

李宏毅 (Hung-yi Lee) · HYLEE | Machine Learning (2021)

HYLEE(2021) · 课程资料包 @ShowMeAI



视频

中英双语字幕



课件

一键打包下载



笔记

官方笔记翻译



代码

作业项目解析



视频 · B 站 [扫码或点击链接]

<https://www.bilibili.com/video/BV1fM4y137M4>



课件 & 代码 · 博客 [扫码或点击链接]

<http://blog.showmeai.tech/ntu-hylee-ml>

机器学习 Auto-encoder 生成式对抗网络 学习率
深度学习 卷积神经网络 GAN 自监督 自注意力机制
批次标准化 神经网络压缩 强化学习 元学习 Transformer

Awesome AI Courses Notes Cheatsheets 是 [ShowMeAI](#) 资料库的分支系列，覆盖最具知名度的 **TOP50+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets · 持续更新中			
知识图谱	图机器学习	深度强化学习	自动驾驶
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094



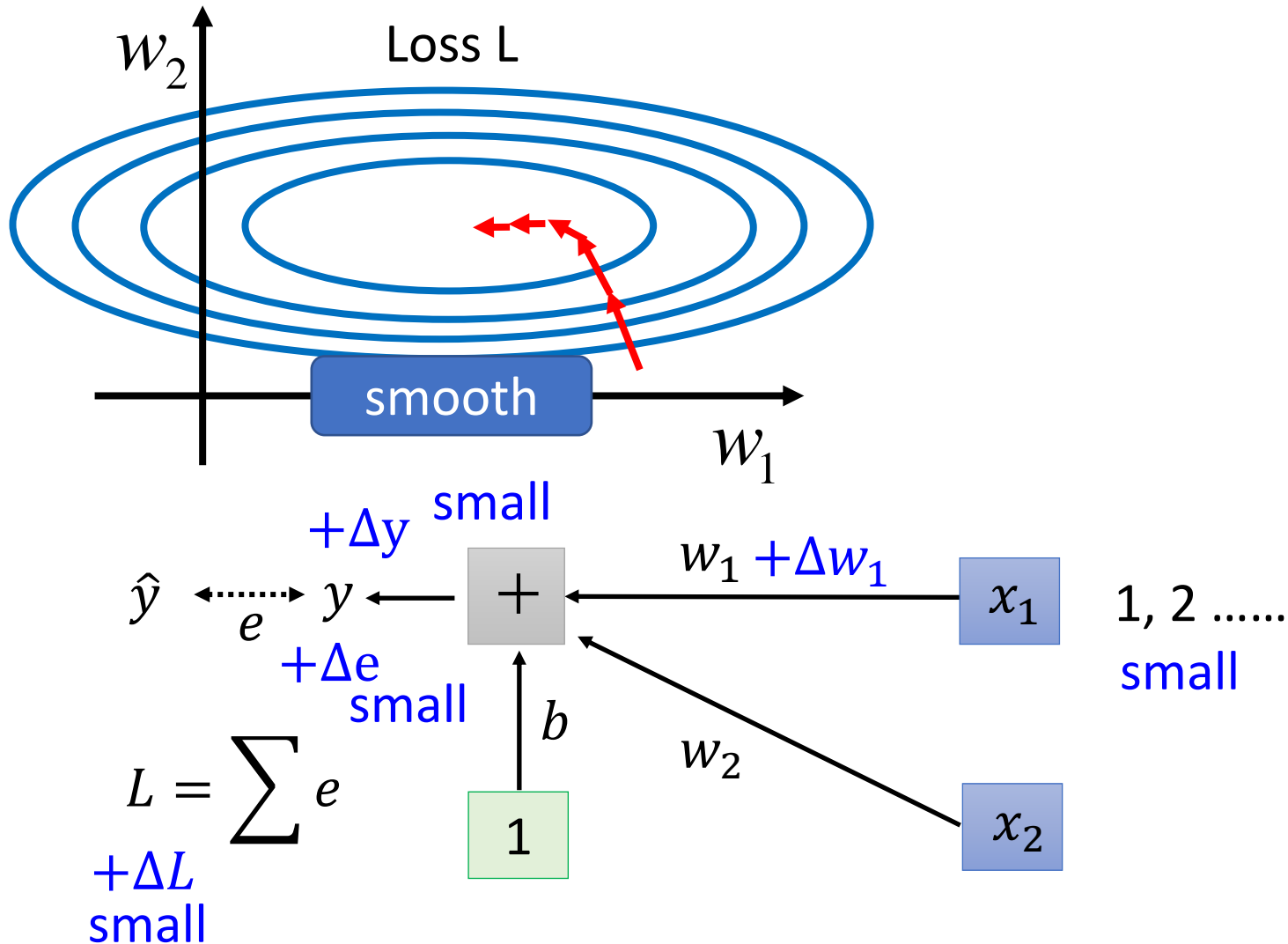
微信公众号

资料下载方式 2: 扫码点击**底部菜单栏**
称为 **AI 内容创作者**? 回复 [添砖加瓦]

