

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

HYLEE(2021) · 课程资料包 @ShowMeAI



视频

课件

笔记

代码

中英双语字幕

一键打包下载

官方笔记翻译

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]

General Guidance

Hung-yi Lee 李宏毅

Framework of ML

Training data: $\{(\mathbf{x}^1, \hat{y}^1), (\mathbf{x}^2, \hat{y}^2), \dots, (\mathbf{x}^N, \hat{y}^N)\}$

Testing data: $\{\mathbf{x}^{N+1}, \mathbf{x}^{N+2}, \dots, \mathbf{x}^{N+M}\}$

Speech Recognition


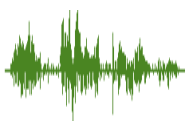
\mathbf{x} :  \hat{y} : phoneme

Image Recognition

\mathbf{x} :  \hat{y} : soup

Speaker Recognition

\mathbf{x} :  \hat{y} : John
(speaker)

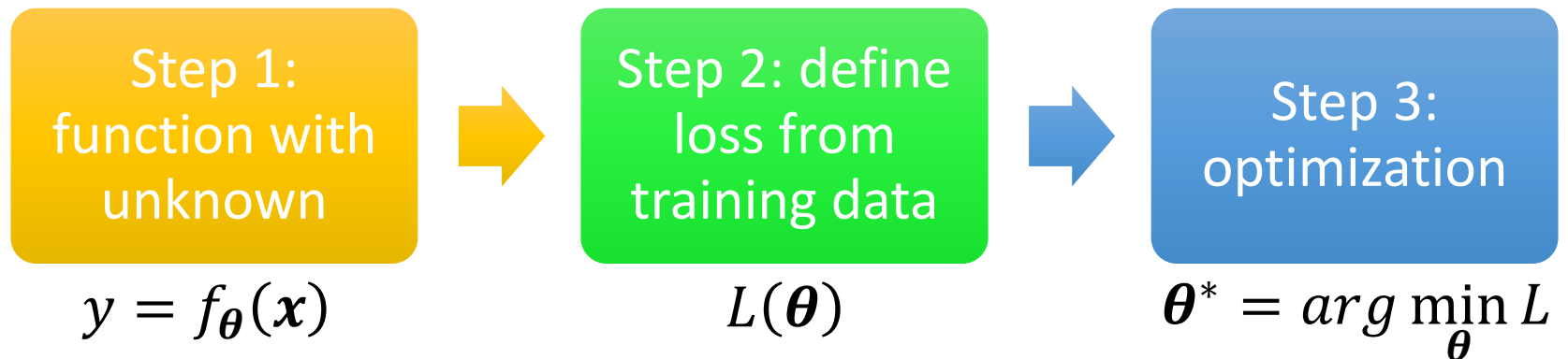
Machine Translation

\mathbf{x} : 痛みを知れ
 \hat{y} : 了解痛苦吧

Framework of ML

Training data: $\{(\mathbf{x}^1, \hat{y}^1), (\mathbf{x}^2, \hat{y}^2), \dots, (\mathbf{x}^N, \hat{y}^N)\}$

Training:

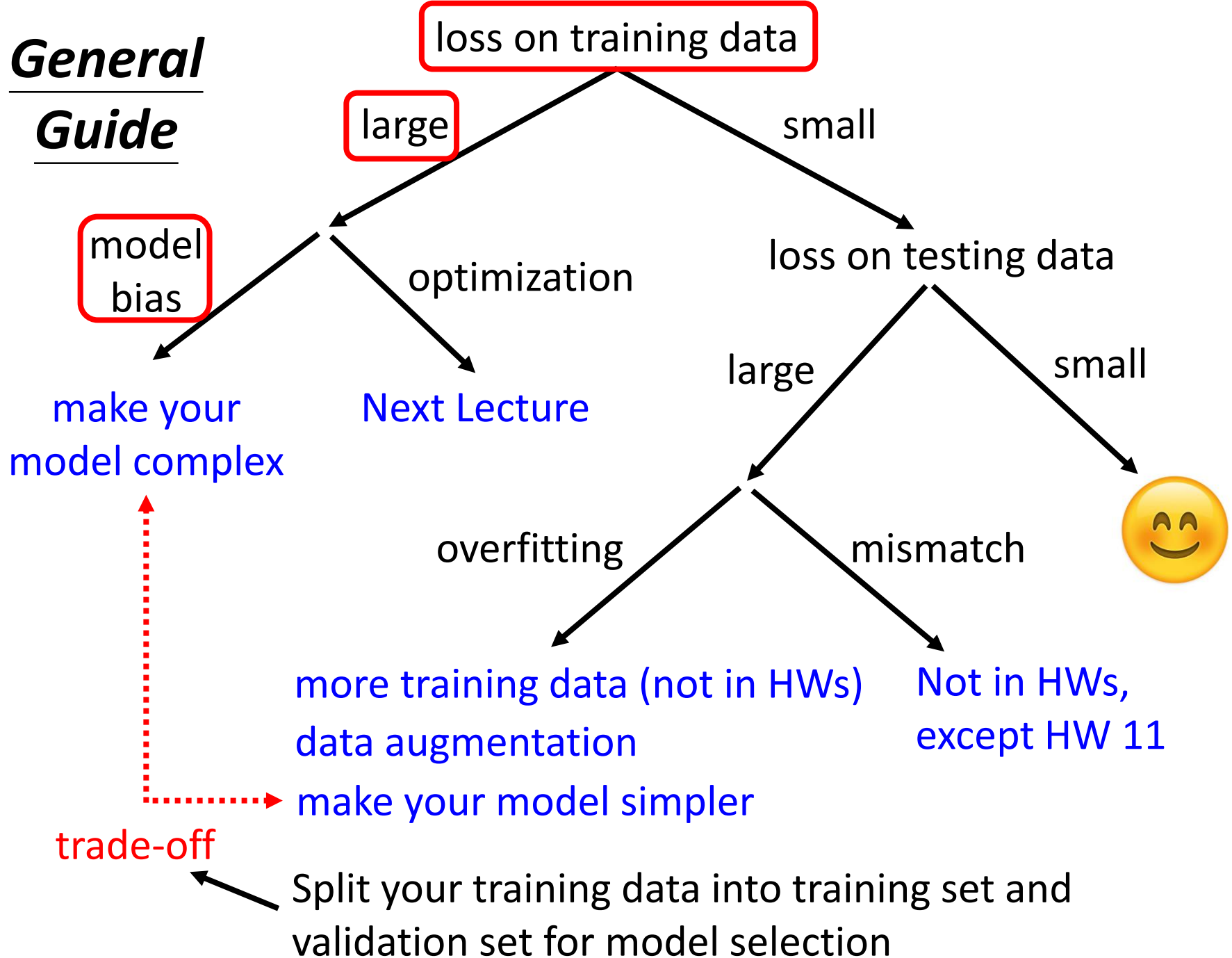


Testing data: $\{\mathbf{x}^{N+1}, \mathbf{x}^{N+2}, \dots, \mathbf{x}^{N+M}\}$

Use $y = f_{\theta^*}(\mathbf{x})$ to label the testing data

$\{y^{N+1}, y^{N+2}, \dots, y^{N+M}\}$ **➡** Upload to Kaggle

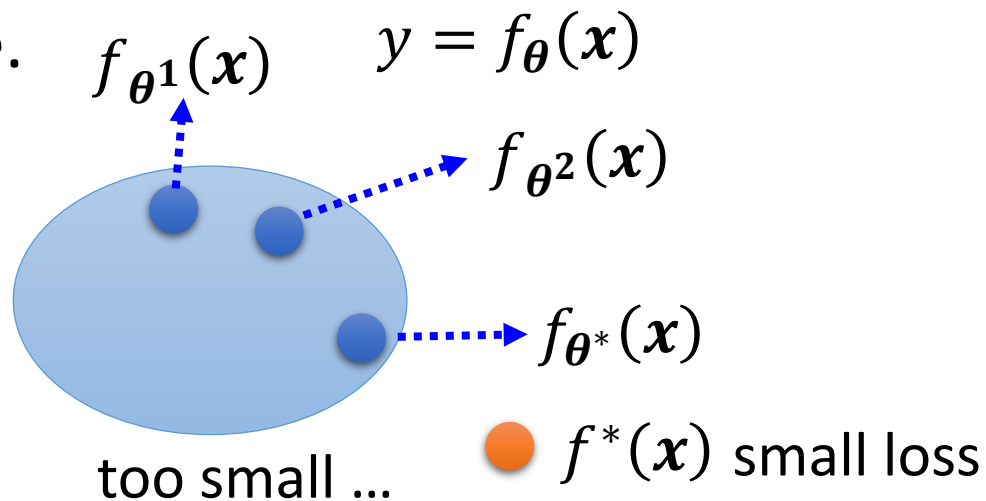
General Guide



Model Bias

- The model is too simple.

find a needle in a haystack ...
... but there is no needle



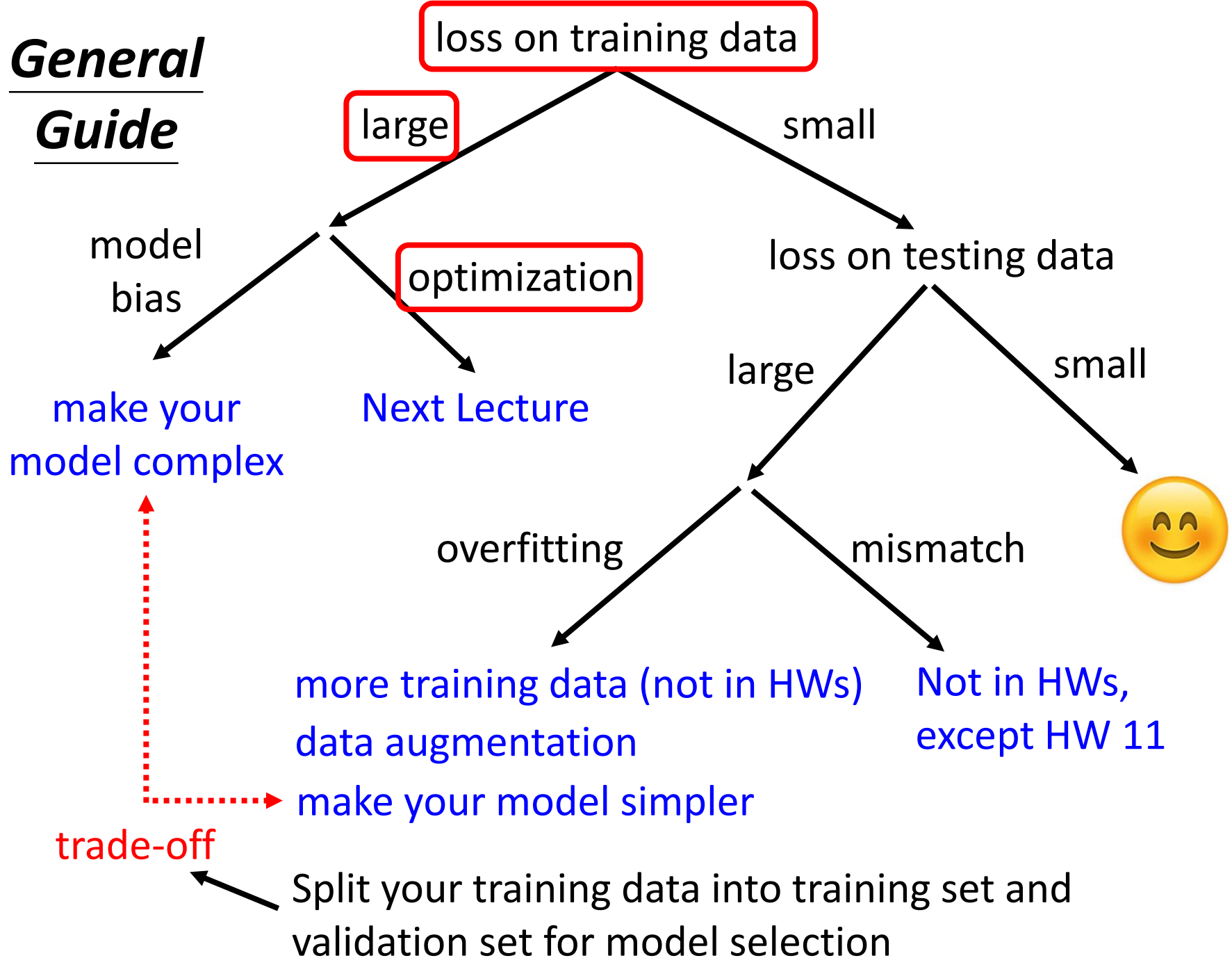
- Solution: redesign your model to make it more flexible

