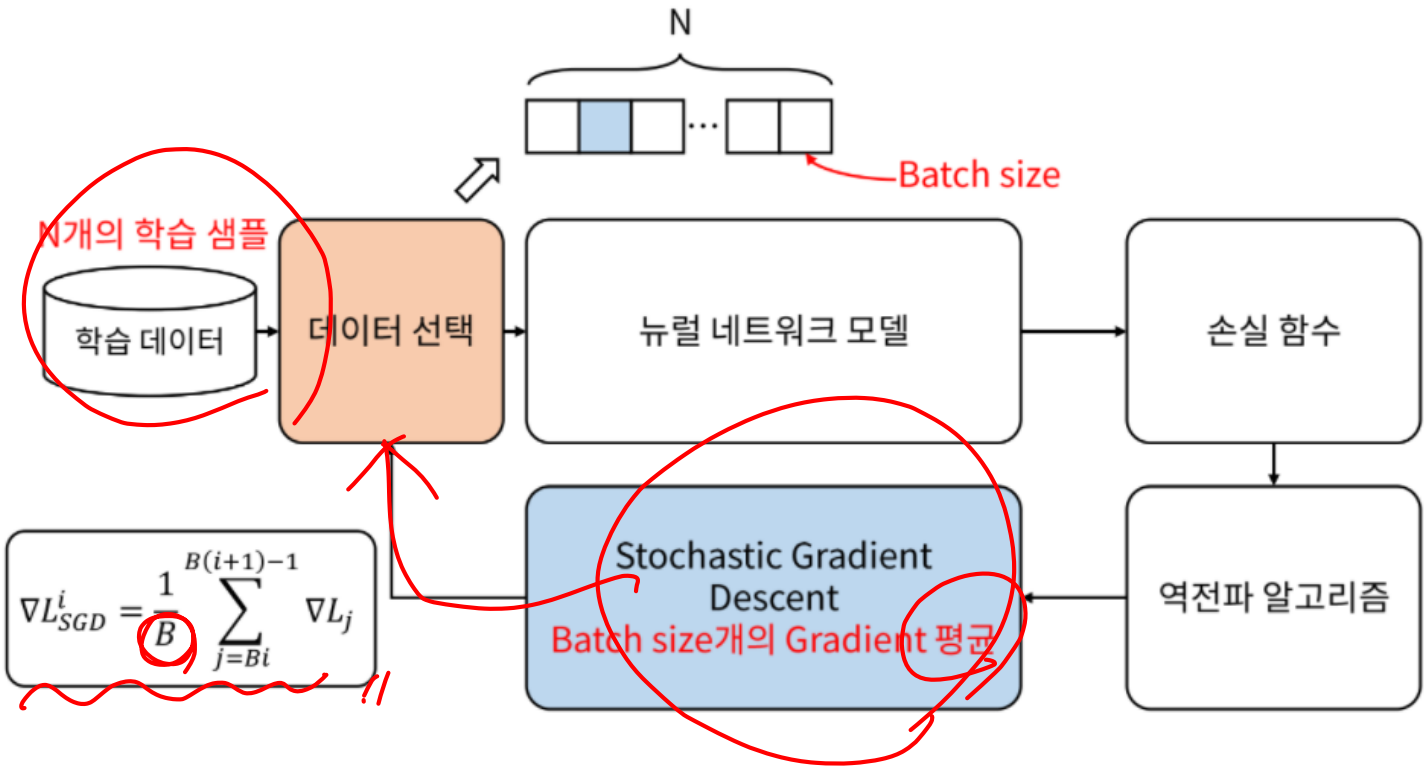
Chapter 4-1

Contents

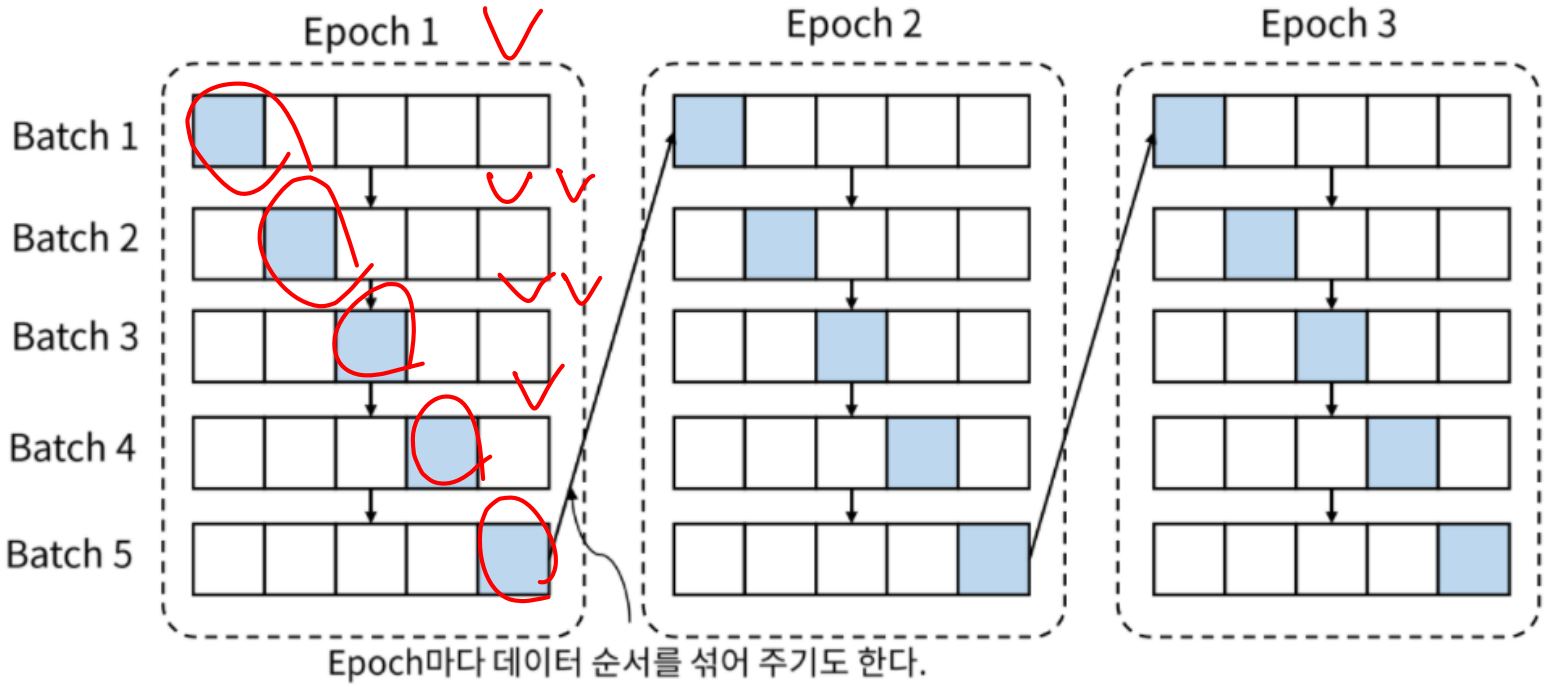
Part 1. Batch Normalization

- Stochastic Gradient Decent (Batch)
- Internal Covarient Shift
- Batch Normalization

