作业通关大攻略

General Guidance

Hung-yi Lee 李宏毅

Framework of ML

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

Testing data: $\{\boldsymbol{x}^{N+1}, \boldsymbol{x}^{N+2}, \ldots, \boldsymbol{x}^{N+M}\}$

本课：
ŷ：真实值
y：预测值

Speech Recognition 作业2

$\boldsymbol{x}$: 

$\hat{y}$: phoneme

Image Recognition 作业3

$\boldsymbol{x}$: 

$\hat{y}$: soup

Speaker Recognition 作业4

$\boldsymbol{x}$: 

$\hat{y}$: John (speaker)

Machine Translation 作业5

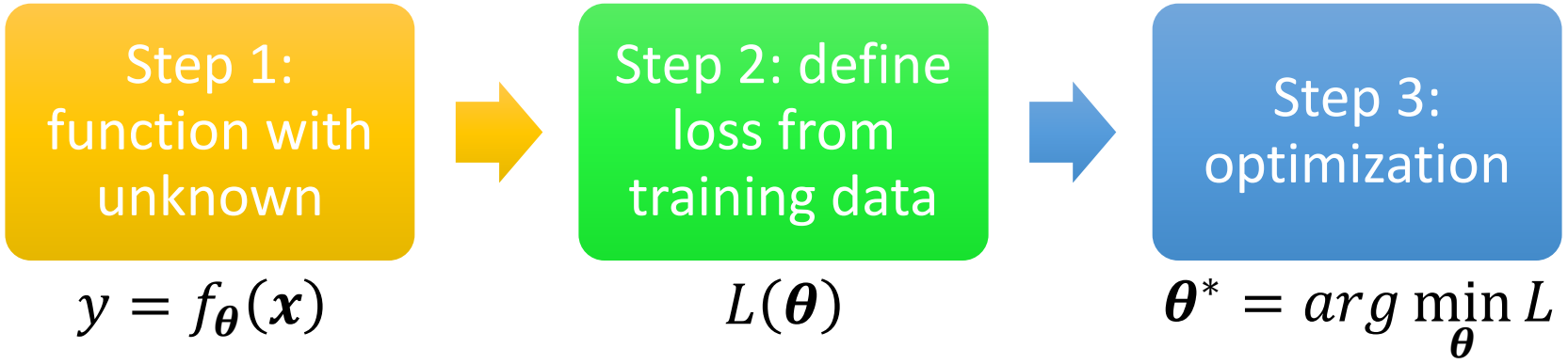
$\boldsymbol{x}$: 痛みを知れ

$\hat{y}$: 了解痛苦吧

Framework of ML

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

Training:

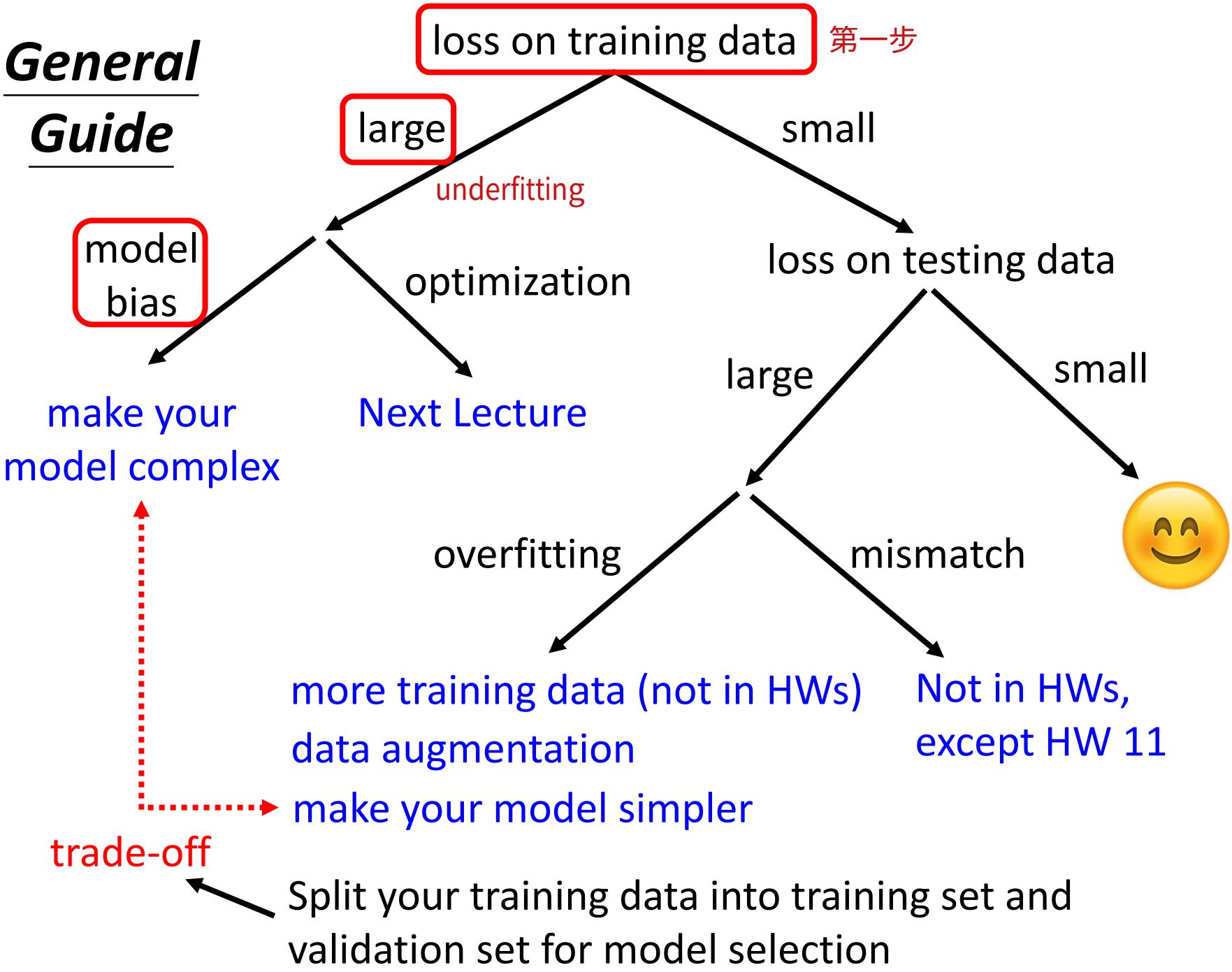


Testing data: $\{\boldsymbol{x^{N+1}}, \boldsymbol{x^{N+2}}, \ldots, \boldsymbol{x^{N+M}}\}$

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

$\{y^{N+1}, y^{N+2}, \ldots, y^{N+M}\}$ → Upload to Kaggle

General Guide



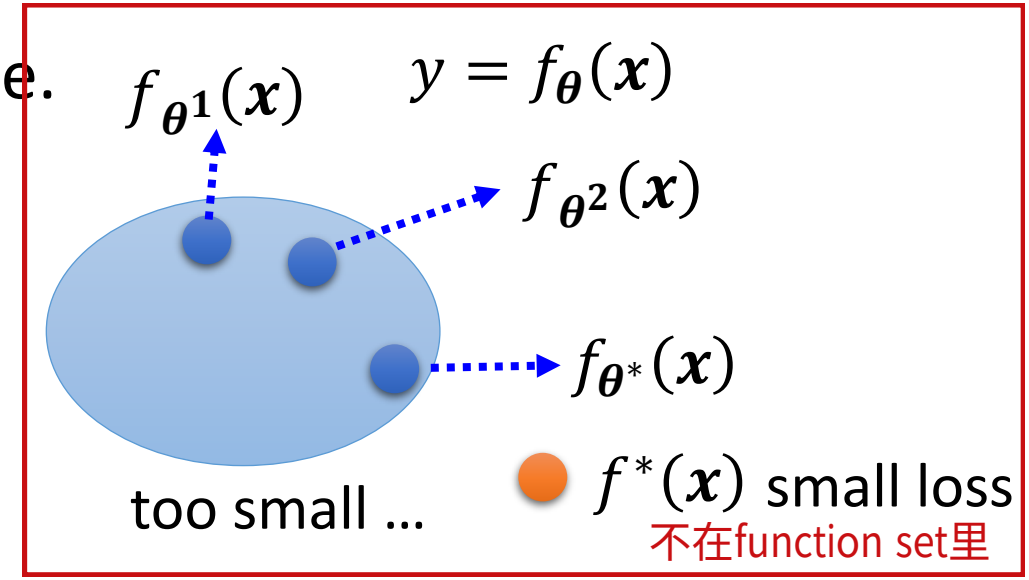
Model Bias

underfitting的第一种可能：model bias

- The model is too simple.