$$y = b + wx_1 \xrightarrow{\text{More features}} y = b + \sum_{j=1}^{56} w_j x_j$$

Deep Learning
(more neurons, layers)

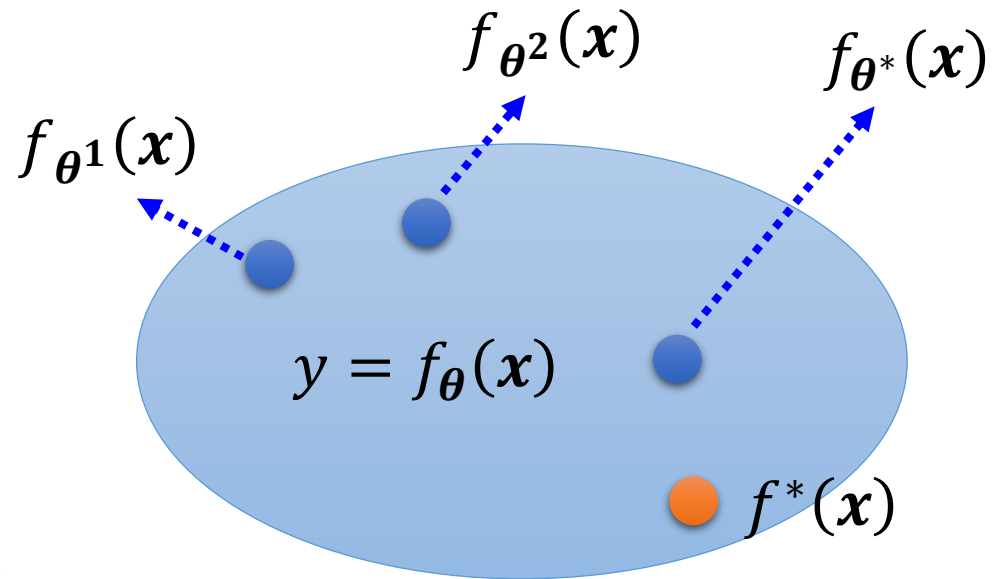
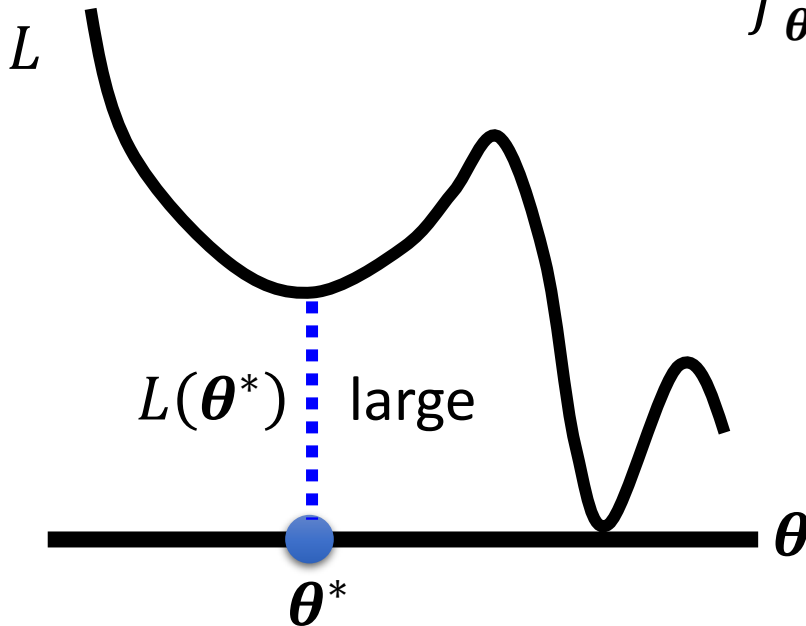
$$y = b + \sum_i c_i \operatorname{sigmoid} \left(b_i + \sum_j w_{ij} x_j \right)$$

General Guide



Optimization Issue

- Large loss not always imply model bias. There is another possibility ...

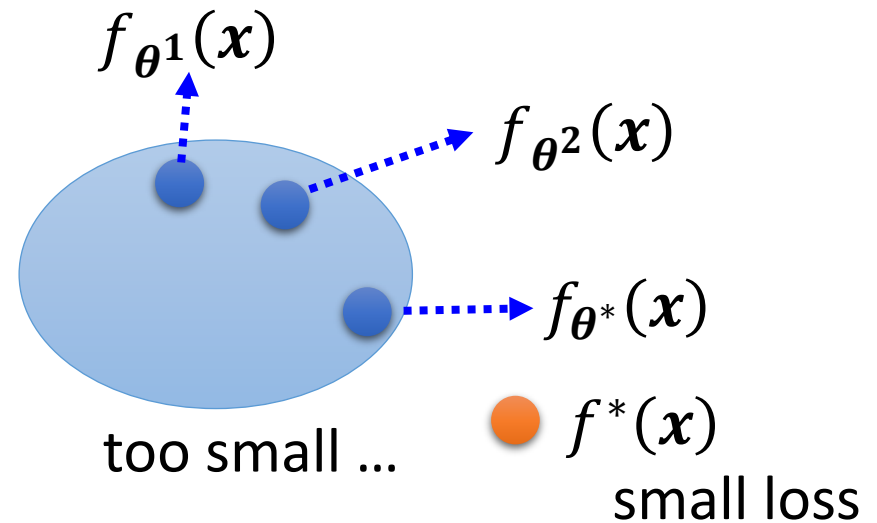


A needle is in a haystack ...

... Just cannot find it.

Model Bias

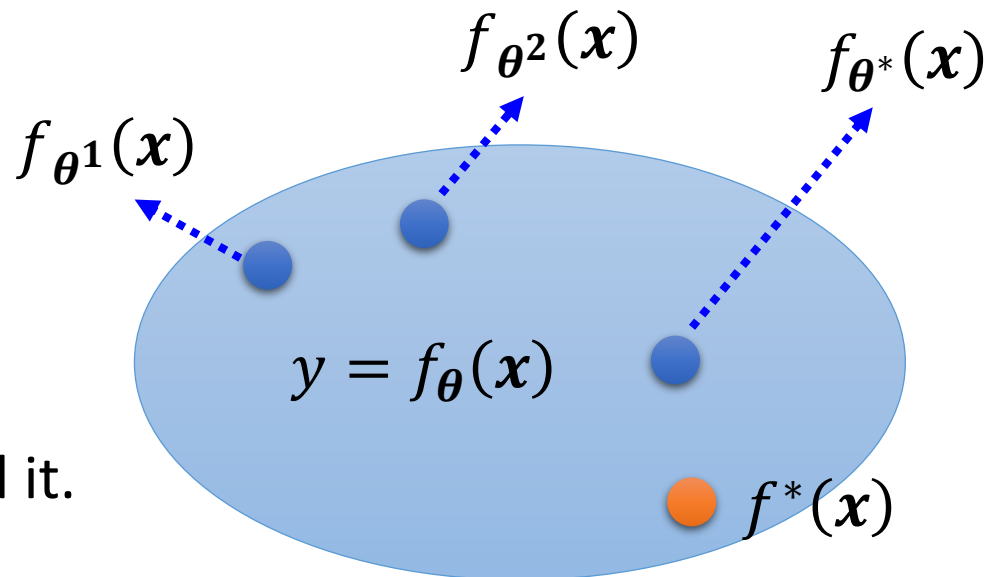
find a needle in a haystack ...
... but there is no needle



Which one???

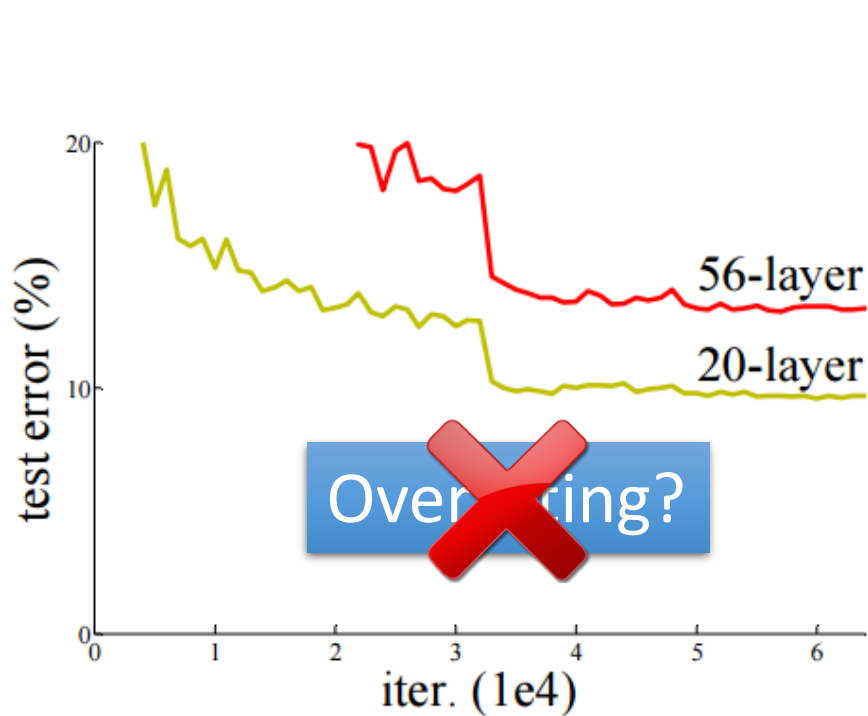
Optimization Issue

A needle is in a haystack ...
... Just cannot find it.