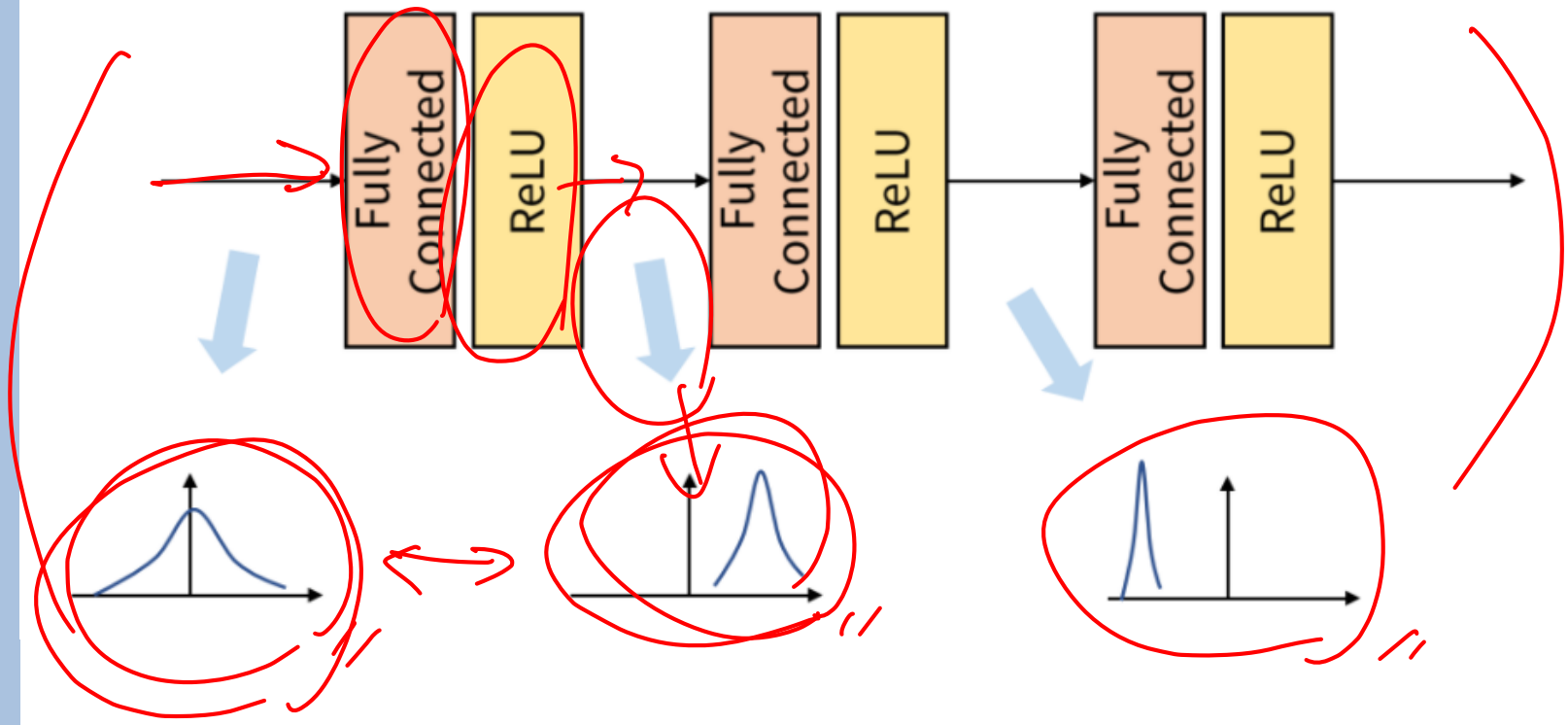
Stochastic Gradient Decent (Batch)



