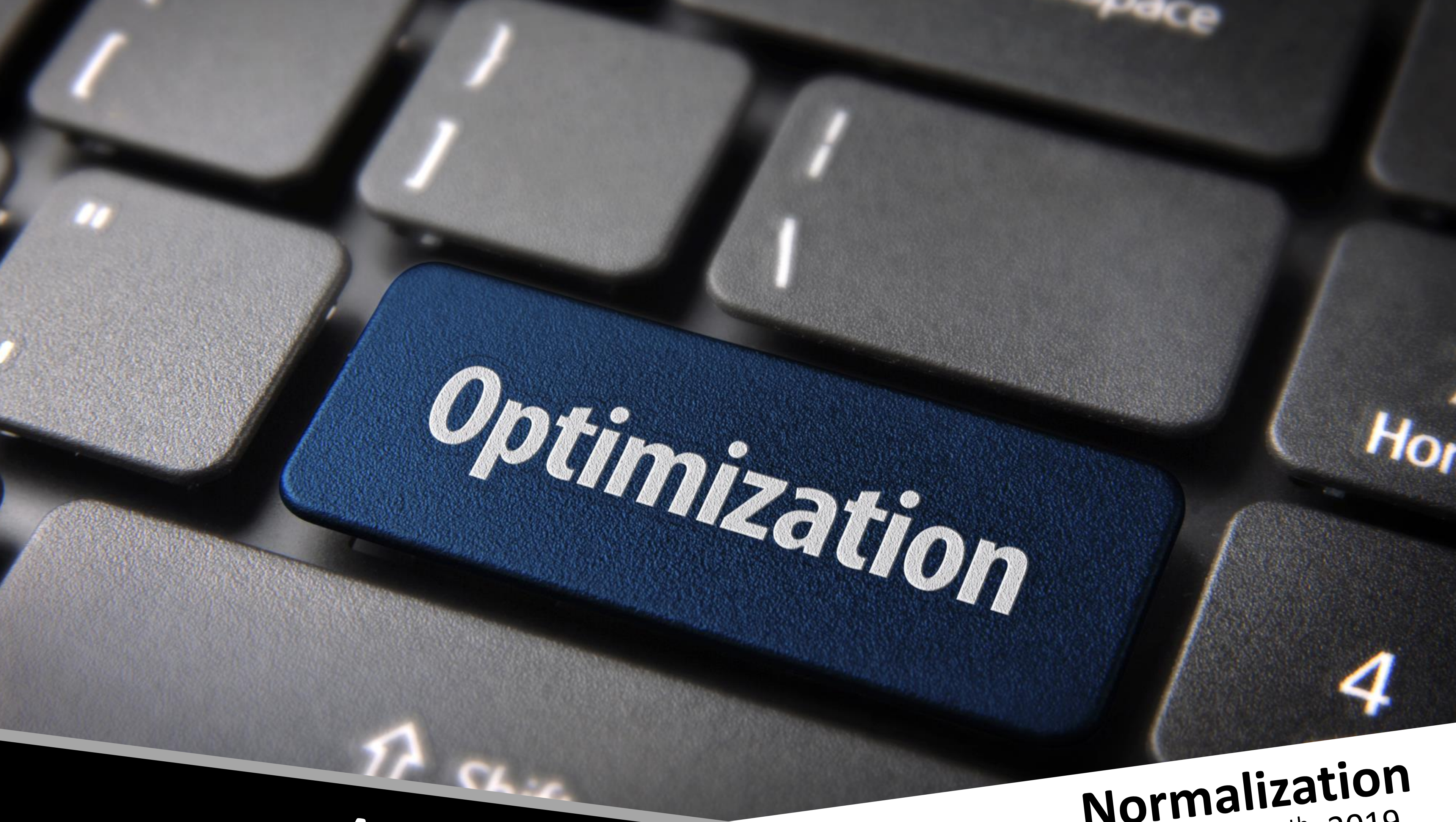


Normalization

Mar 19th, 2019

Applied Deep Learning

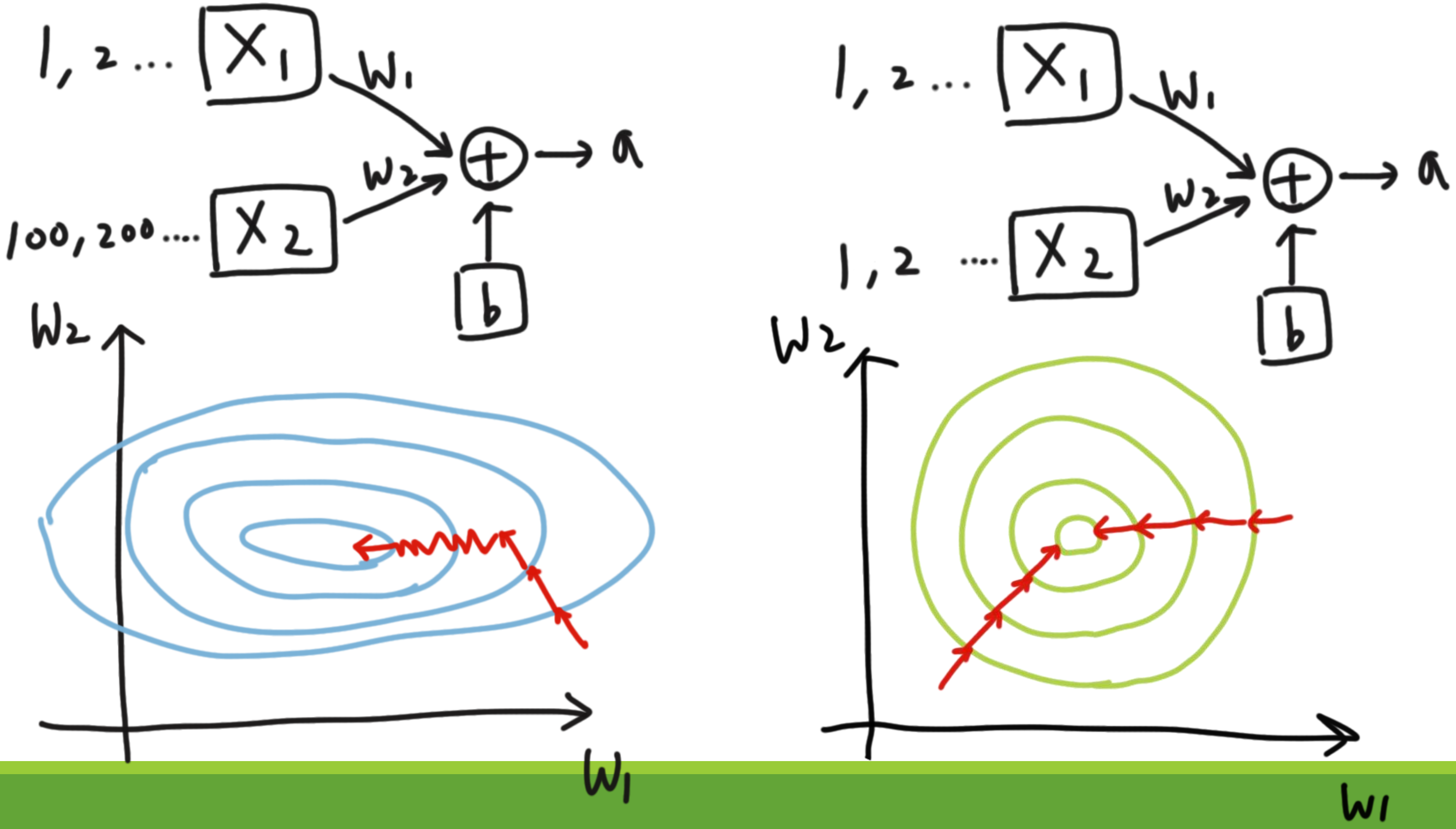
SHANG-YU SU

HTTP://ADL.MIULAB.TW