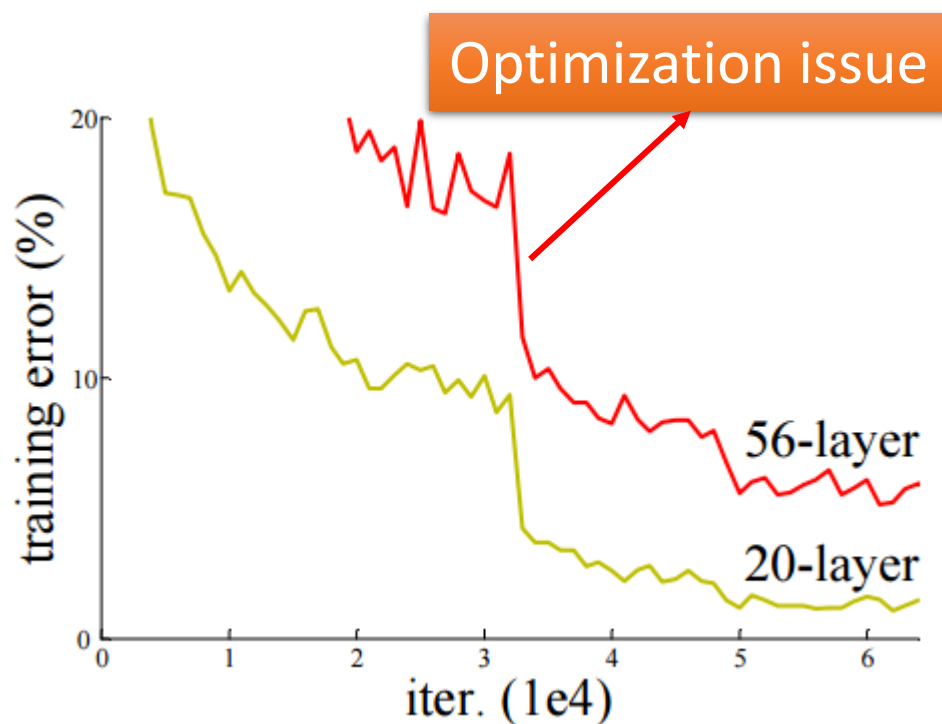


Model Bias v.s. Optimization Issue

- Gaining the insights from comparison



Testing Data



Training Data

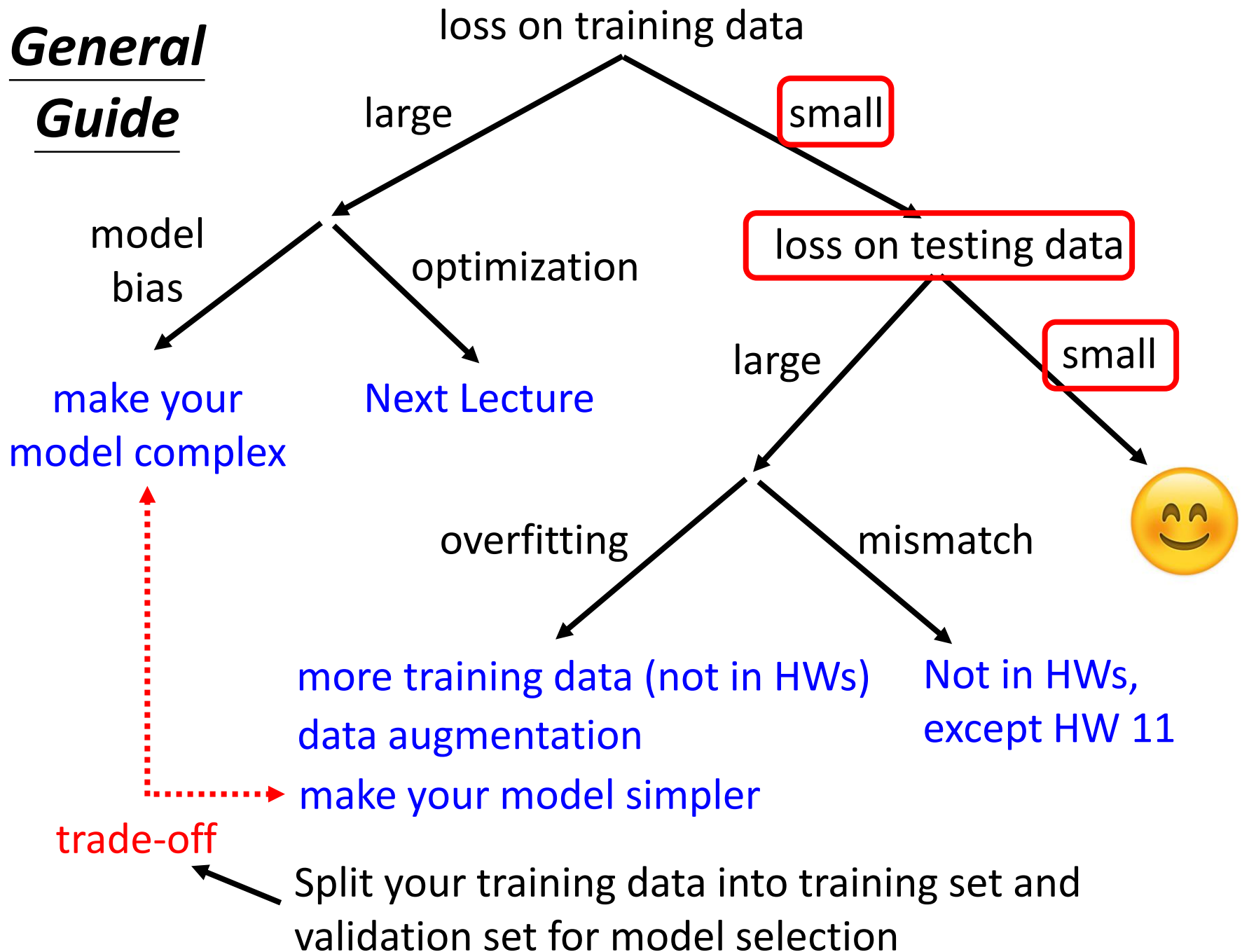
Optimization Issue

- Gaining the insights from comparison
- Start from shallower networks (or other models), which are easier to optimize.
- If deeper networks do not obtain smaller loss on **training data**, then there is optimization issue.

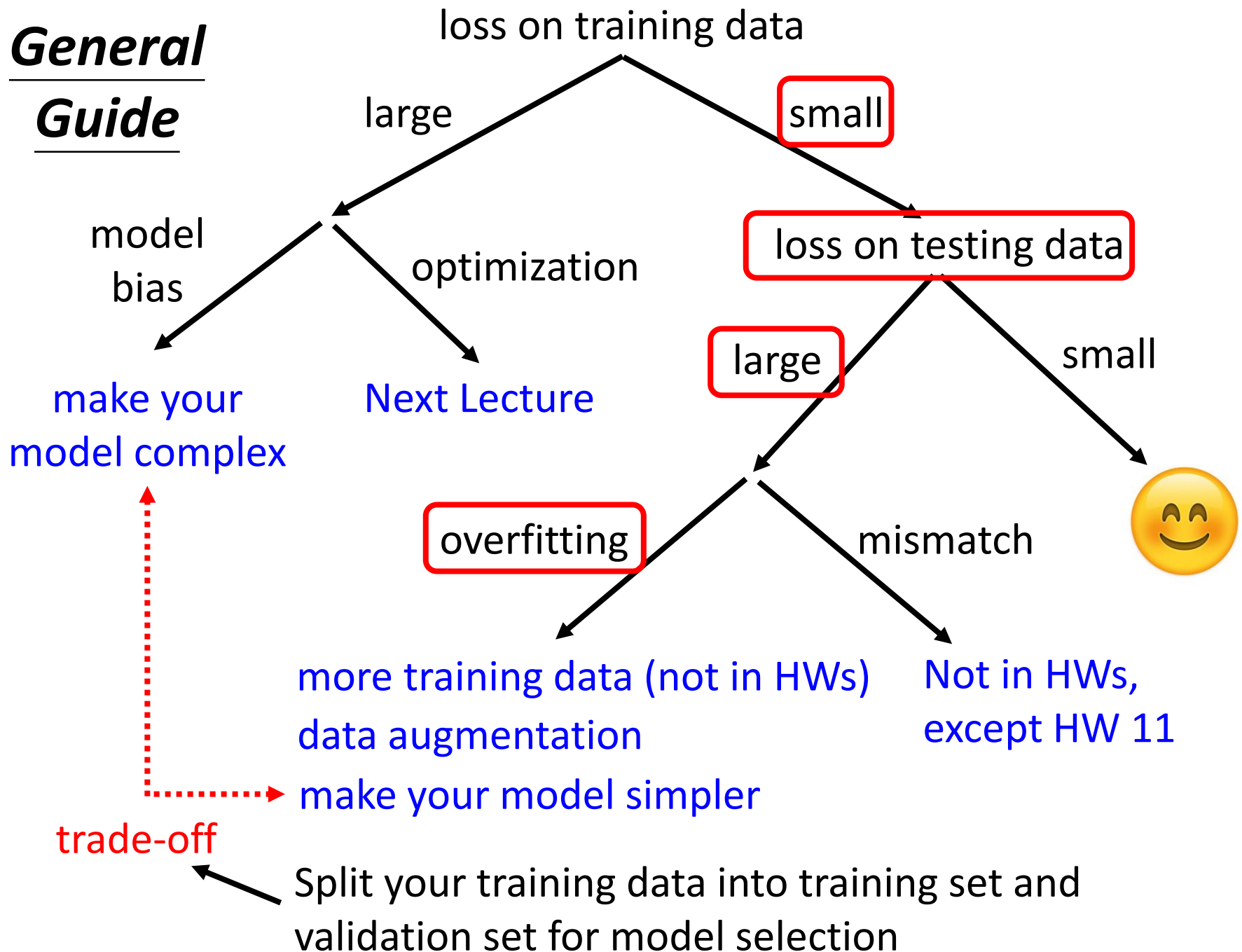
	1 layer	2 layer	3 layer	4 layer	5 layer
2017 – 2020	0.28k	0.18k	0.14k	0.10k	0.34k

- Solution: More powerful optimization technology (next lecture)

General Guide



General Guide



Overfitting

- Small loss on training data, large loss on testing data. Why?