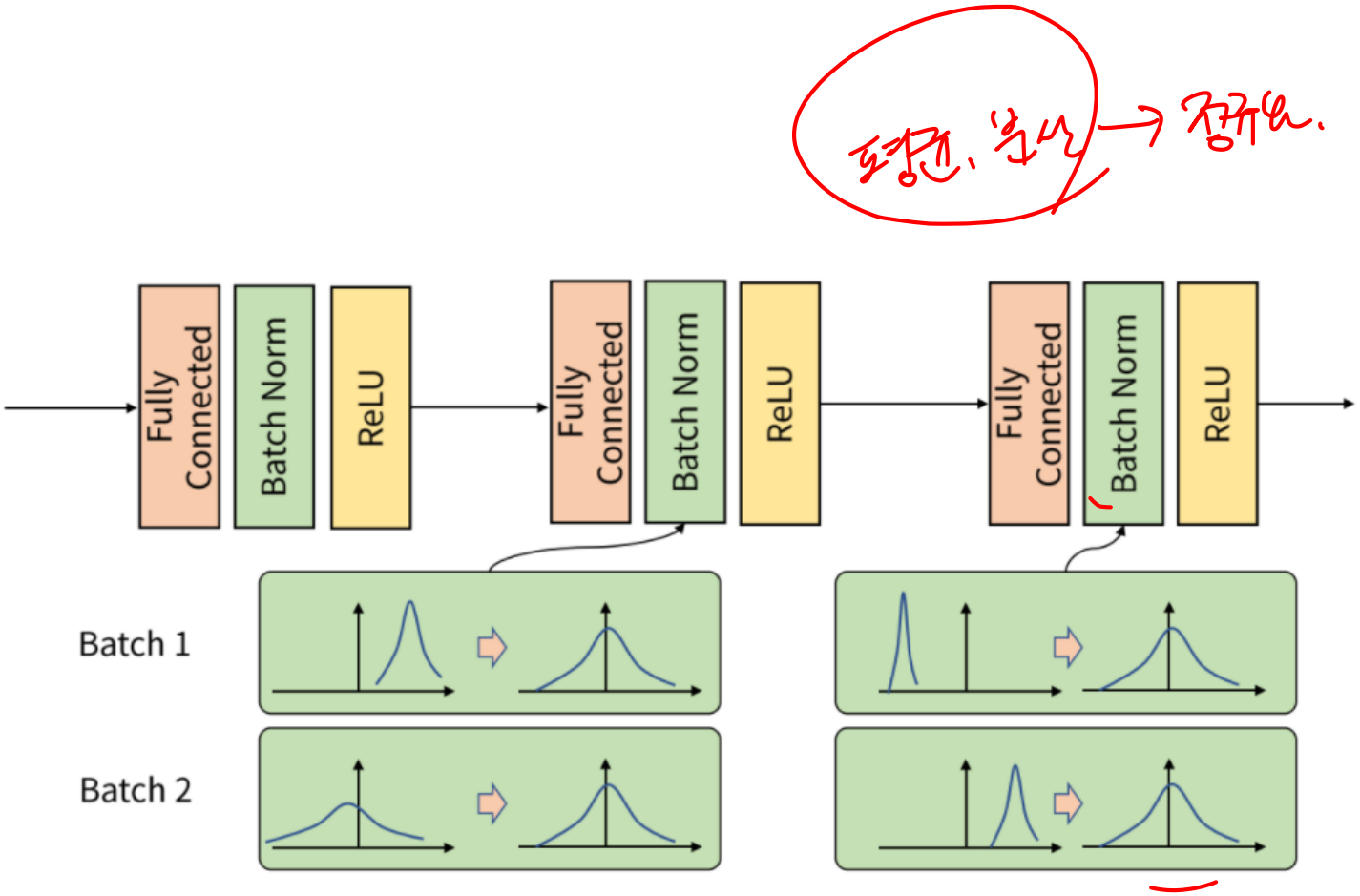
Stochastic Gradient Decent (Batch)