find a needle in a haystack …

… but there is no needle



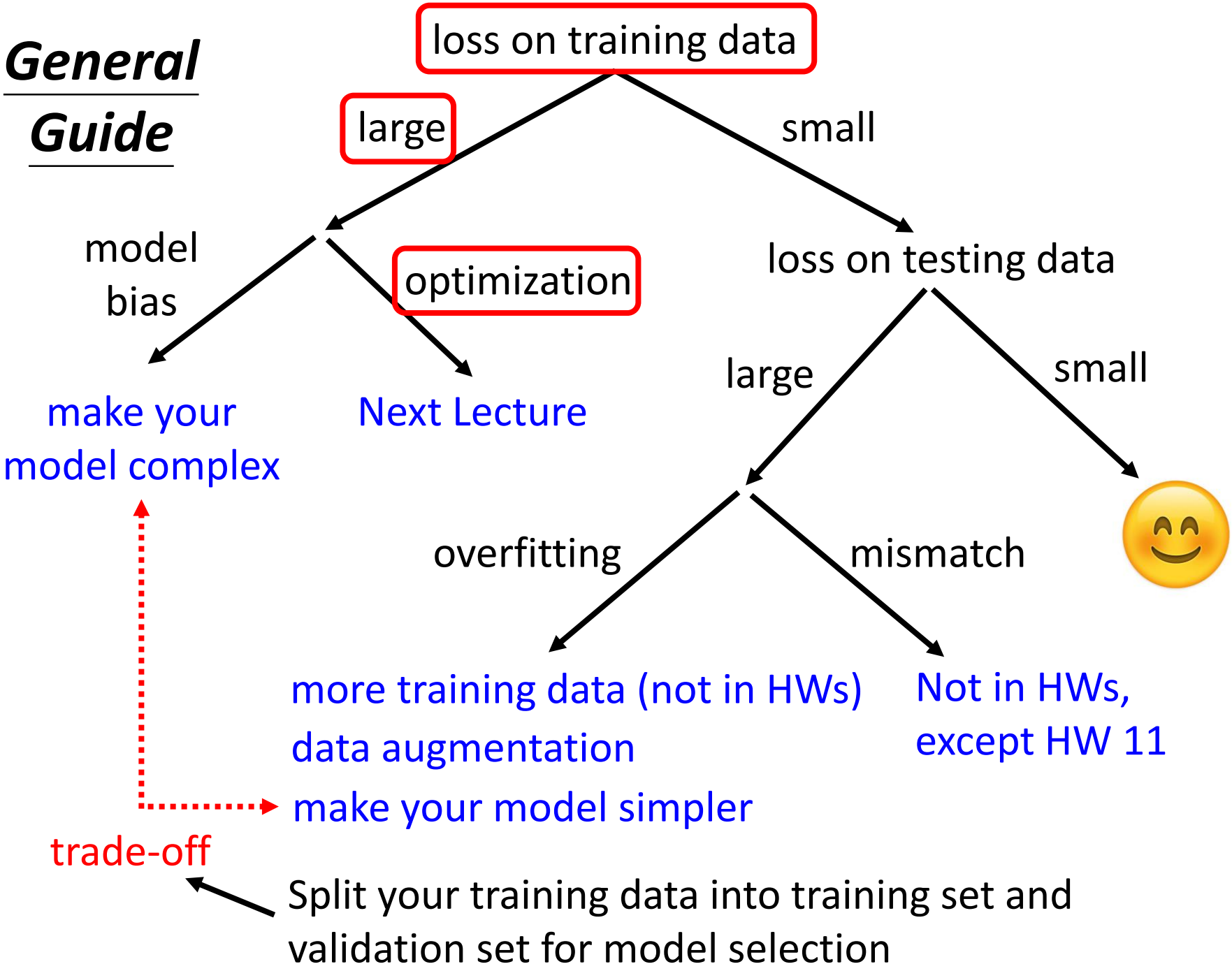
- Solution: redesign your model to make it more flexible

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

Deep Learning
(more neurons, layers)