國立臺灣大學
National Taiwan University

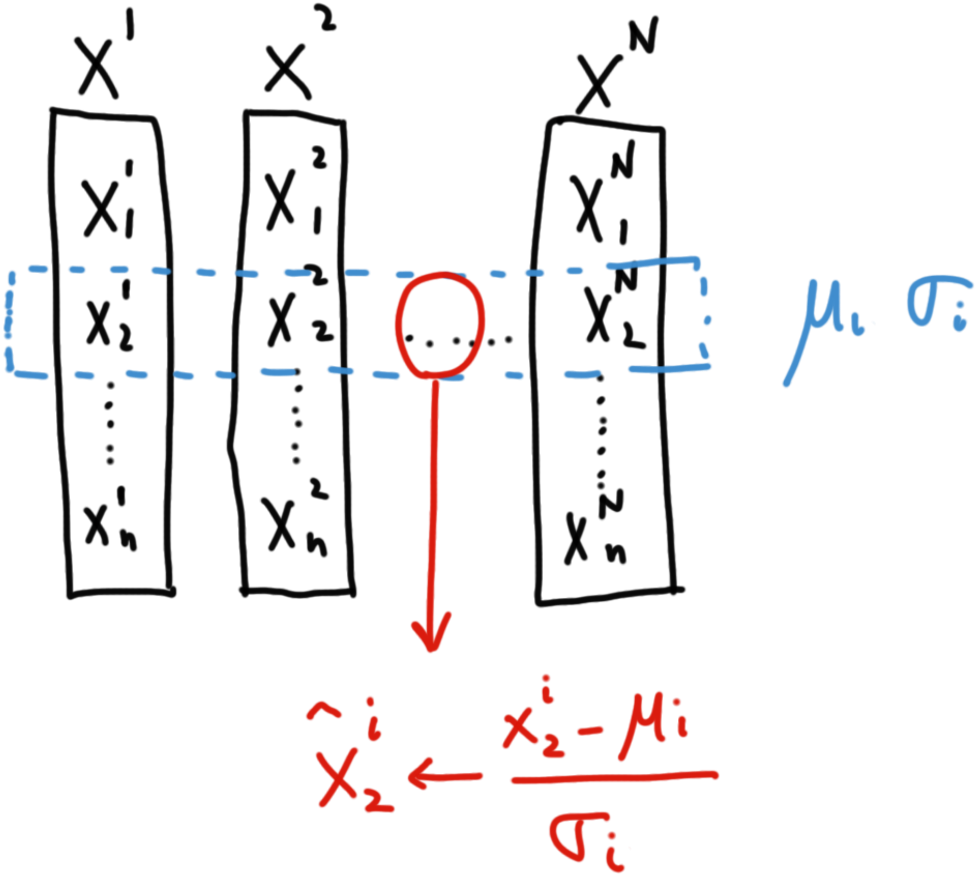
Feature Scaling

- Idea: make sure features are on the same scale



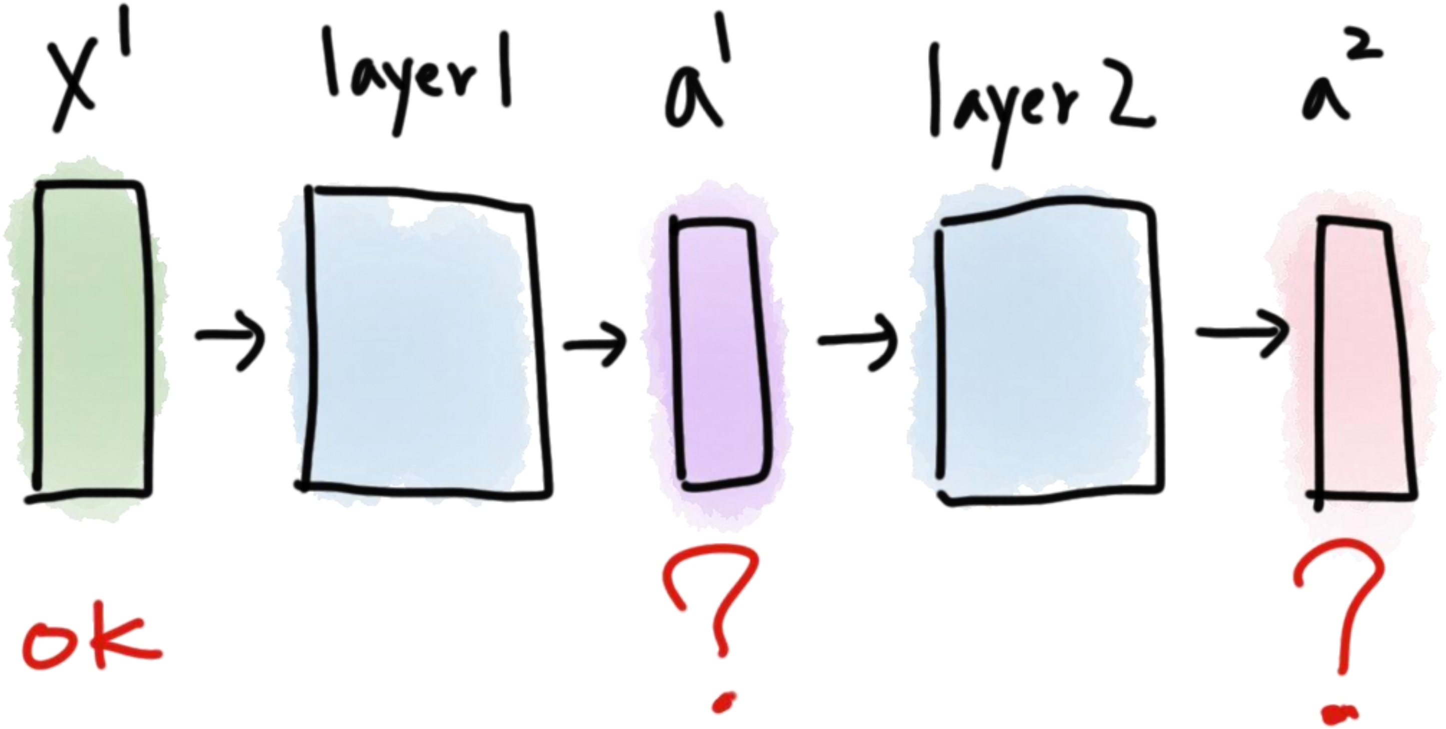
Feature Scaling

- for each dimension, compute mean and standard deviation
- the means of normalized feature vectors are all 0 and the variances are all 1