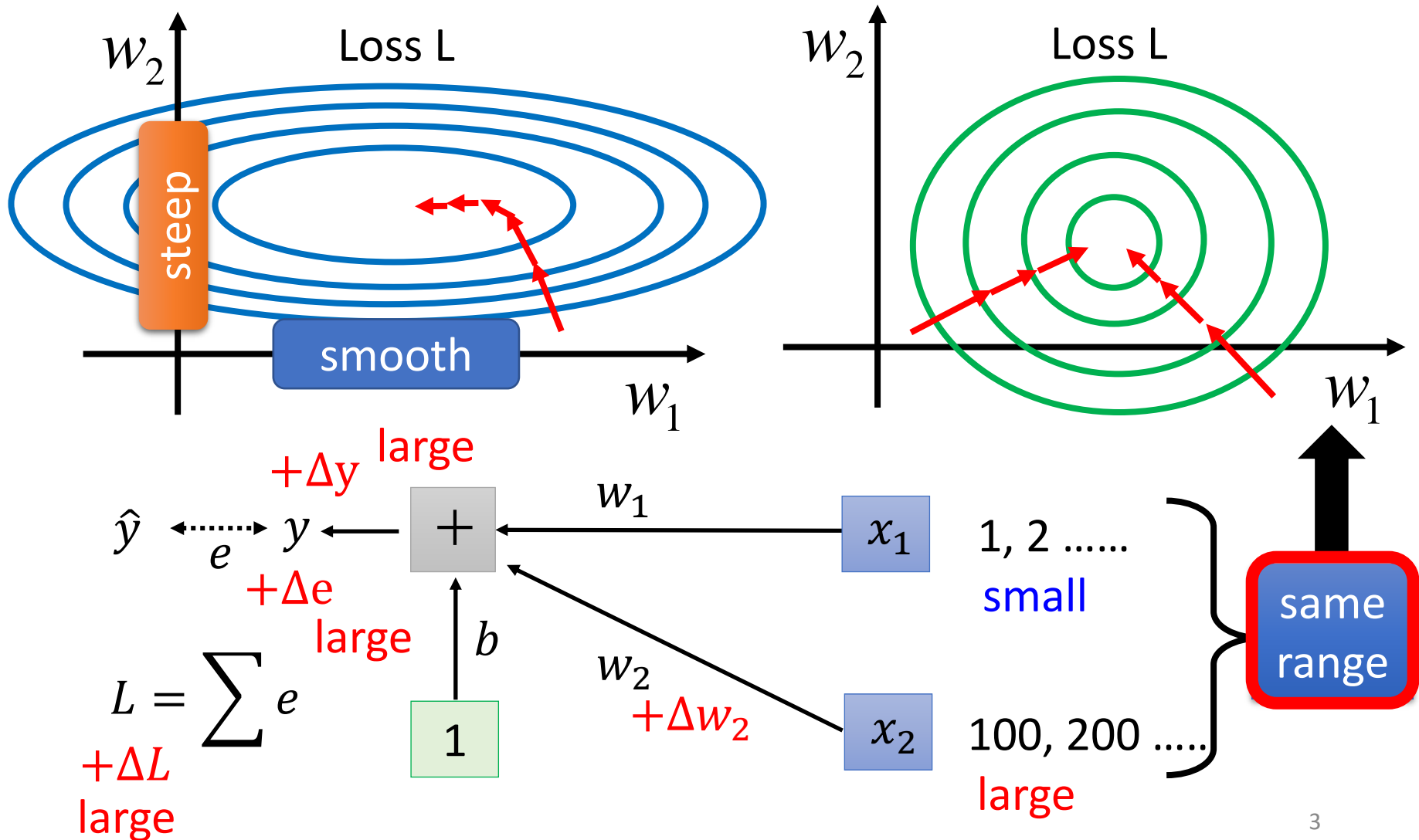
Quick Introduction of Batch Normalization