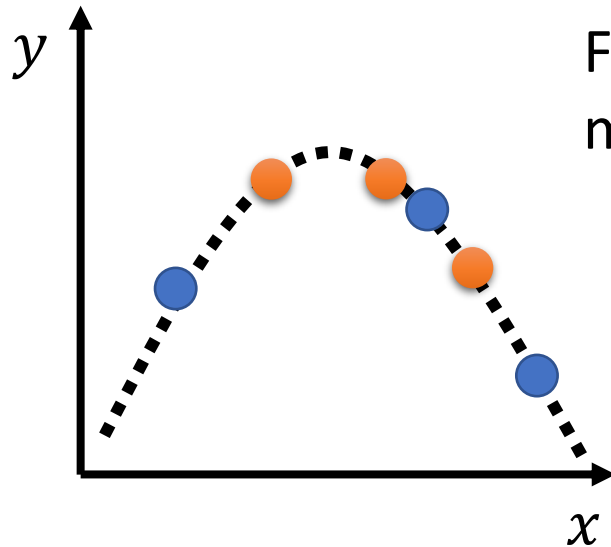
An extreme example

Training data: $\{(\mathbf{x}^1, \hat{y}^1), (\mathbf{x}^2, \hat{y}^2), \dots, (\mathbf{x}^N, \hat{y}^N)\}$

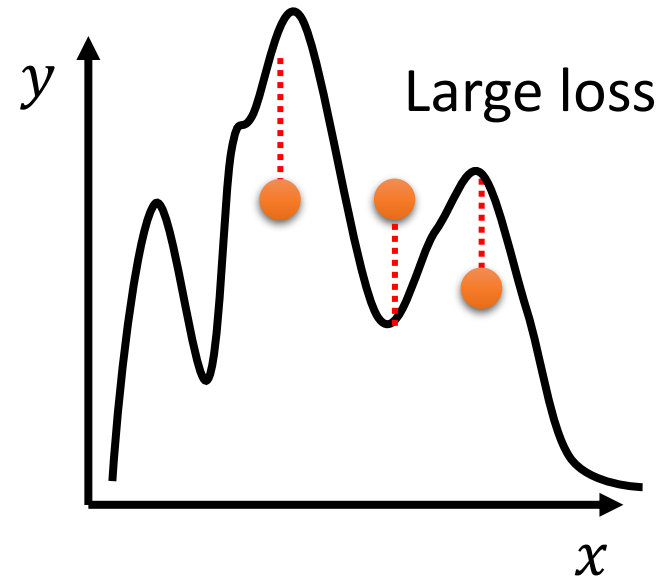
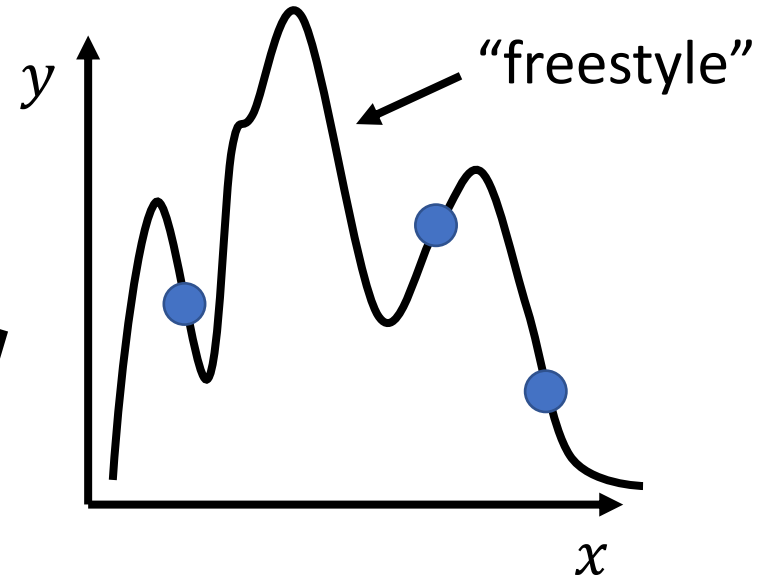
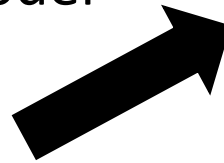
$$f(\mathbf{x}) = \begin{cases} \hat{y}^i & \exists \mathbf{x}^i = \mathbf{x} \\ random & otherwise \end{cases} \quad \text{Less than useless ...}$$

This function obtains **zero training loss**, but **large testing loss**.

Overfitting



Flexible
model

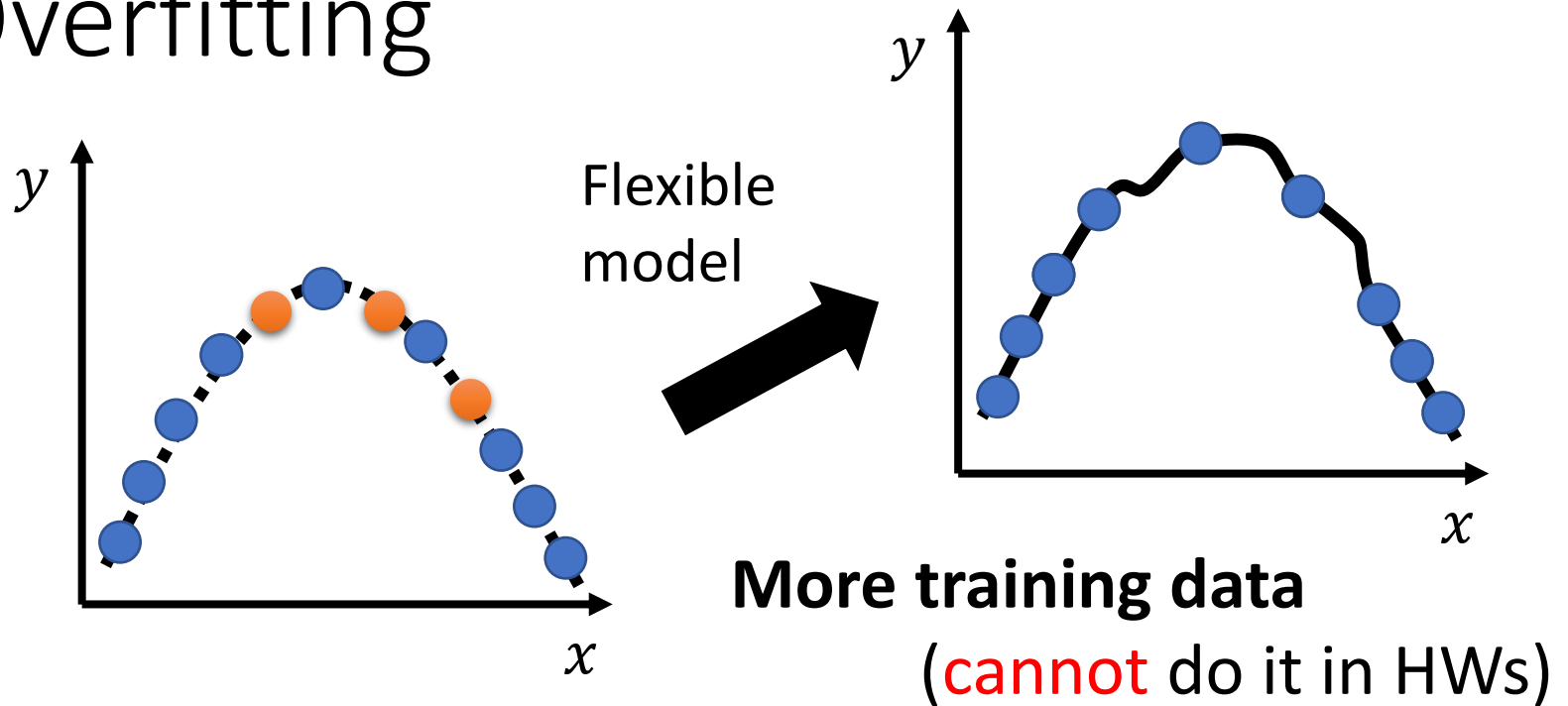


---- Real data distribution
(not observable)

● Training data

● Testing data

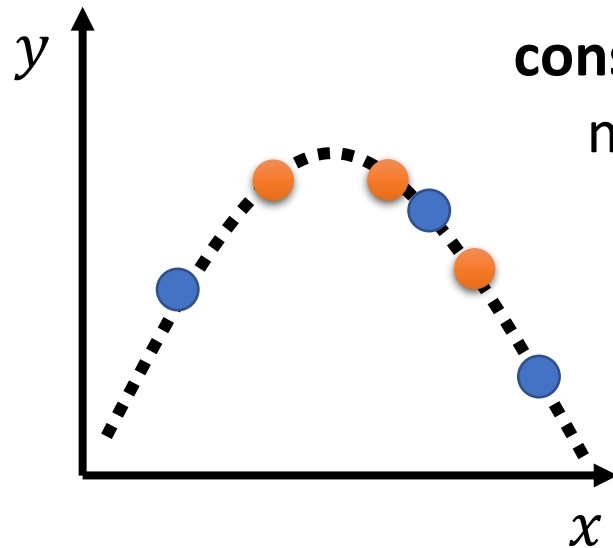
Overfitting



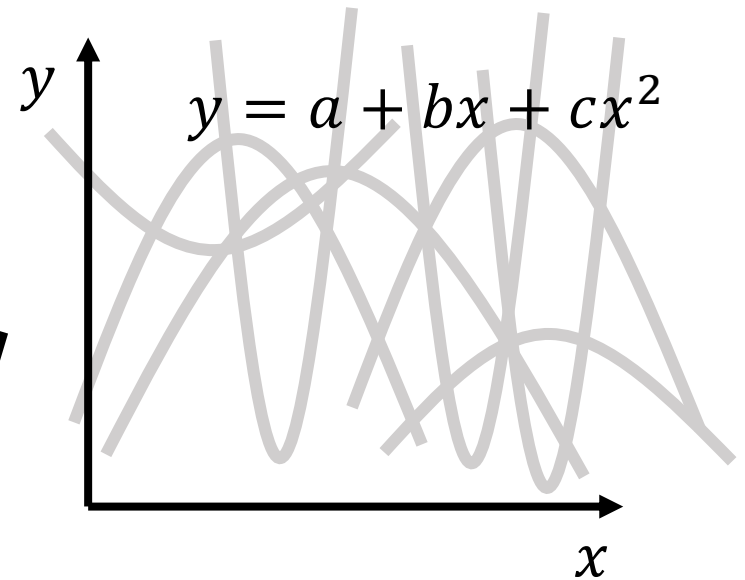
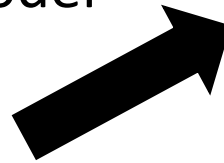
Data augmentation (you can do that in HWs)



