Batch_Normalization

Batch
Normalization

Batch_Normalization

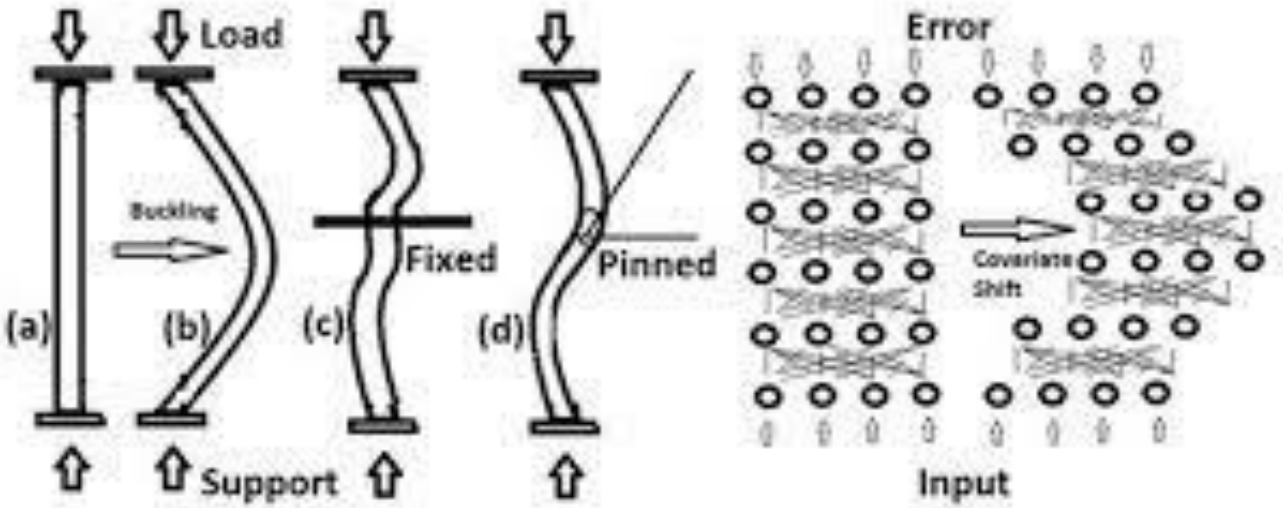
Batch_norm

Why use batch_norm?