Hung-yi Lee 李宏毅

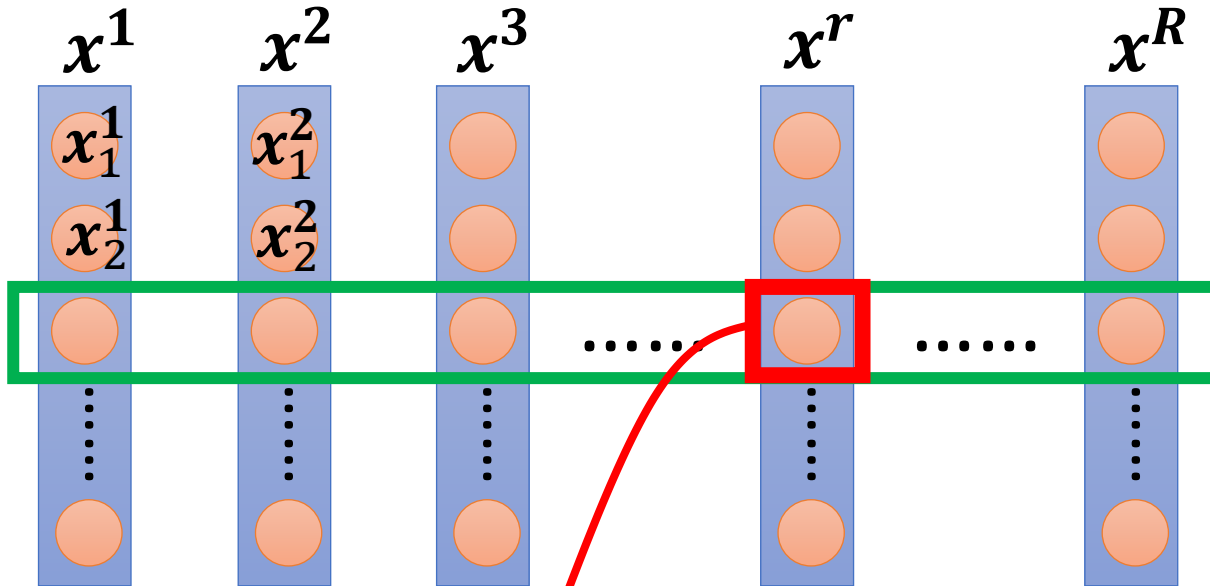
Changing Landscape



Changing Landscape



Feature Normalization



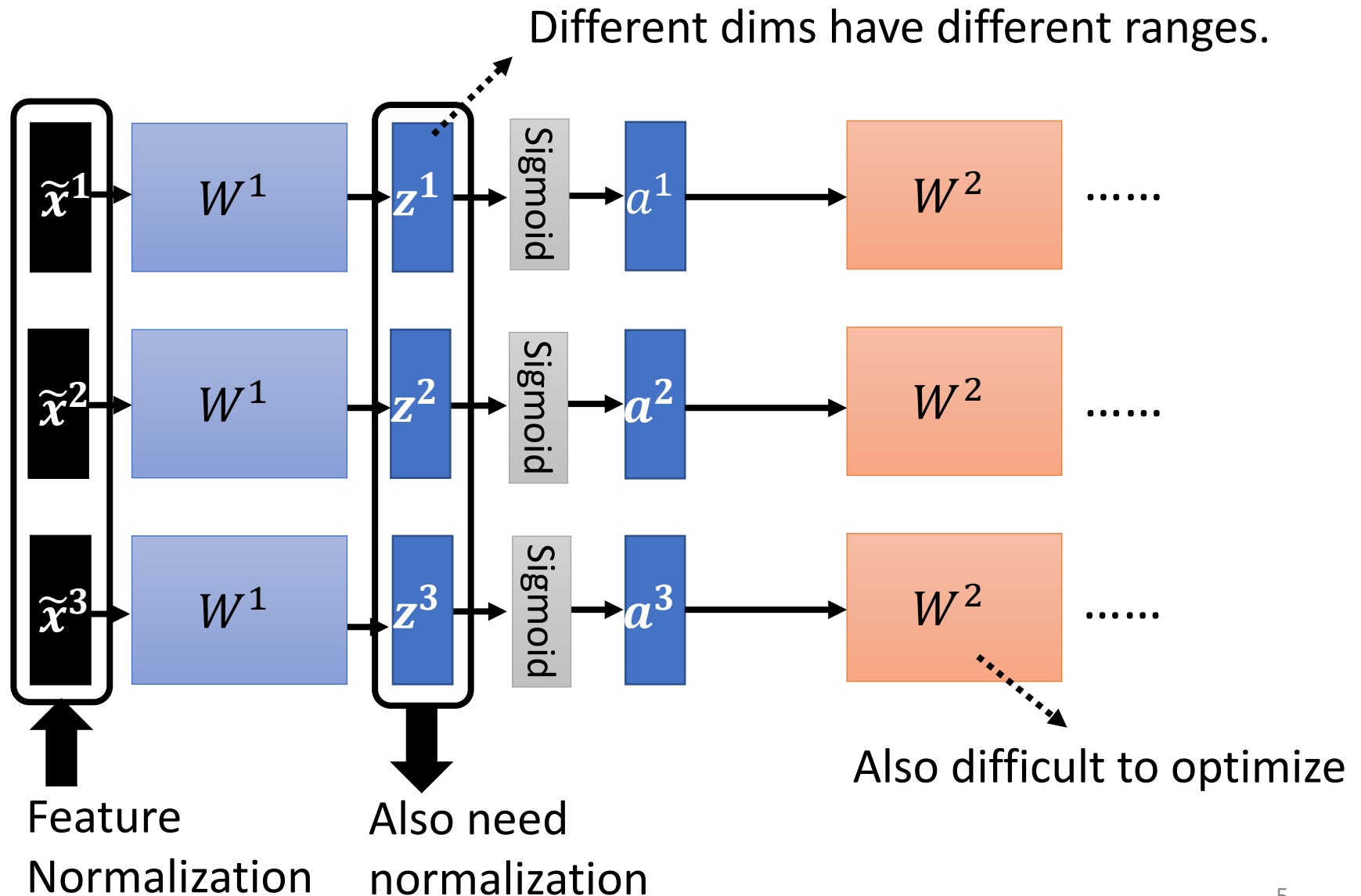
For each
dimension i :
mean: m_i
standard
deviation: σ_i