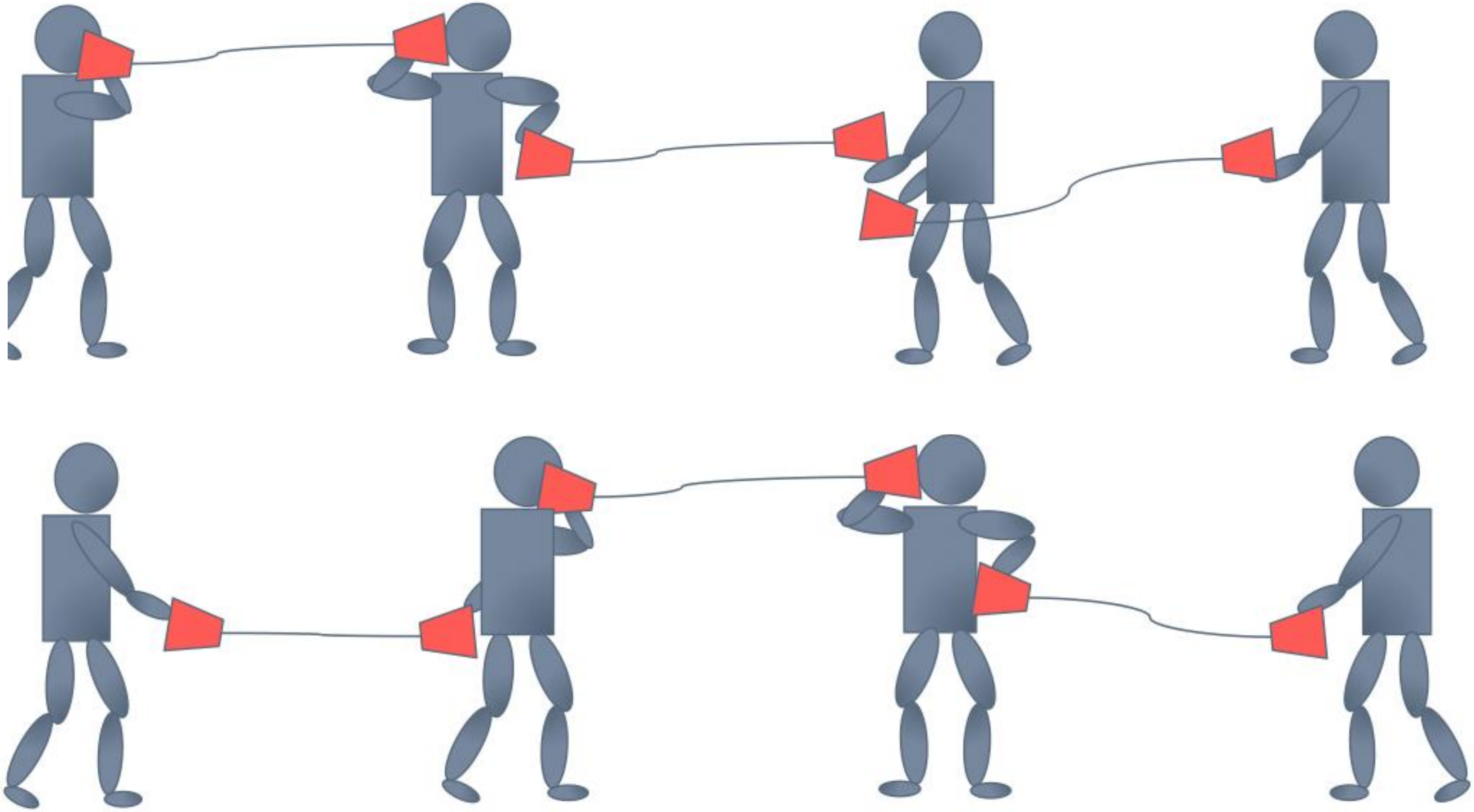


Hidden States

- statistics of hidden states keep changing during training

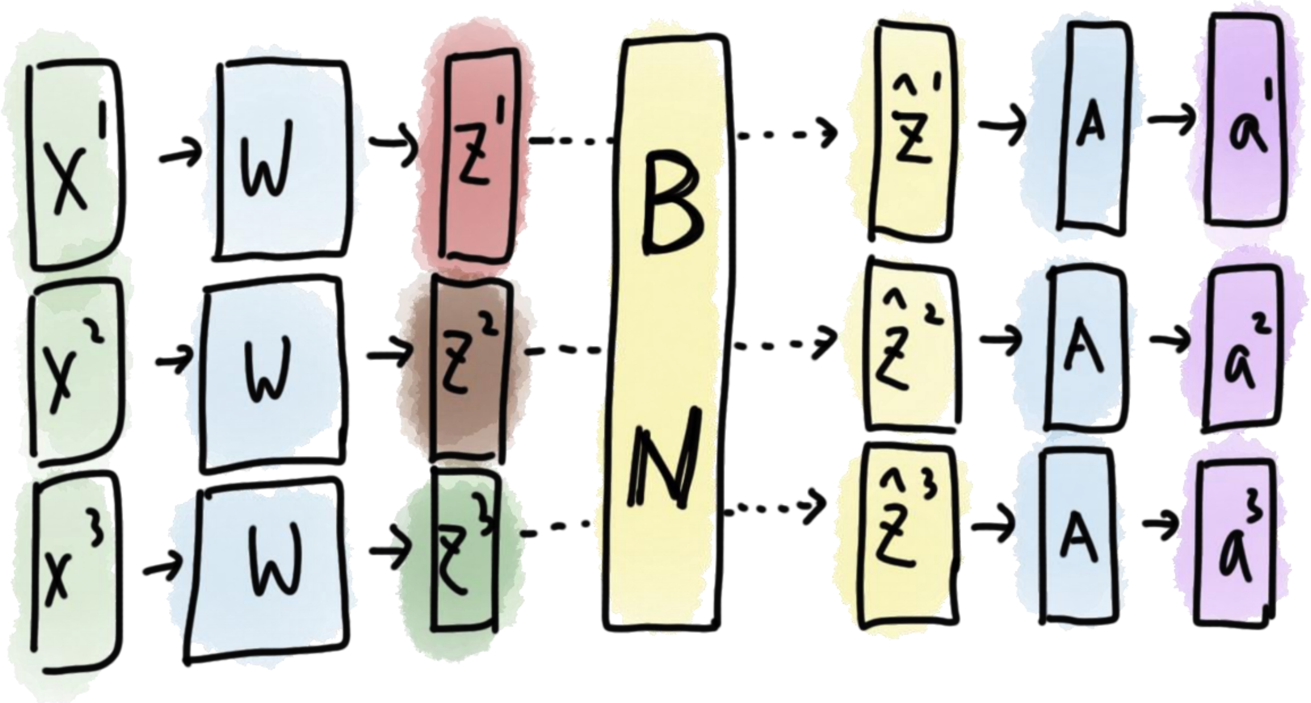


Internal Covariate Shift