Internal covariate shift

Batch_norm

Internal covariate shift

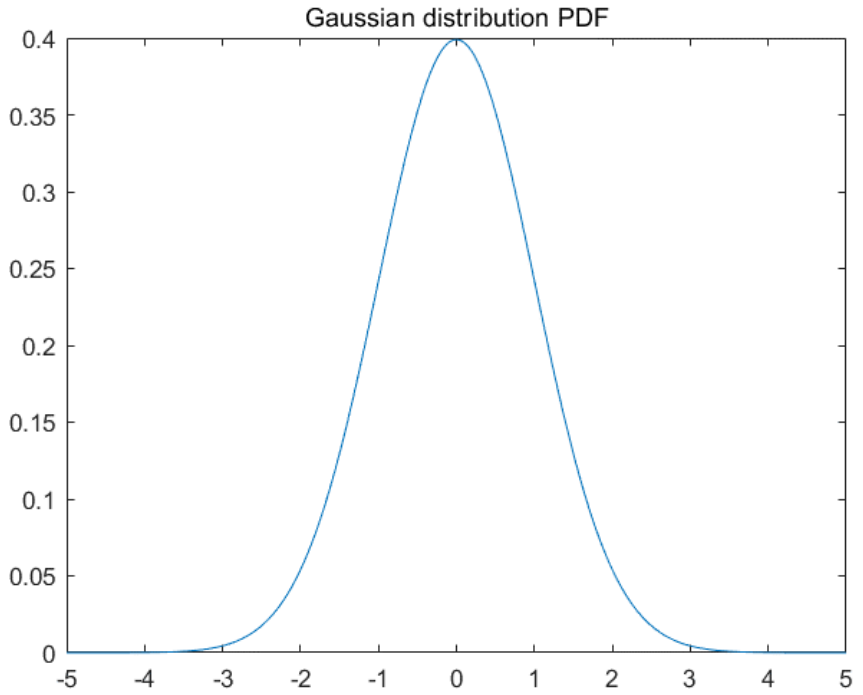


Solve Gradient vanishing

출처 :
http://blog.naver.com/PostView.nhn?blogId=laonple&logNo=220808903260

Batch_norm

All Samples follow a Gaussian distribution



Mean = 0 , Covariance =1

Batch_norm

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

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

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

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

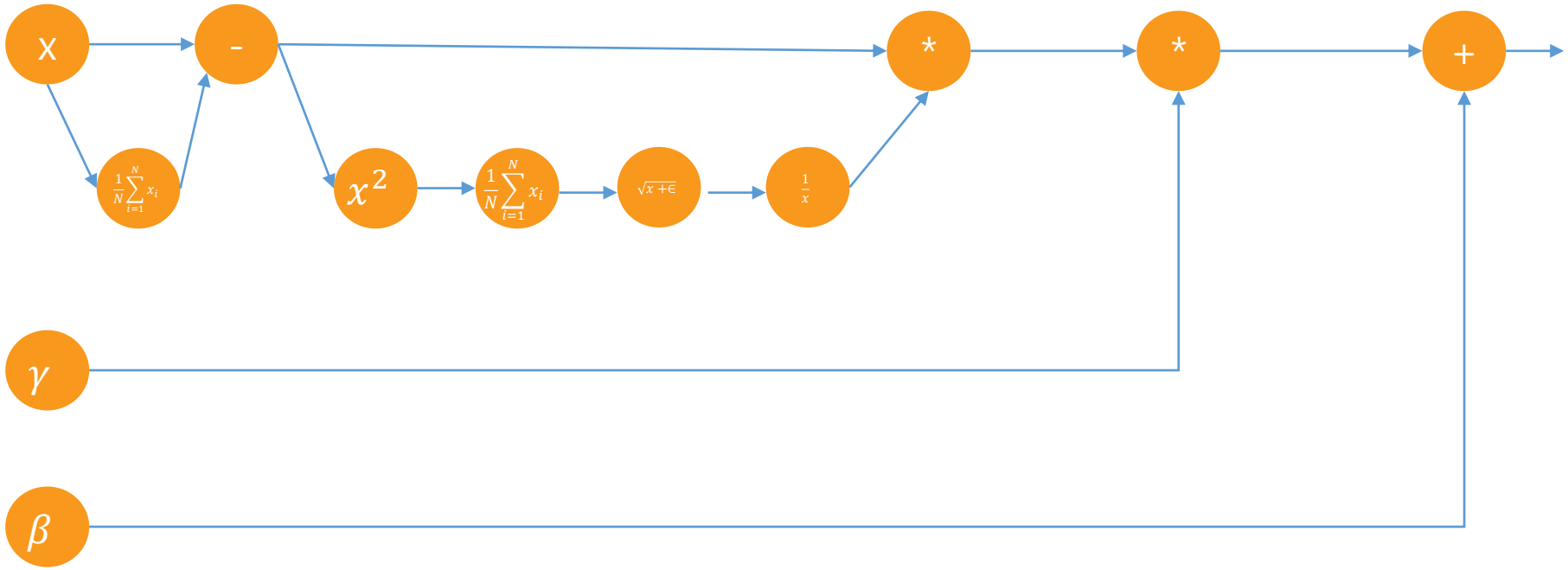
$$y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift

Batch_norm

How to opimize gamma, beta

Chain Rule



참고 : https://kratzert.github.io/2016/02/12/understanding-the-gradient-flow-through-the-batch-normalization-layer.html

Batch_norm

$$\frac{\partial l}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i} * x_i$$

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i}$$

Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift

Batch_norm

Sample Batch

$$E[x] = E_B[\mu_B]$$

$$Var[x] = E_B[\sigma_B^2] * \frac{m}{m-1}$$

$$y = \frac{\gamma}{\sqrt{Var[x] + \epsilon}} x + (\beta - \frac{\gamma E[x]}{\sqrt{Var[x] + \epsilon}})$$

Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift

Batch_norm

Advantages

1. Gradient Vanishing X(internal covariate shift)

2. Learning Rate Upscale

3. You don`t have to use Dropout

Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift

Batch_norm

Discussion

Use ReLU, Batch Norm

Batch_norm

Discussion

감사합니다

THANK YOU