$$\tilde{x}_i^r \leftarrow \frac{x_i^r - m_i}{\sigma_i}$$

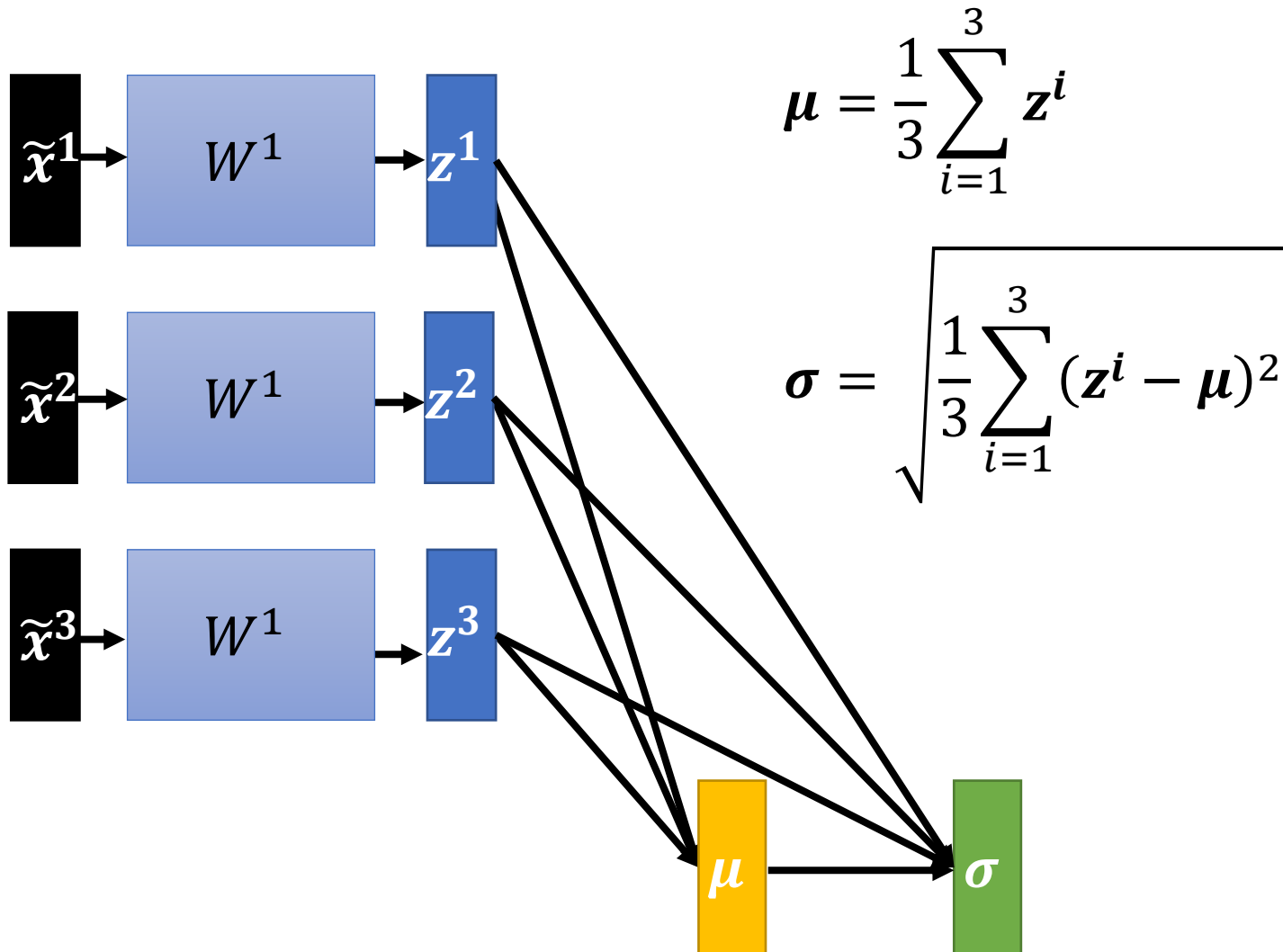
The means of all dims are 0,
and the variances are all 1

In general, feature normalization makes gradient descent converge faster.