Overfitting

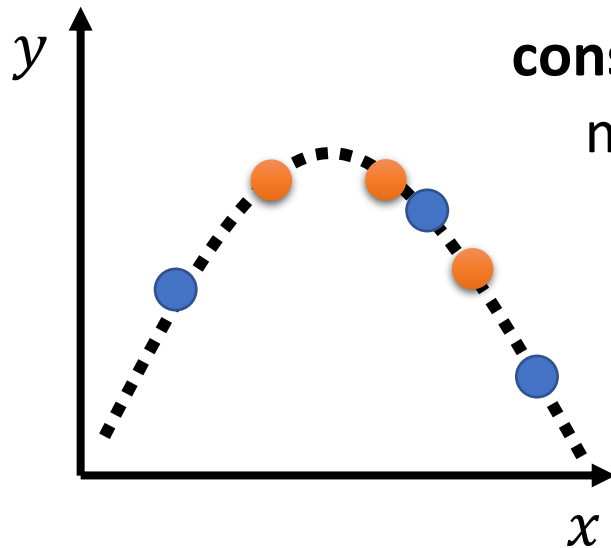


**constrained
model**

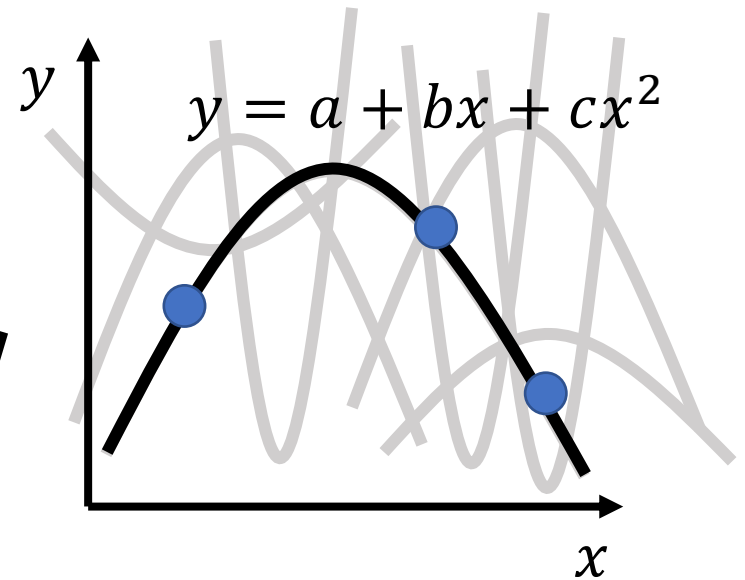
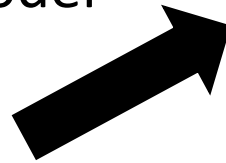


- Real data distribution
(not observable)
- Training data
- Testing data

Overfitting



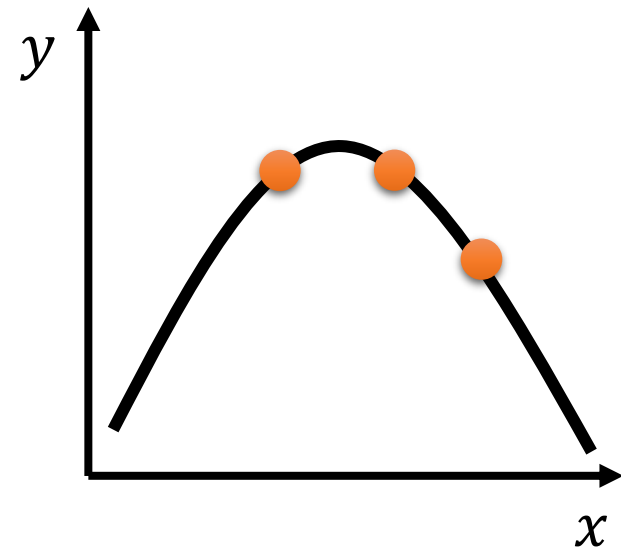
**constrained
model**



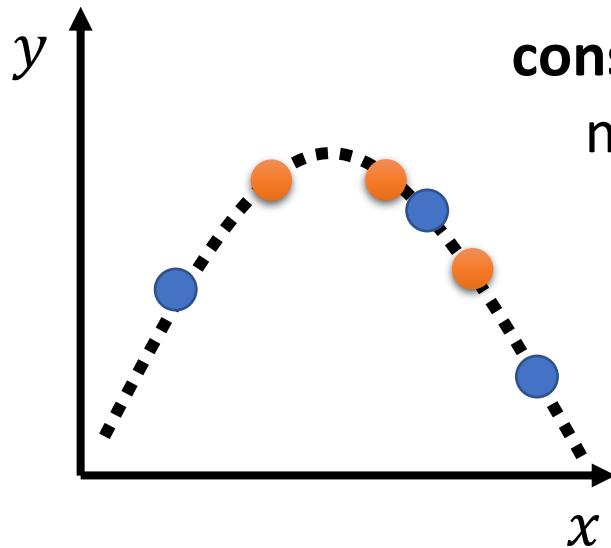
---- Real data distribution
(not observable)

● Training data

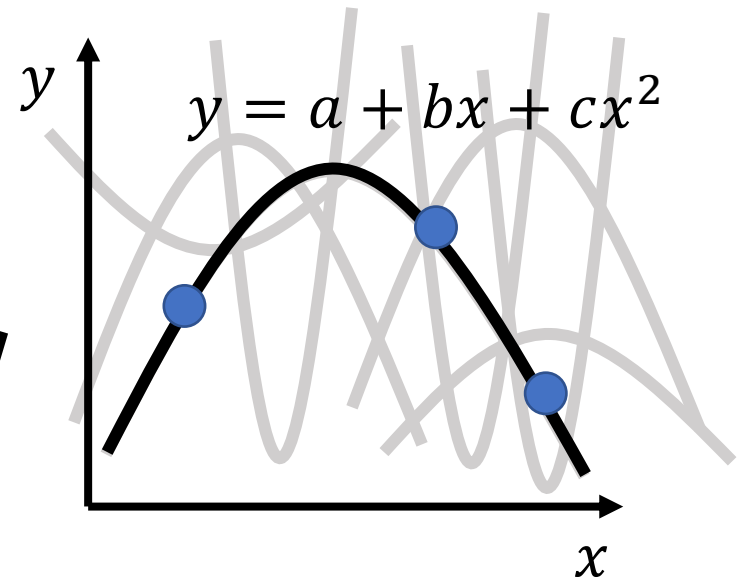
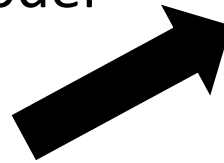
● Testing data



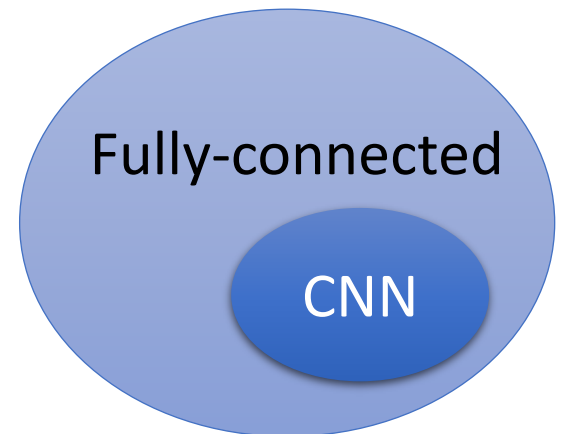
Overfitting



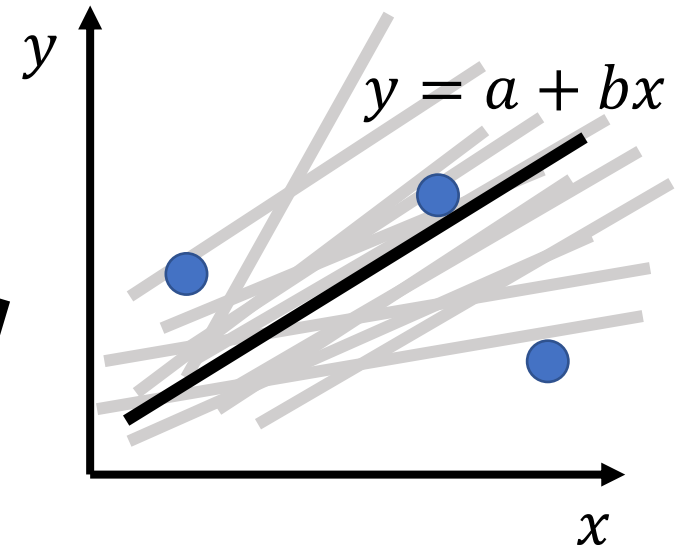
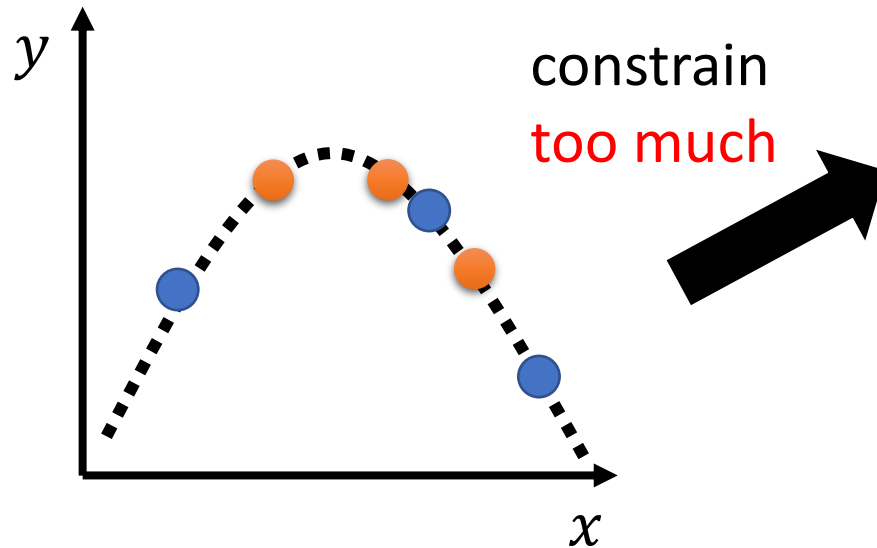
**constrained
model**



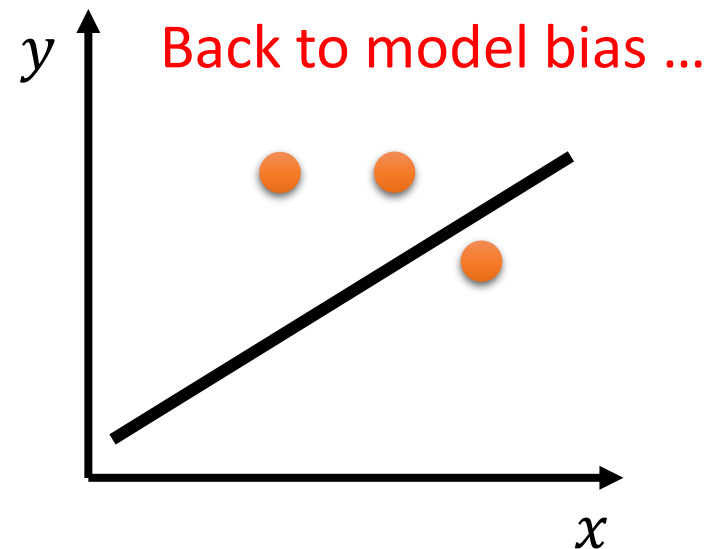
- Less parameters, sharing parameters
- Less features
- Early stopping
- Regularization
- Dropout