Internal Covarient Shift



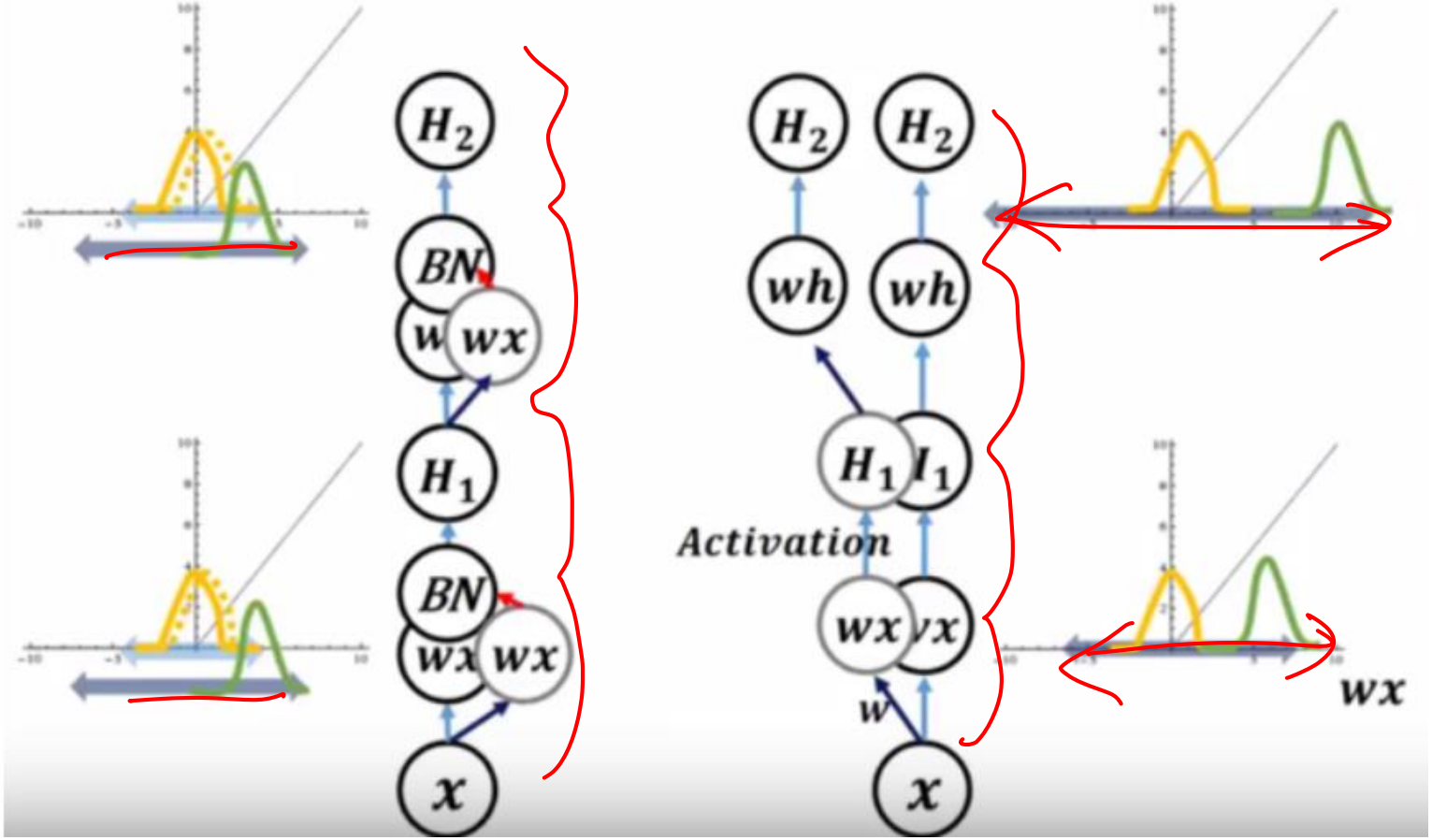
Batch Normalization



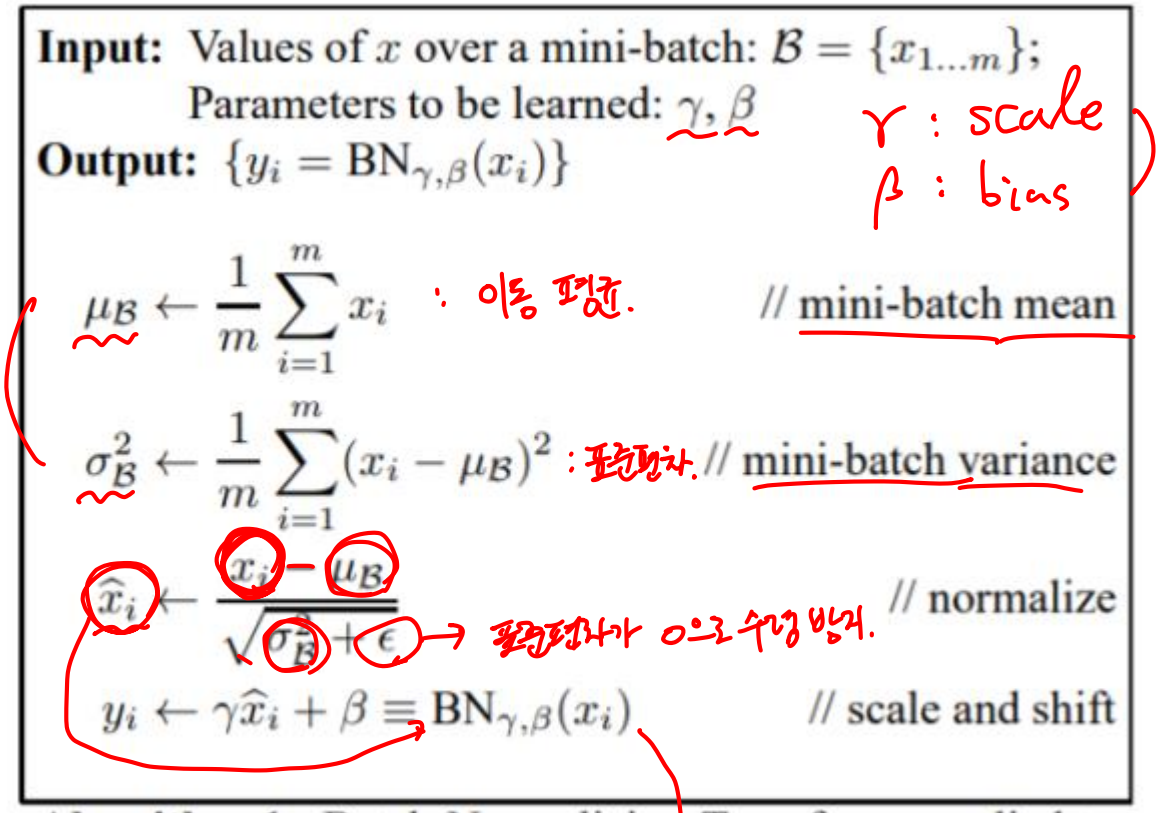
Batch Normalization

x축 : wx.

y축 : activation.



Batch Normalization



Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

Algorithm 1: Batch Normalizing Transform, applied to activation x over a mini-batch.

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

Batch Normalization

Effect

- Training을 빠르게 해준다. Time↓↓
- Learning Rate이 작아야 하는 부담 감소↓↓
- 모델의 Regularization → Overfitting 부담감소↓↓
- Weight Initialization의 부담 감소↓↓

Caution

- Batch size가 너무 크거나 작으면 X (보통 128,256,512 ...) 1024

$$\begin{cases} \mu_{\text{batch}} = \frac{1}{B}\sum_{i} x_i & \text{평균} \\ \sigma^2_{\text{batch}} = \frac{1}{B}\sum_{i}(x_i - \mu_{\text{batch}})^2 & \text{표준편차.} \end{cases}$$

Thank you..!!!

본 강의 자료는 JINSOL KIM님 자료를 참고 하였습니다. 감사합니다.
Reference : https://gaussian37.github.io/dl-concept-batchnorm/