Considering Deep Learning



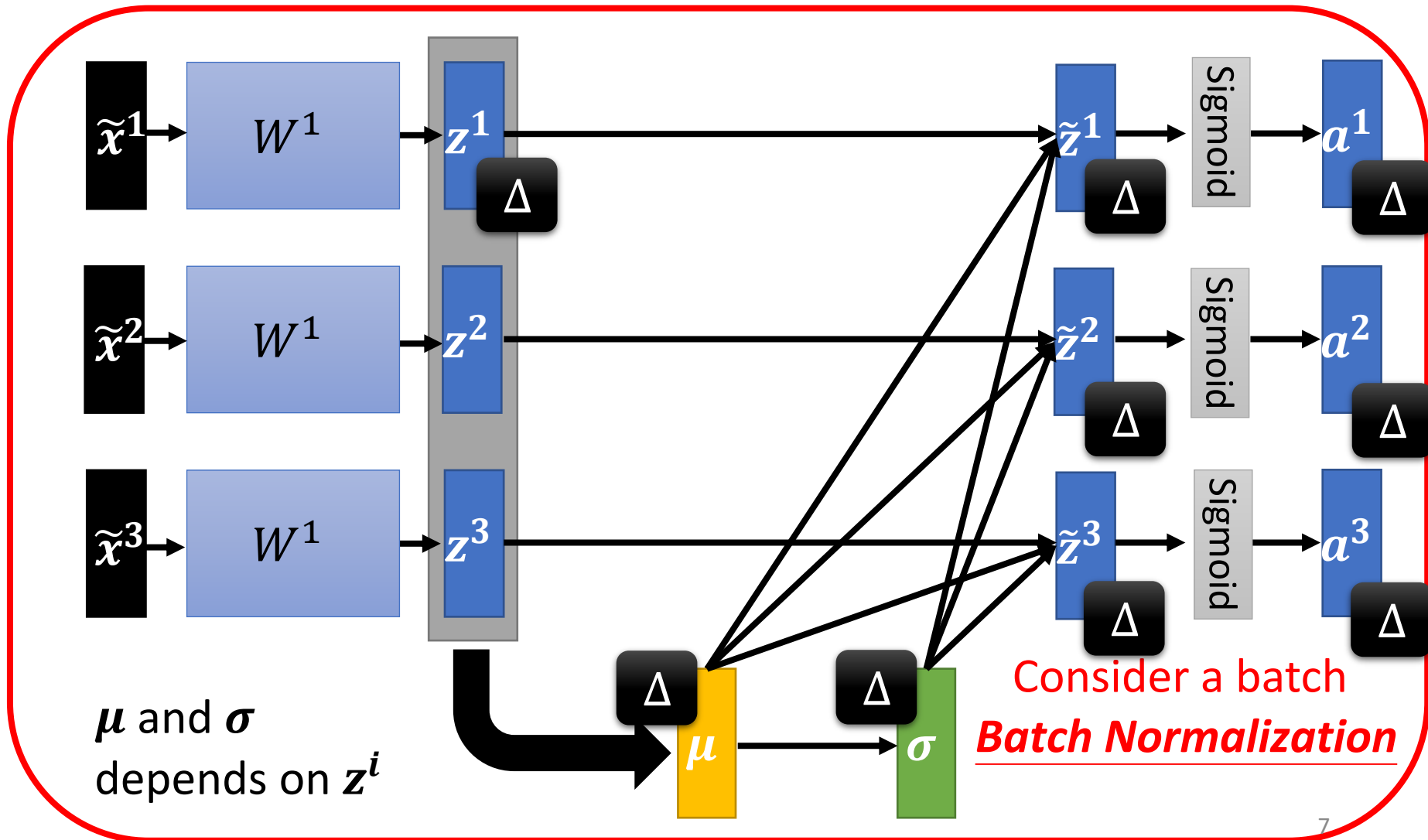
Considering Deep Learning



Considering Deep Learning

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

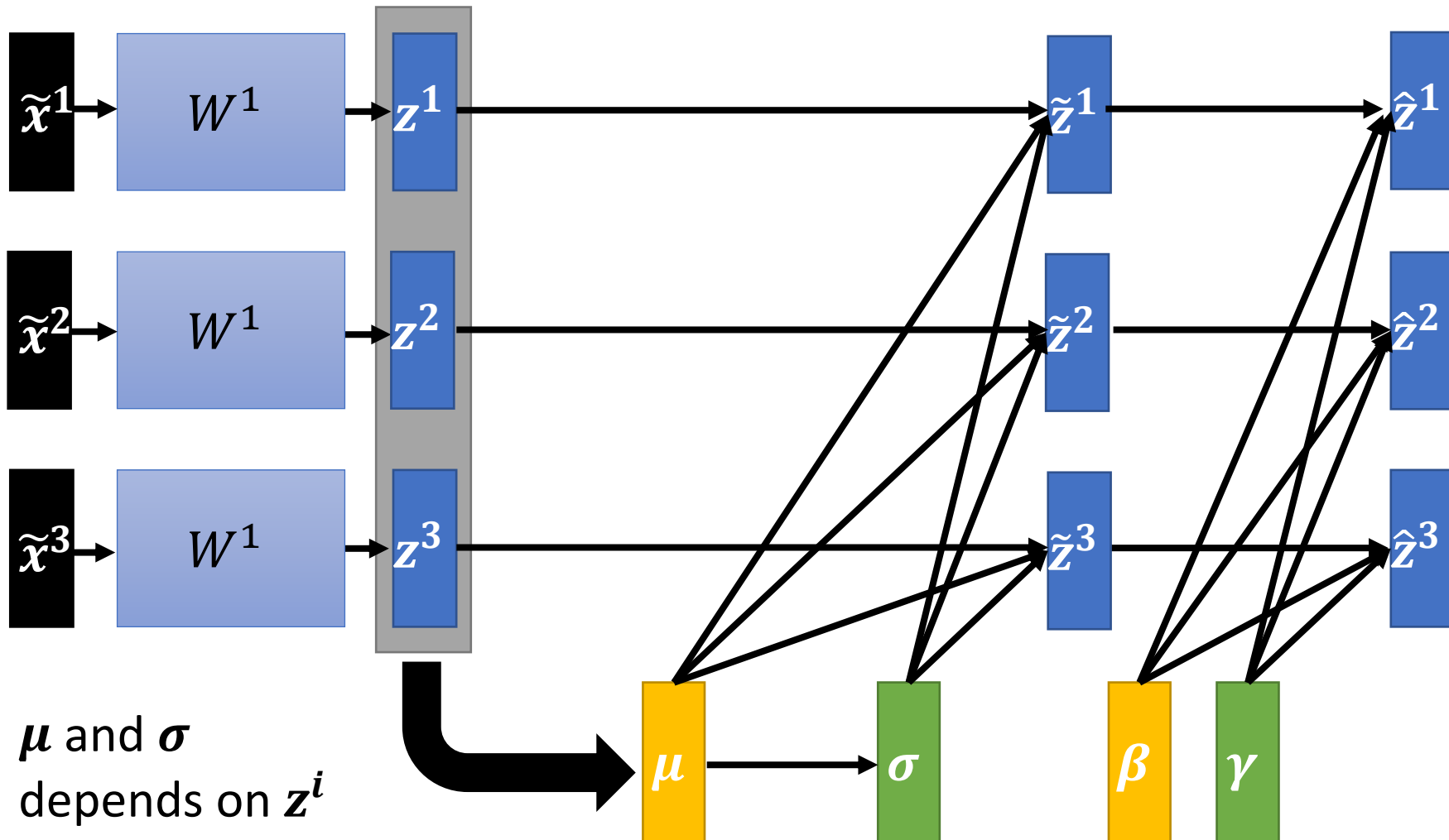
This is a large network!