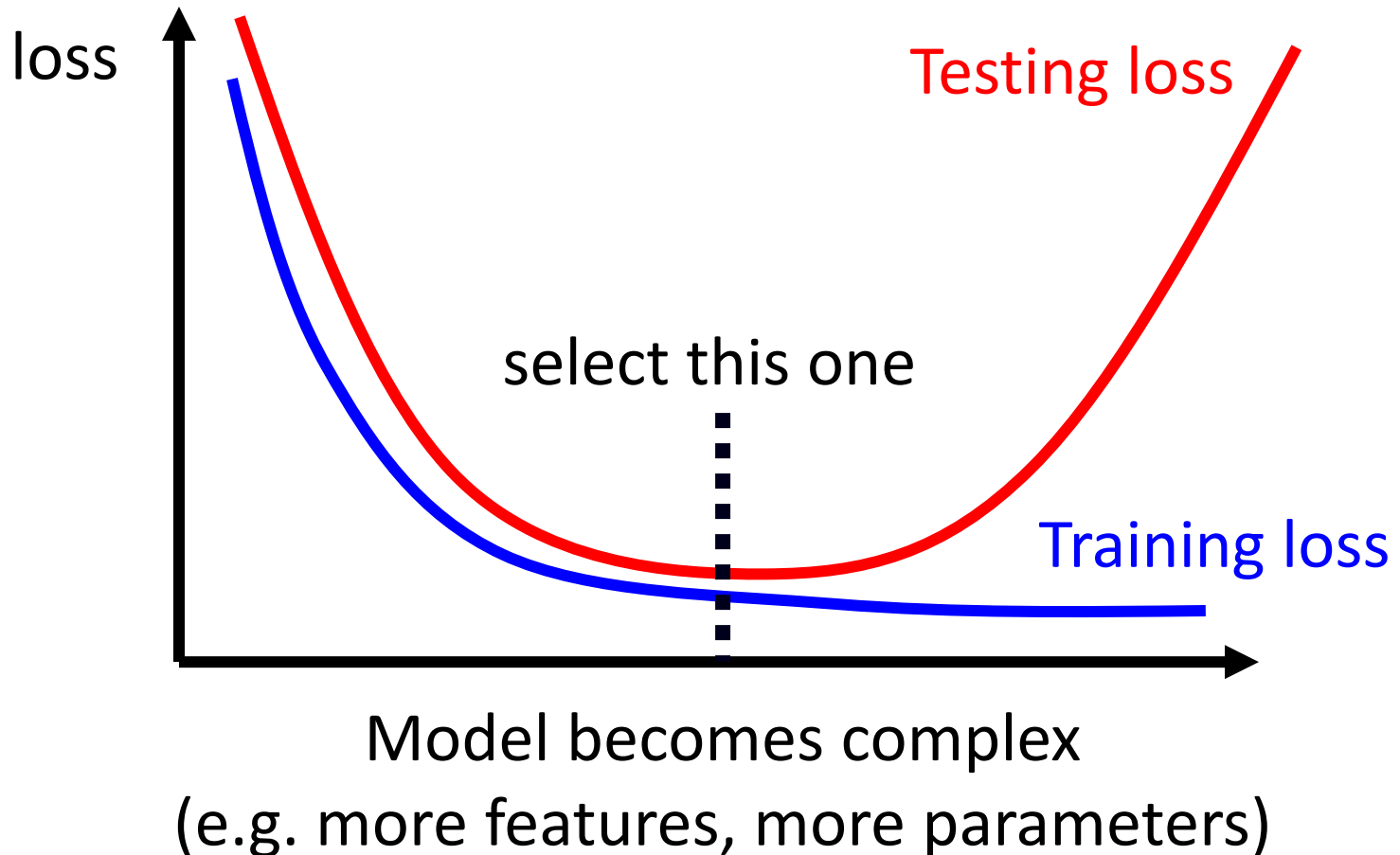
Overfitting



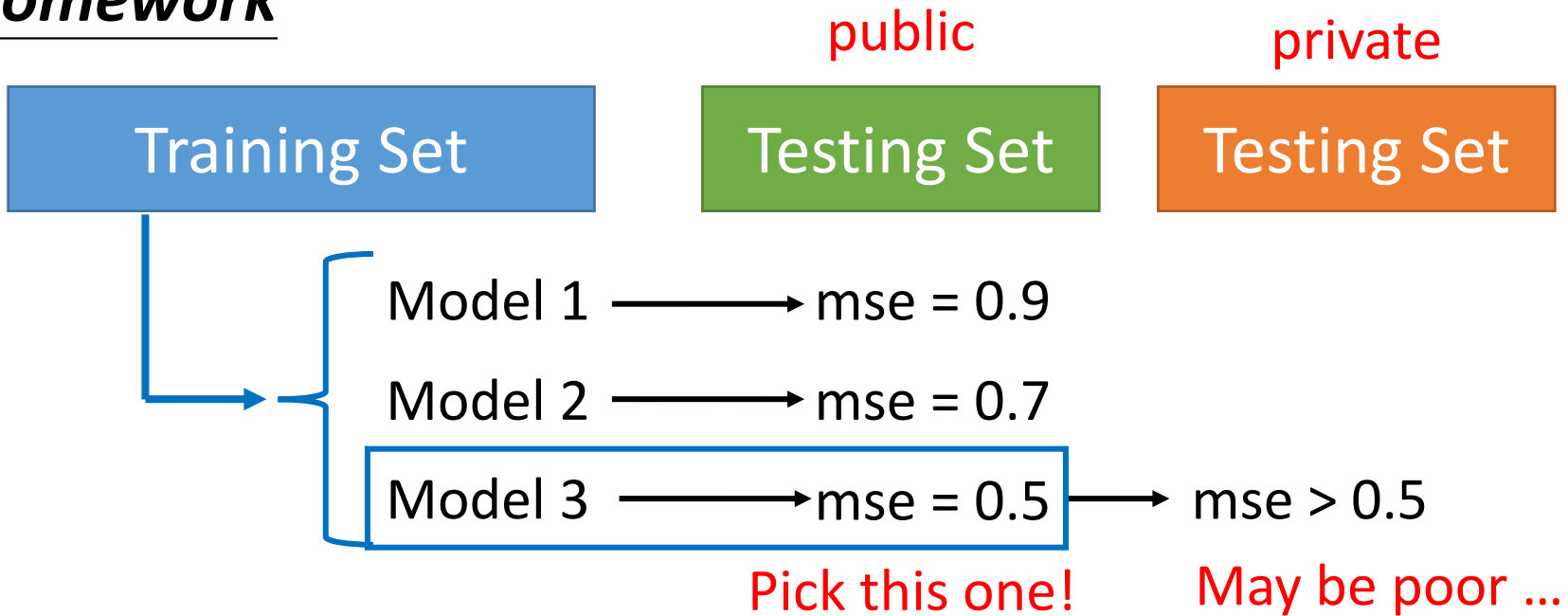
- Real data distribution (not observable)
- Training data
- Testing data



Bias-Complexity Trade-off



Homework



The extreme example again

$$f_{\textcolor{red}{k}}(\mathbf{x}) = \begin{cases} \hat{y}^i & \exists \mathbf{x}^i = \mathbf{x} \\ random & otherwise \end{cases} \quad \textcolor{red}{k}: 1 - 1000000000000000000000000$$

It is possible that $f_{56789}(x)$ **happens** to get good performance on public testing set.

So you select $f_{56789}(\mathbf{x}) \dots\dots$ Random on private testing set

Homework

public

private

Training Set

Testing Set

Testing Set

Why?

Model 1 \longrightarrow mse = 0.9

Model 2 \longrightarrow mse = 0.7

Model 3 \longrightarrow mse = 0.5 \longrightarrow mse > 0.5

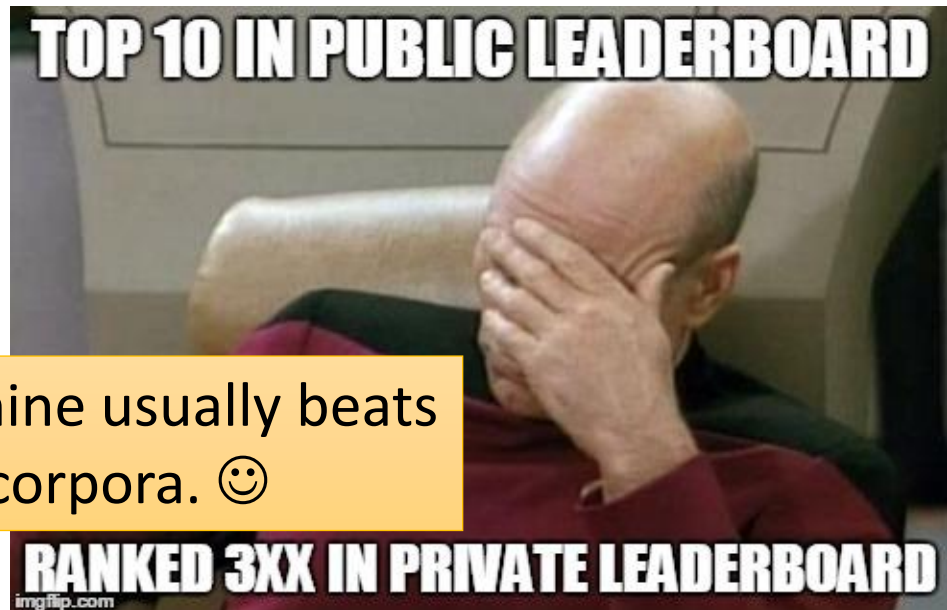
Pick this one!

May be poor ...

What will happen?