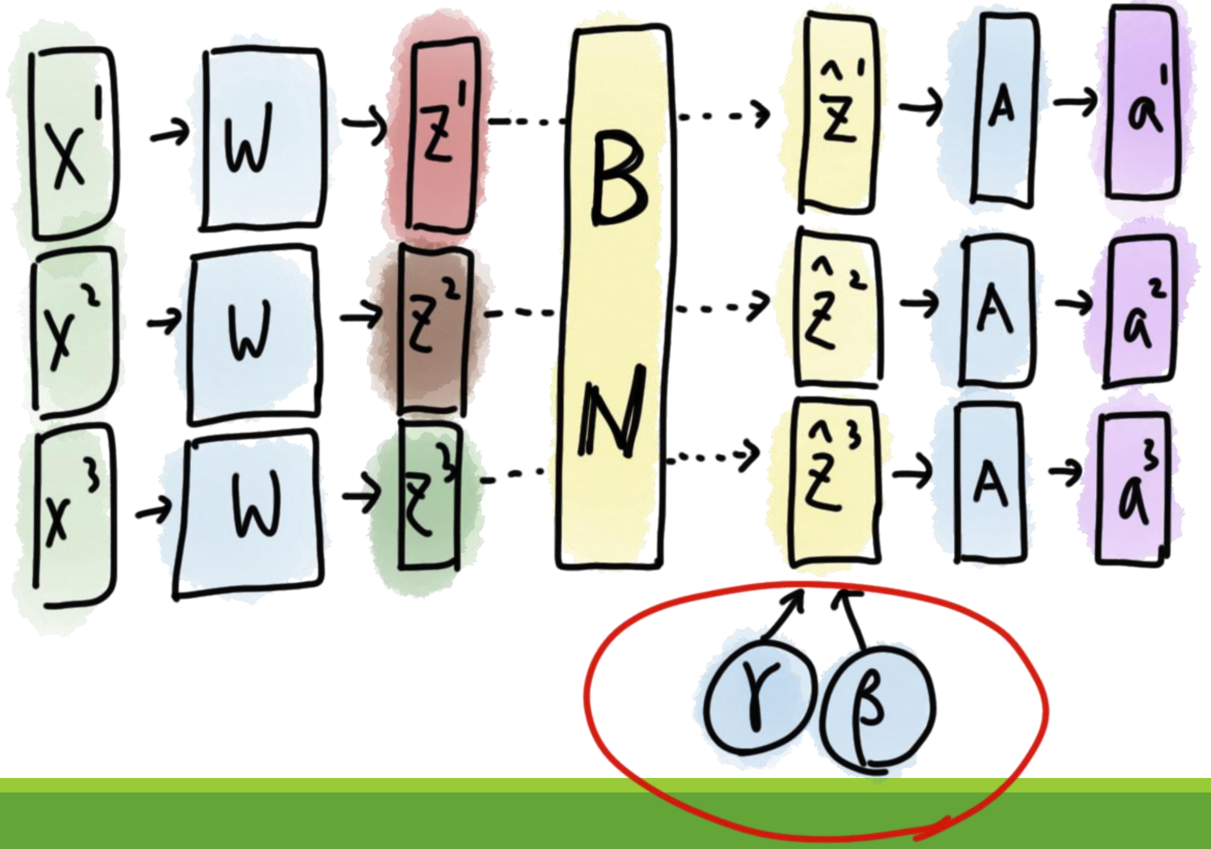
Batch Normalization

$$\hat{z}^i = \frac{z^i - \mu}{\sigma}$$



Batch Normalization

$$\hat{z}^i = \frac{z^i - \mu}{\sigma} \quad \hat{z}^i = \gamma \cdot \hat{z}^i + \beta$$