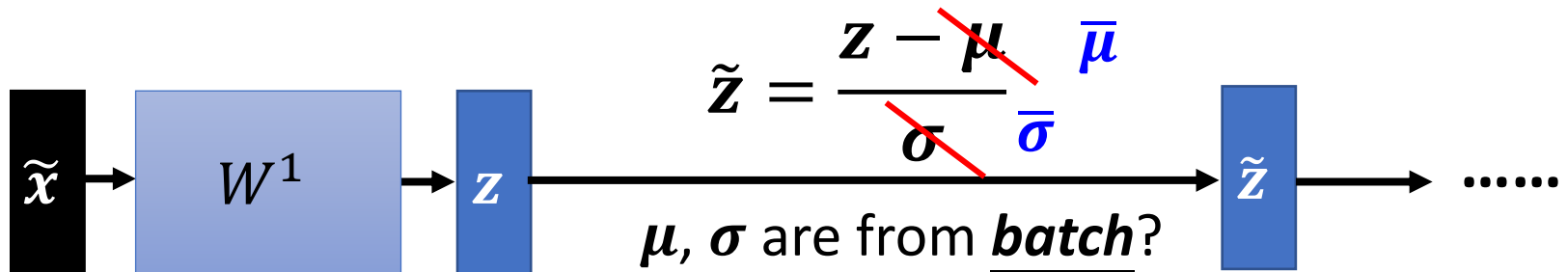
Batch normalization

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

$$\hat{\mathbf{z}}^i = \gamma \odot \tilde{\mathbf{z}}^i + \beta$$



Batch normalization – Testing



We do not always have batch at testing stage.