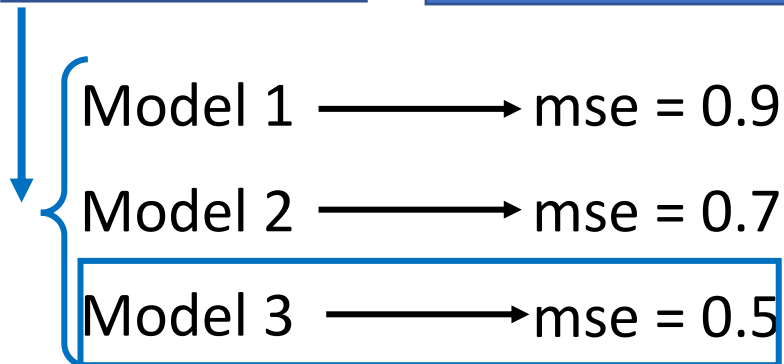
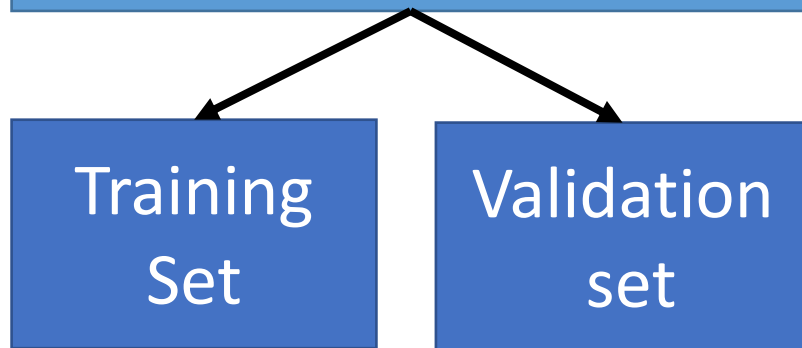
<http://www.chioka.in/how-to-select-your-final-models-in-a-kaggle-competitio/>

This explains why machine usually beats human on benchmark corpora. 😊



Cross Validation

How to split?



public

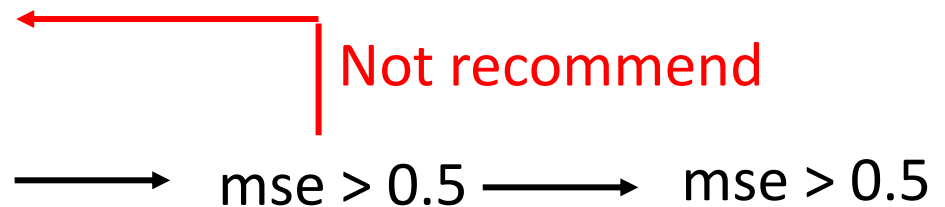


private

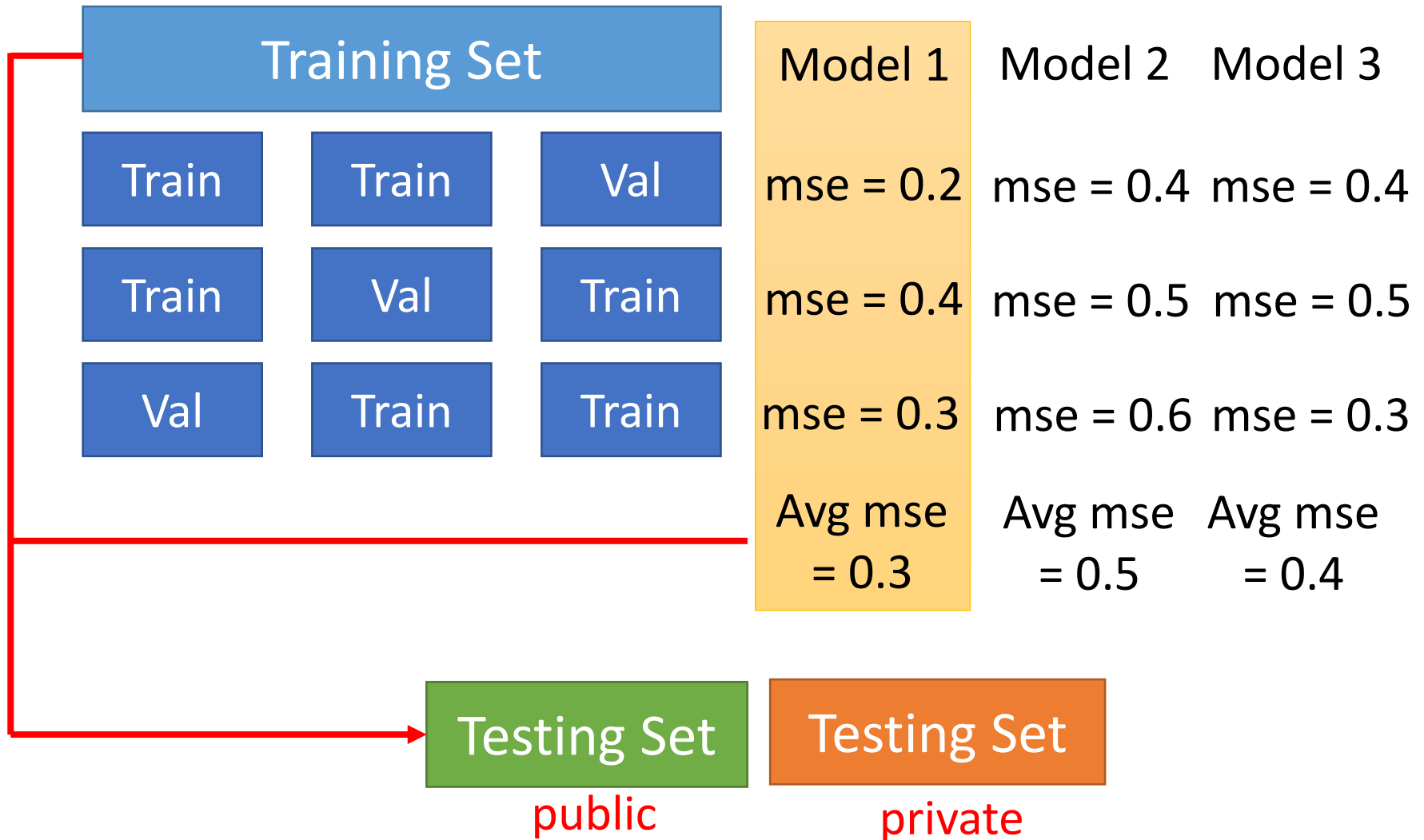


Using the results of public testing data to select your model
You are making public set better than private set.