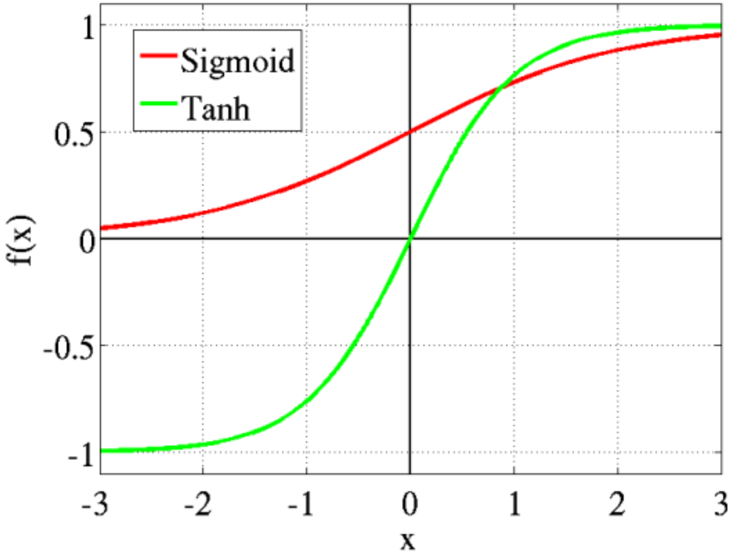


Batch Normalization

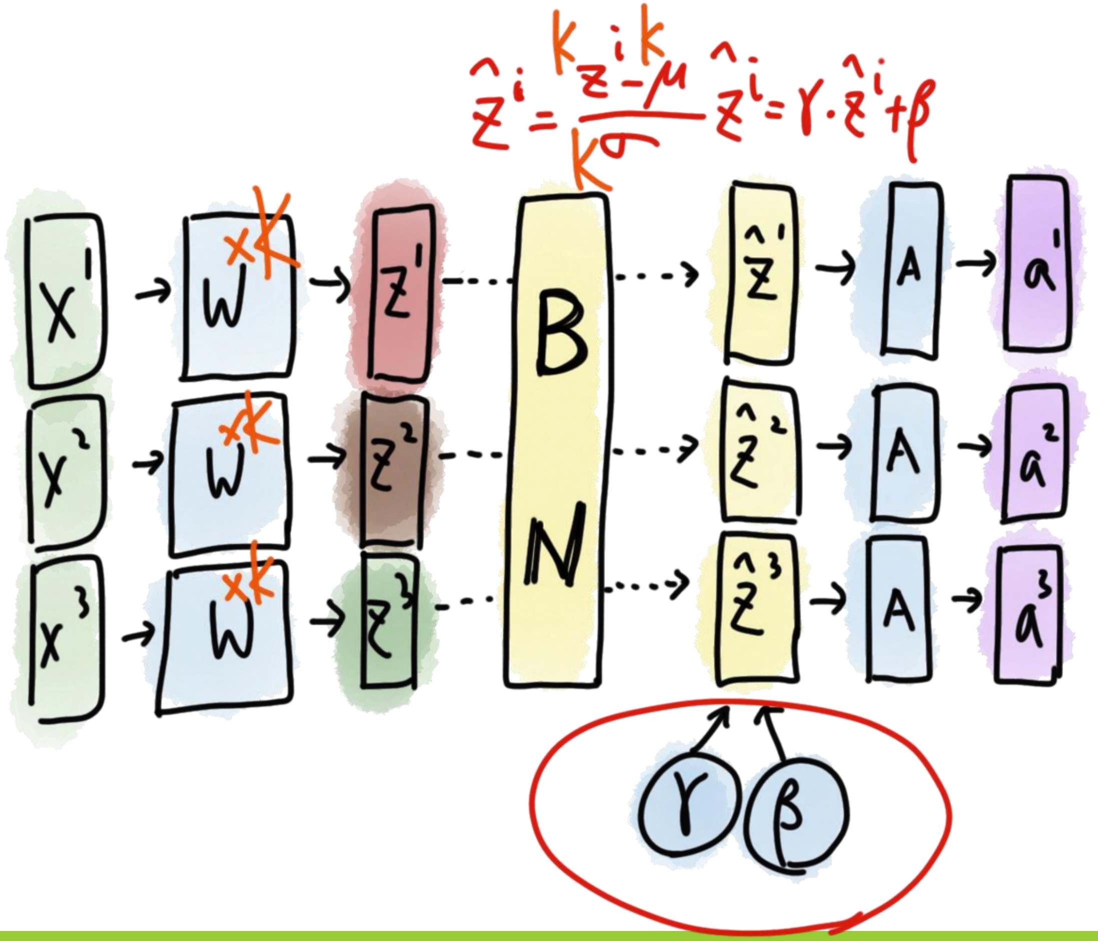
- learnable parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ to rescale and reshift distribution to preserve model capacity
- do not have “batch” in testing phase
- Ideal solution: computing mean and variance based on the whole training set
- practical solution: computing moving average of mean and variance of batches after convergence

Closer Look...

- Interval Covariate Shift?
- **avoid exploding/vanishing gradients**, especially for sigmoid and tanh activation functions
- usually apply before activation function
- when batch size is large