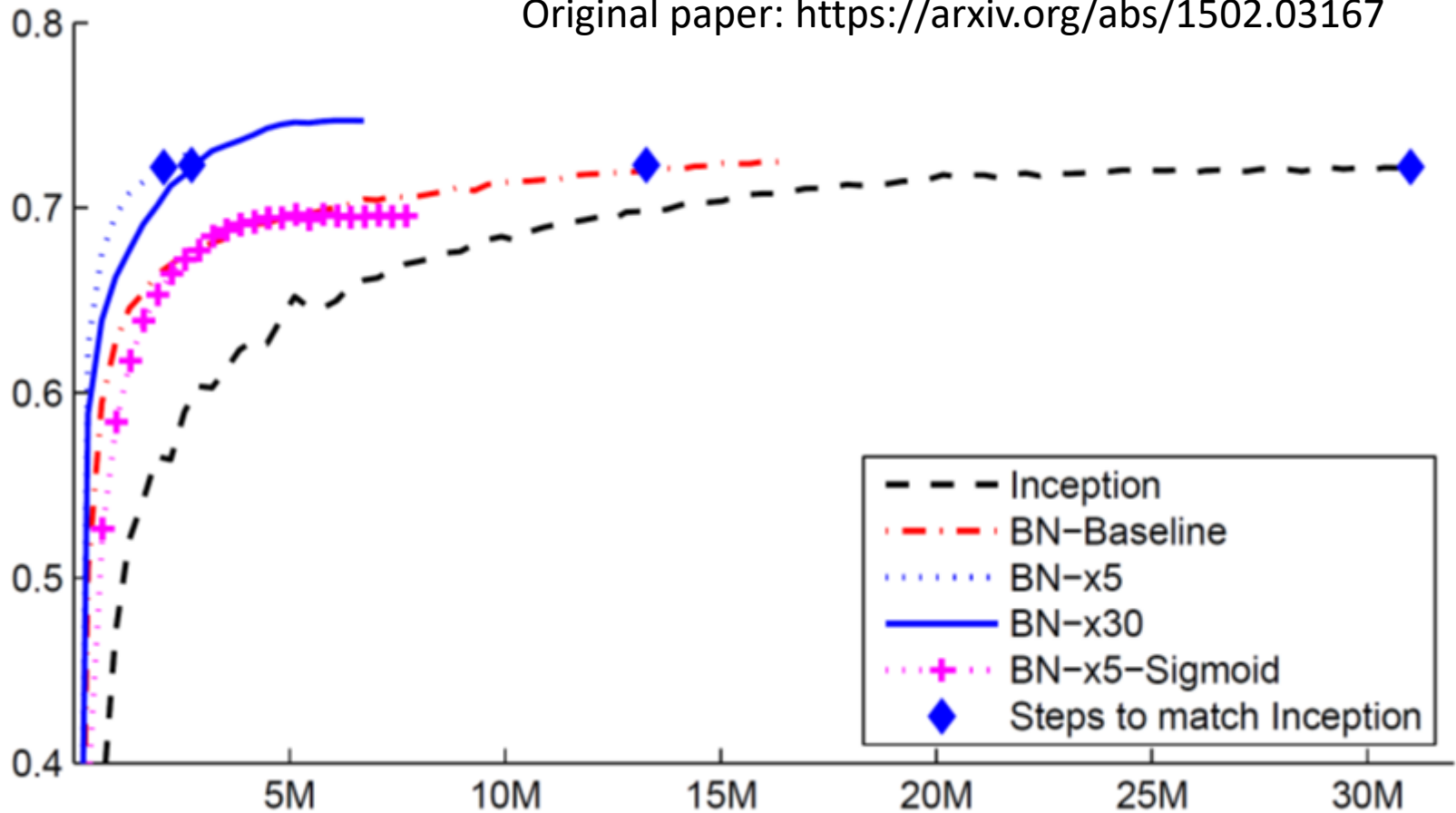
Computing the moving average of μ and σ of the batches during training.