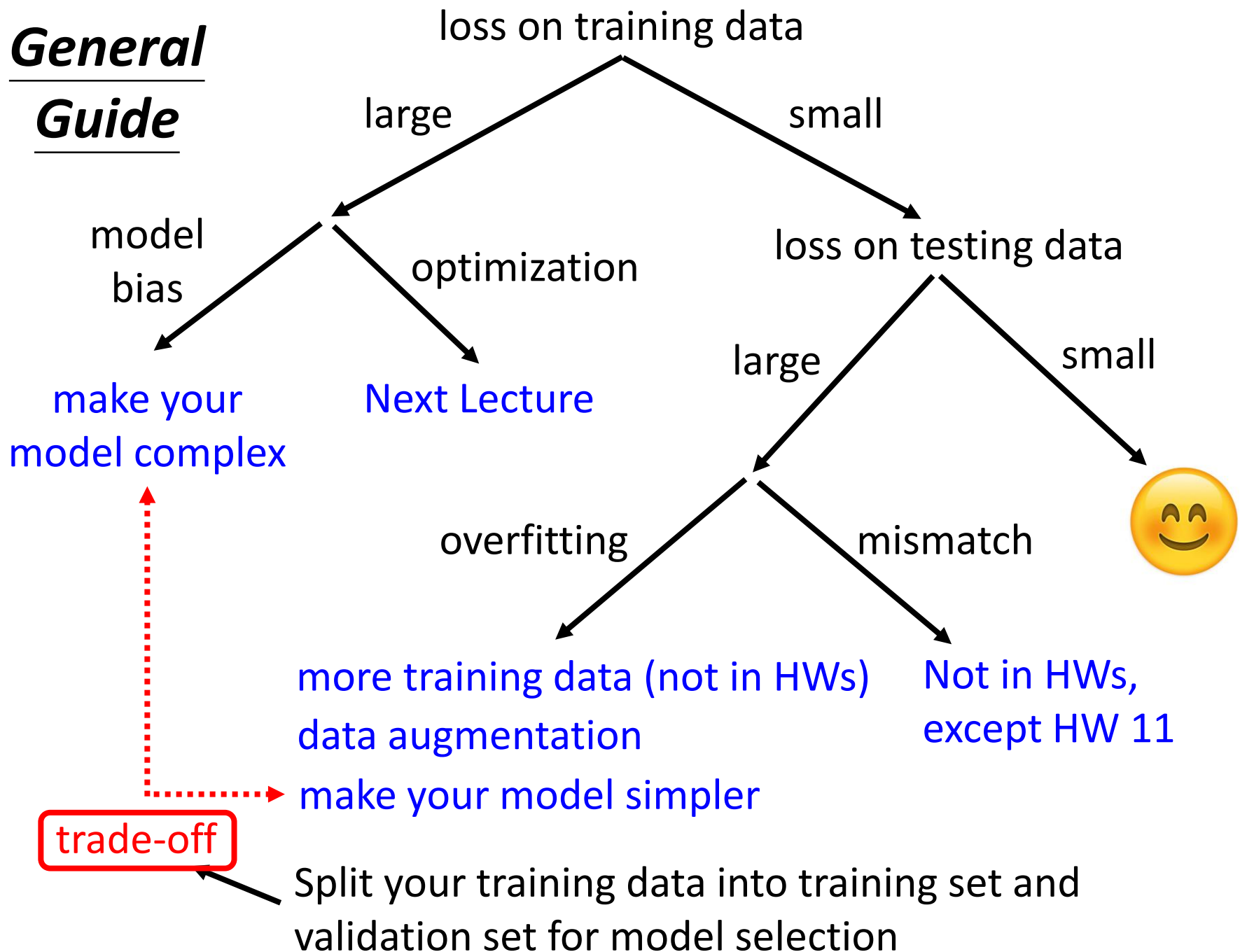
Not recommend



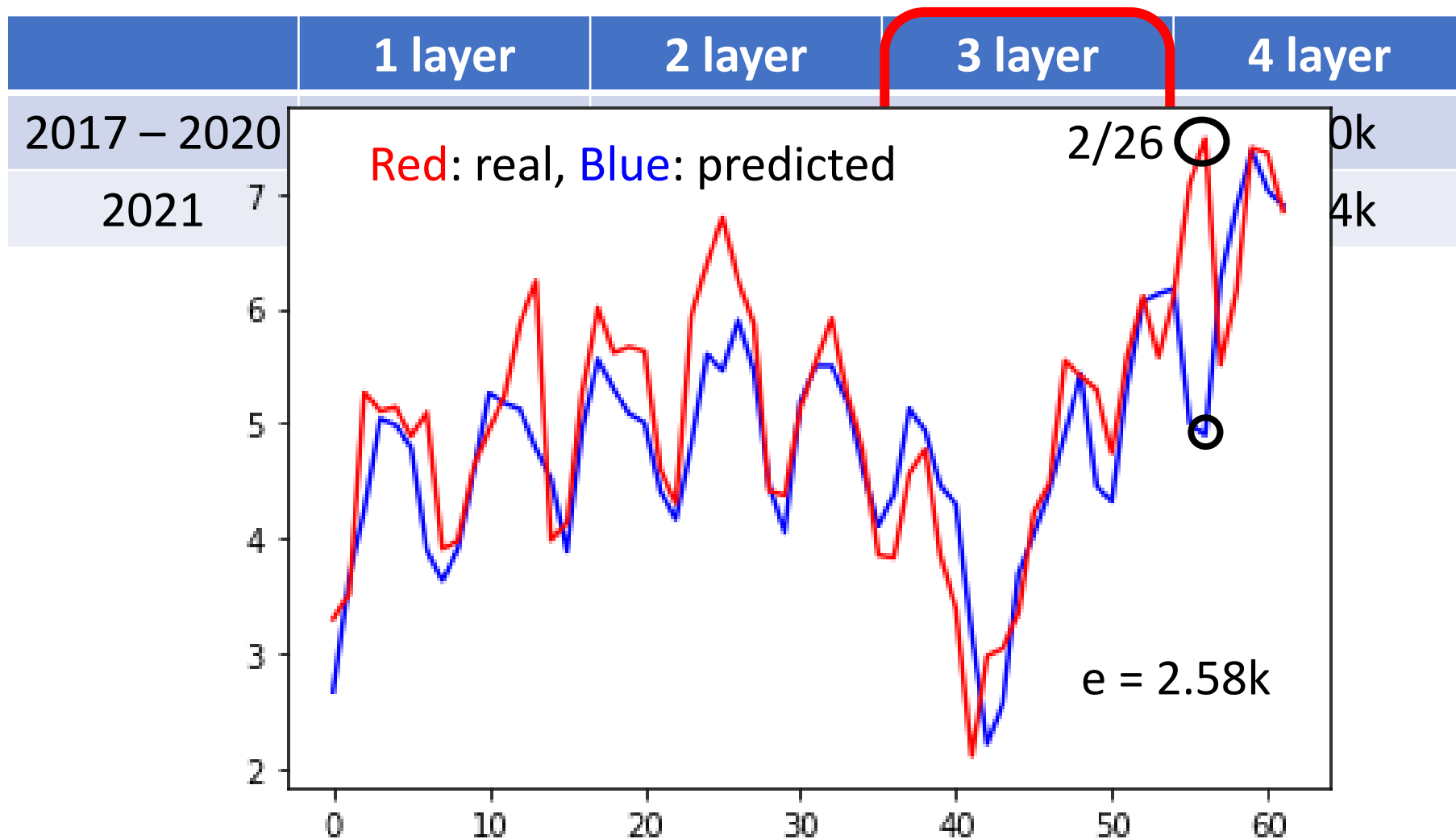
N-fold Cross Validation



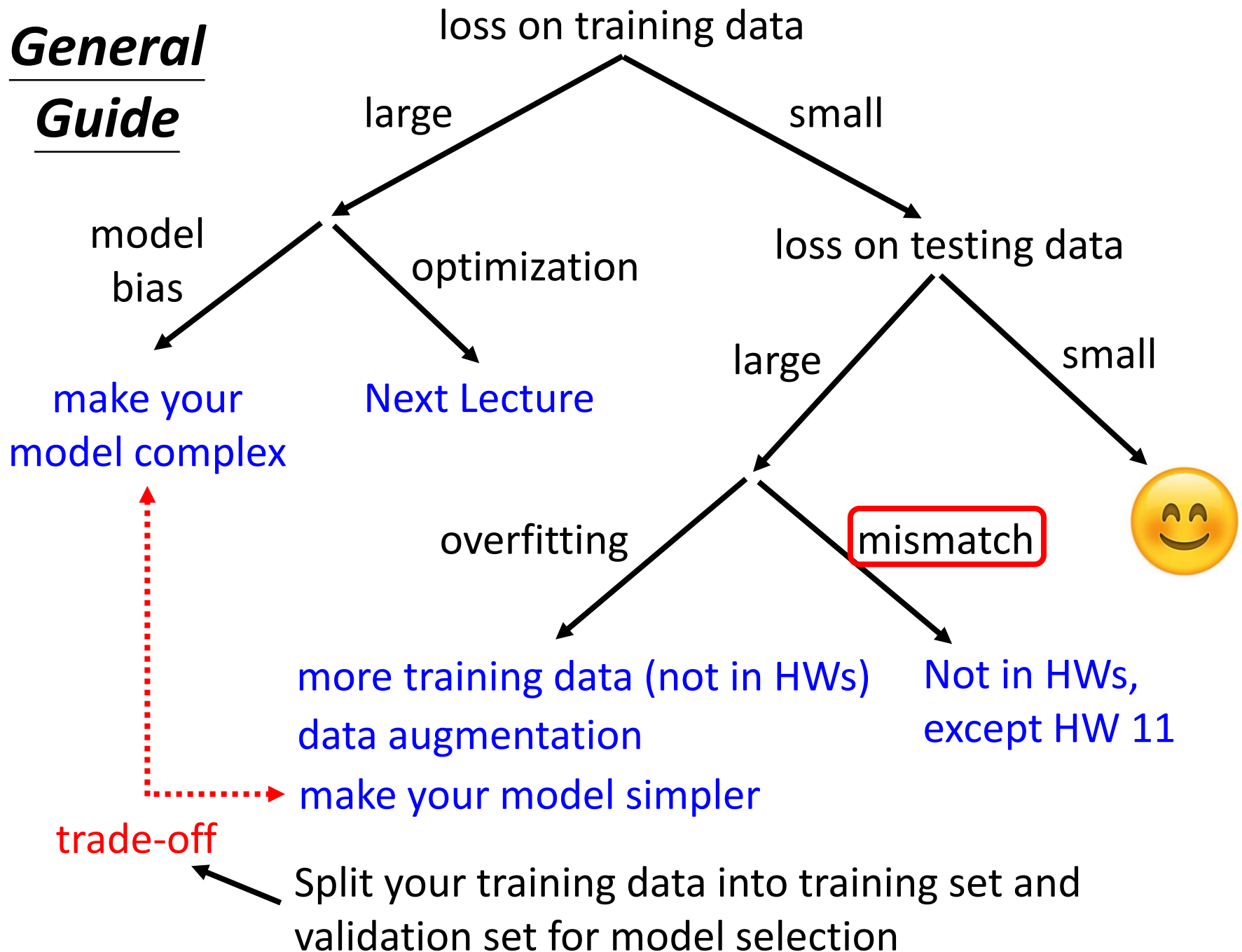
General Guide



Let's predict no. of views of 2/26!



General Guide



Mismatch

- Your training and testing data have different distributions. Be aware of how data is generated.

Most HWs do not have this problem, except HW11

Training Data

horse



bed



clock



apple



cat



plane



television



dog



dolphin



spider

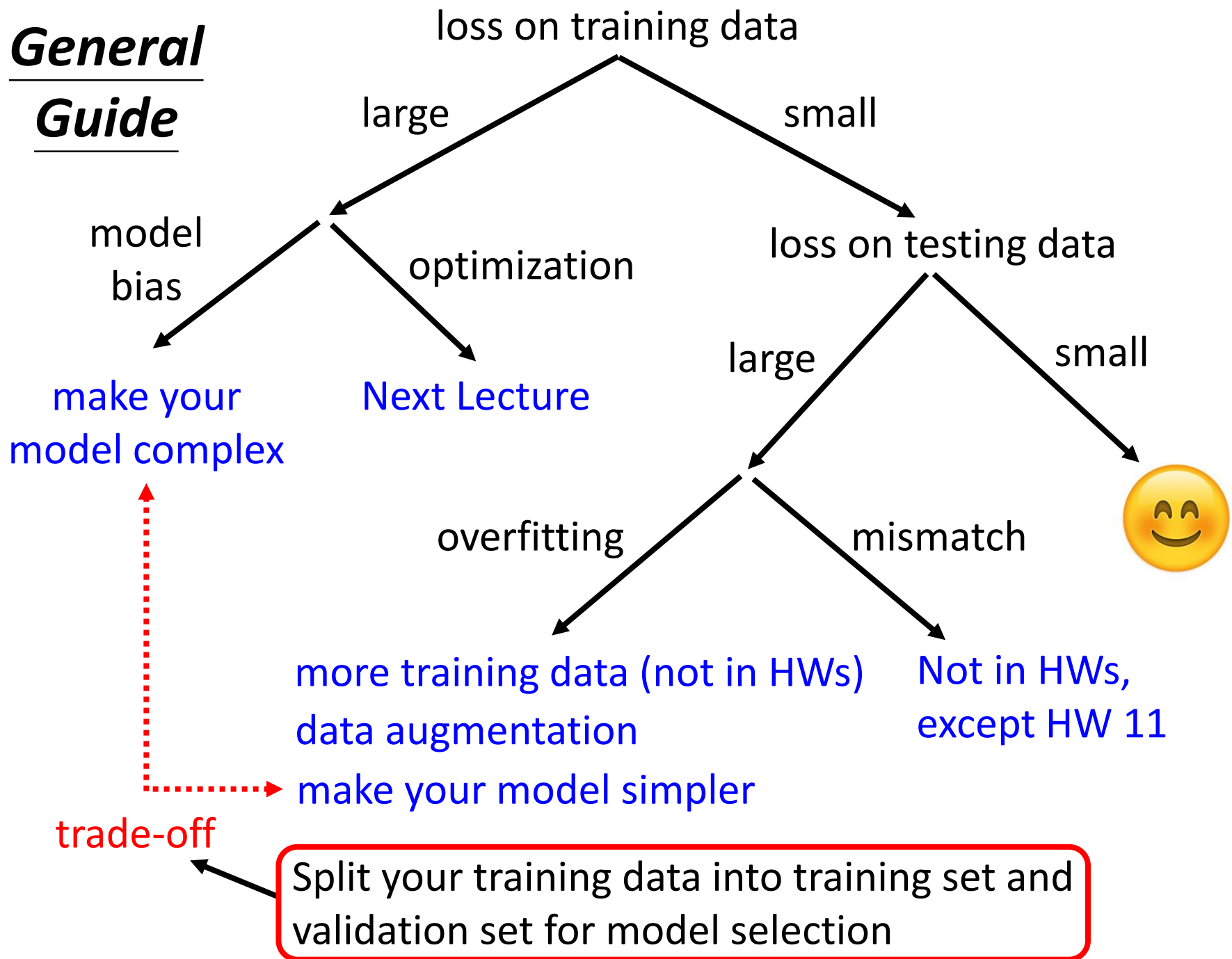


Simply increasing the training data will not help.

Testing Data



General Guide



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

HYLEE(2021) · 课程资料包 @ShowMeAI



视频

课件

笔记

代码

中英双语字幕

一键打包下载

官方笔记翻译

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]