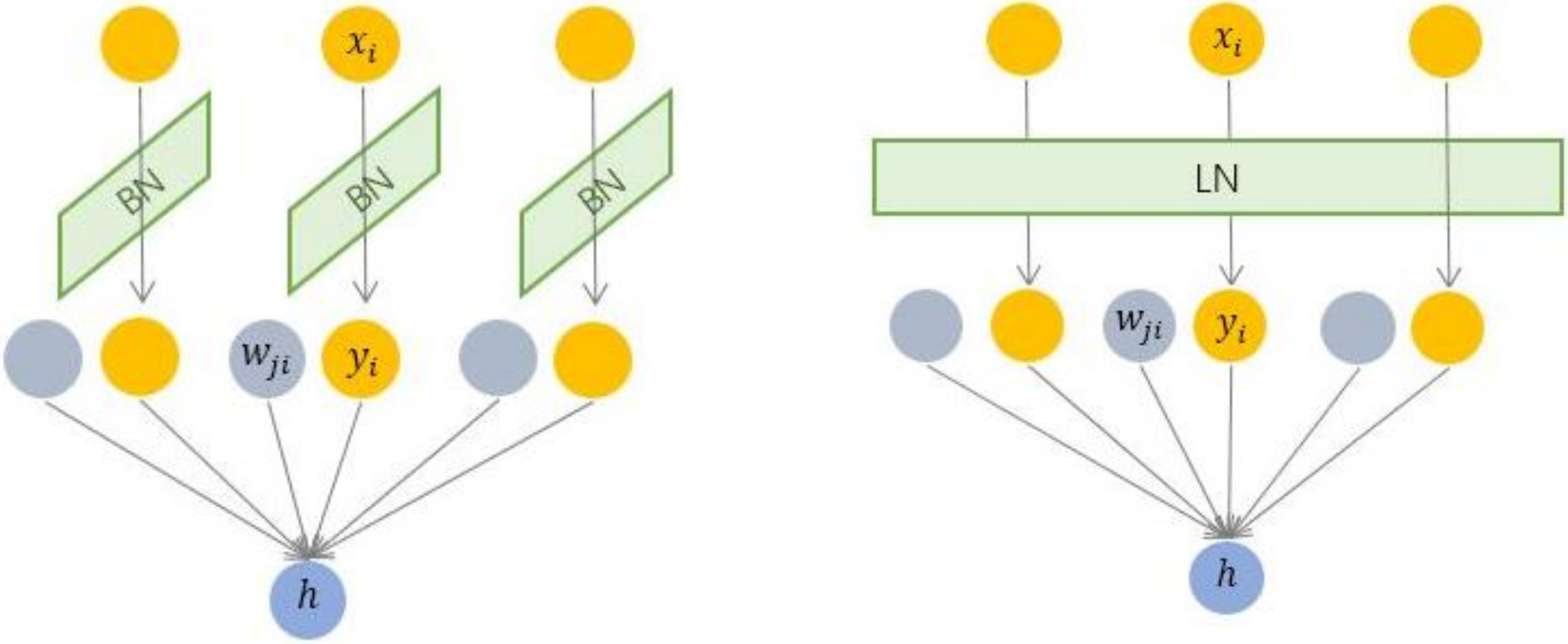


Closer Look...


$\hat{z}^i = \frac{kz^i - k\mu}{k\sigma}$
$\hat{z}^i = \gamma \cdot \hat{z}^i + \beta$
x^1
x^2
x^3
W xk
W xk
W xk
z^1
z^2
z^3
BN
$\hat{z}^1$
$\hat{z}^2$
$\hat{z}^3$
A
A
A
a^1
a^2
a^3
γ
β

Layer Normalization

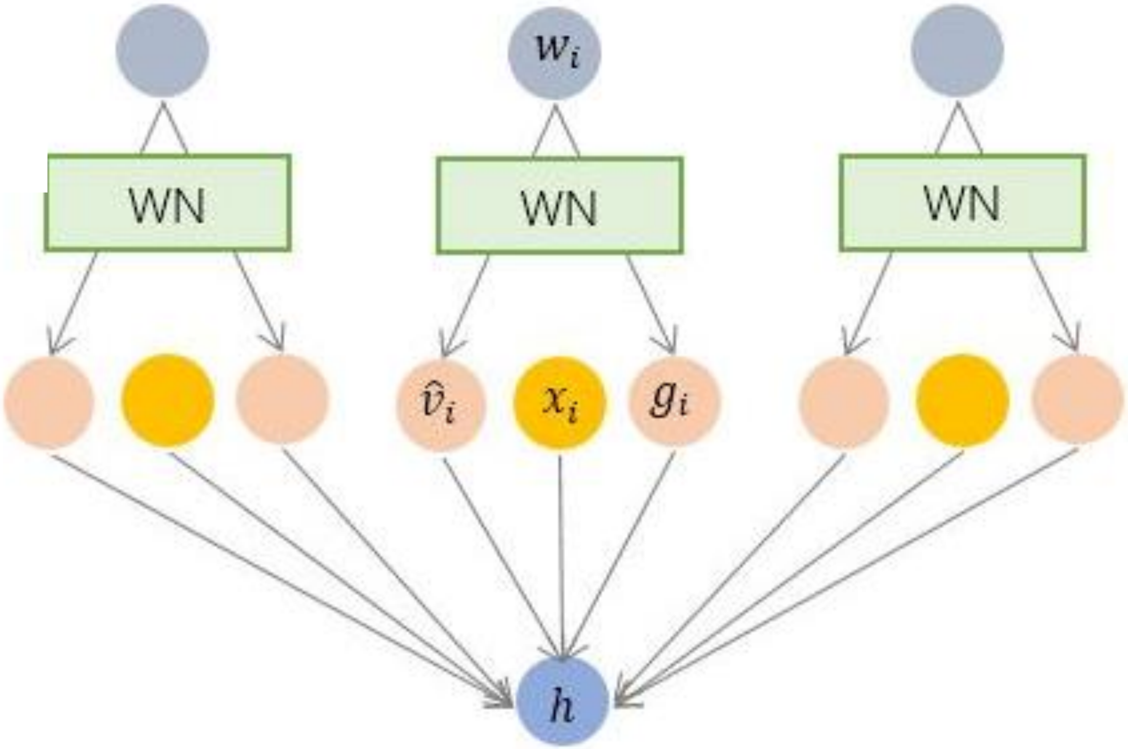
- can be used in (1) small batch scenario, even a single data sample and (2) dynamic network structures like RNN



Weight Normalization

- Reparameterization on weights

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



More

- Instance Normalization
- Group Normalization
- Spectral Normalization

references

- https://www.csie.ntu.edu.tw/~yvchen/f106-adl/doc/171116+171120_Tip.pdf
- https://zhuanlan.zhihu.com/p/33173246
- https://gab41.lab41.org/batch-normalization-what-the-hey-d480039a9e3b
- https://arxiv.org/pdf/1803.08494.pdf