$$\mu^1 \quad \mu^2 \quad \mu^3 \quad \dots \quad \mu^t$$

$$\bar{\mu} \leftarrow p\bar{\mu} + (1 - p)\mu^t$$

Batch normalization

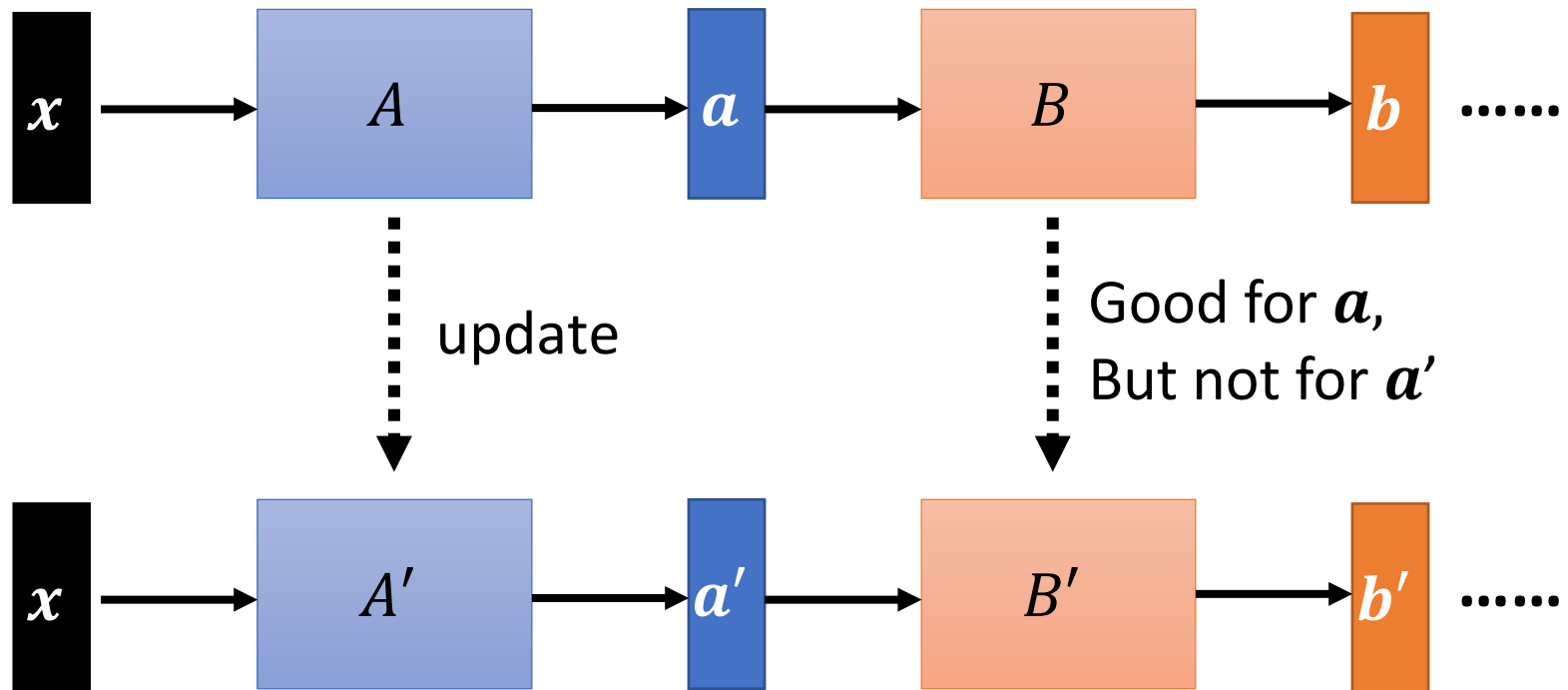
Original paper: <https://arxiv.org/abs/1502.03167>



Internal Covariate Shift?

How Does Batch Normalization Help Optimization?

<https://arxiv.org/abs/1805.11604>



Batch normalization make a and a' have similar statistics.
Experimental results do not support the above idea.

Internal Covariate Shift?

How Does Batch Normalization Help Optimization?

<https://arxiv.org/abs/1805.11604>

Experimental results (and theoretically analysis) support batch normalization change the landscape of error surface.

and 12 of Appendix B.) This suggests that the positive impact of BatchNorm on training might be somewhat serendipitous. Therefore, it might be valuable to perform a principled exploration of the design space of normalization schemes as it can lead to better performance.

serendipitous (偶然的)

penicillin



To learn more

- Batch Renormalization
 - <https://arxiv.org/abs/1702.03275>
- Layer Normalization
 - <https://arxiv.org/abs/1607.06450>
- Instance Normalization
 - <https://arxiv.org/abs/1607.08022>
- Group Normalization
 - <https://arxiv.org/abs/1803.08494>
- Weight Normalization
 - <https://arxiv.org/abs/1602.07868>
- Spectrum Normalization
 - <https://arxiv.org/abs/1705.10941>

Q&A

李宏毅 (Hung-yi Lee) · HYLEE | Machine Learning (2021)

HYLEE(2021) · 课程资料包 @ShowMeAI



视频

中英双语字幕



课件

一键打包下载



笔记

官方笔记翻译



代码

作业项目解析



视频 · B 站 [扫码或点击链接]

<https://www.bilibili.com/video/BV1fM4y137M4>



课件 & 代码 · 博客 [扫码或点击链接]

<http://blog.showmeai.tech/ntu-hylee-ml>

机器学习 Auto-encoder 生成式对抗网络 学习率
深度学习 卷积神经网络 GAN 自监督 自注意力机制
批次标准化 神经网络压缩 强化学习 元学习 Transformer

Awesome AI Courses Notes Cheatsheets 是 [ShowMeAI](#) 资料库的分支系列，覆盖最具知名度的 **TOP50+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets · 持续更新中			
知识图谱	图机器学习	深度强化学习	自动驾驶
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094



微信公众号

资料下载方式 2: 扫码点击**底部菜单栏**
称为 **AI 内容创作者**? 回复 [添砖加瓦]