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

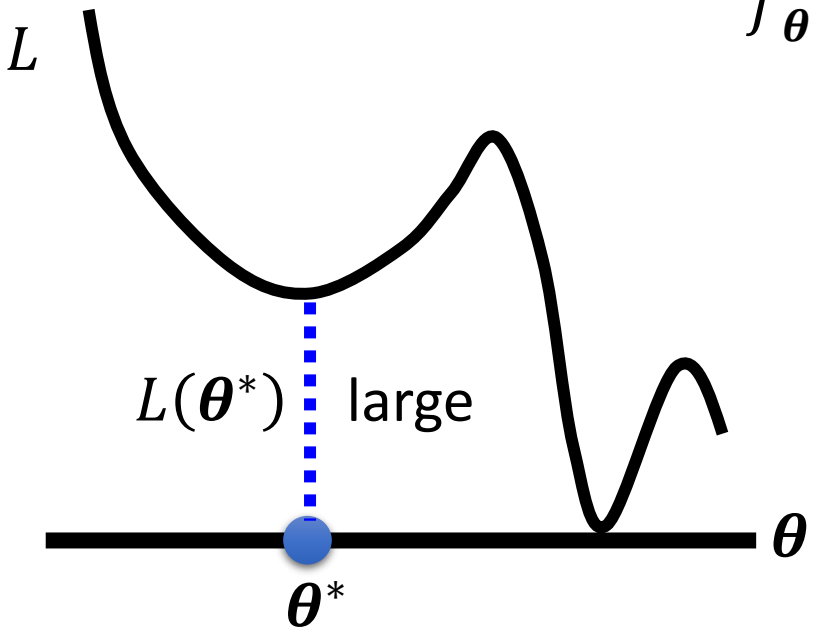
General Guide

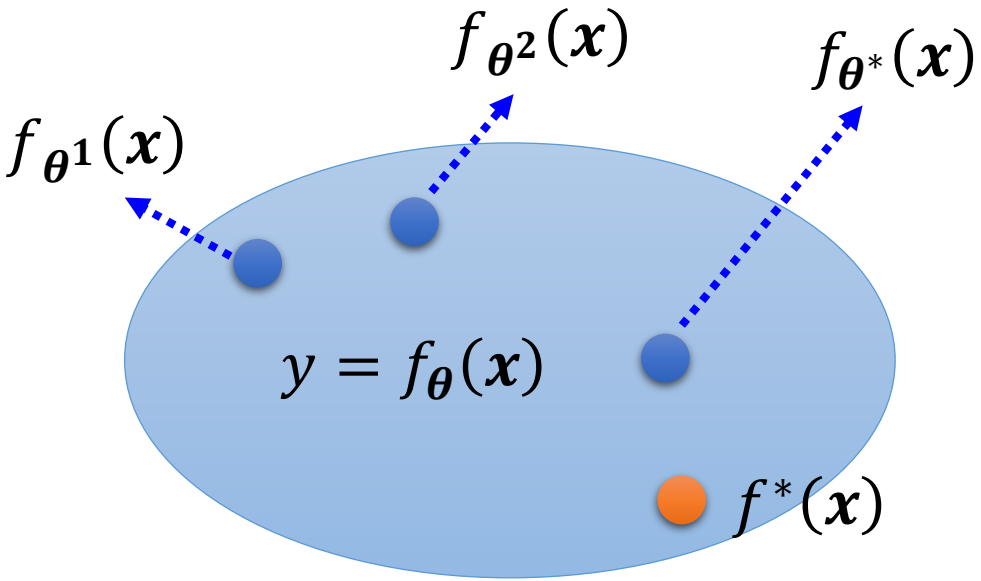


underfitting的第二种可能：
optimization做得不好

Optimization Issue

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





A needle is in a haystack …

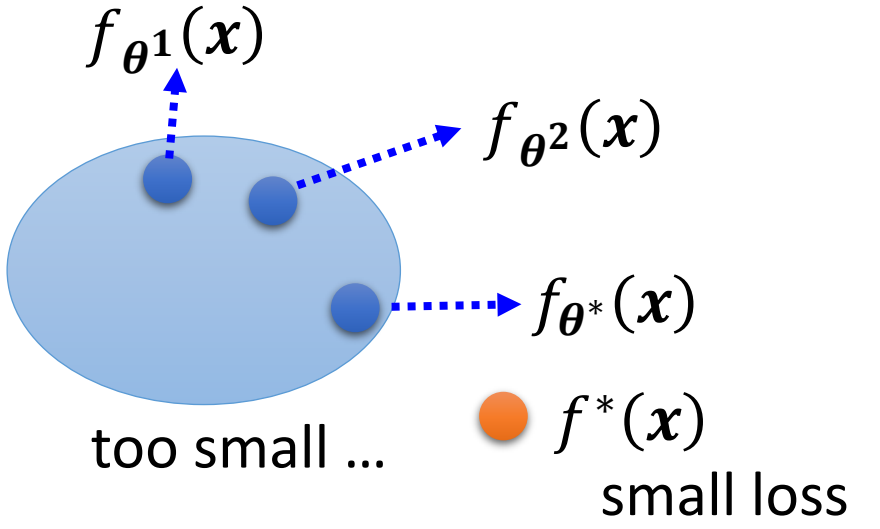
… Just cannot find it.

大海捞针，针确实在海里，但是就是找不到

Model Bias

find a needle in a haystack …

… but there is no needle

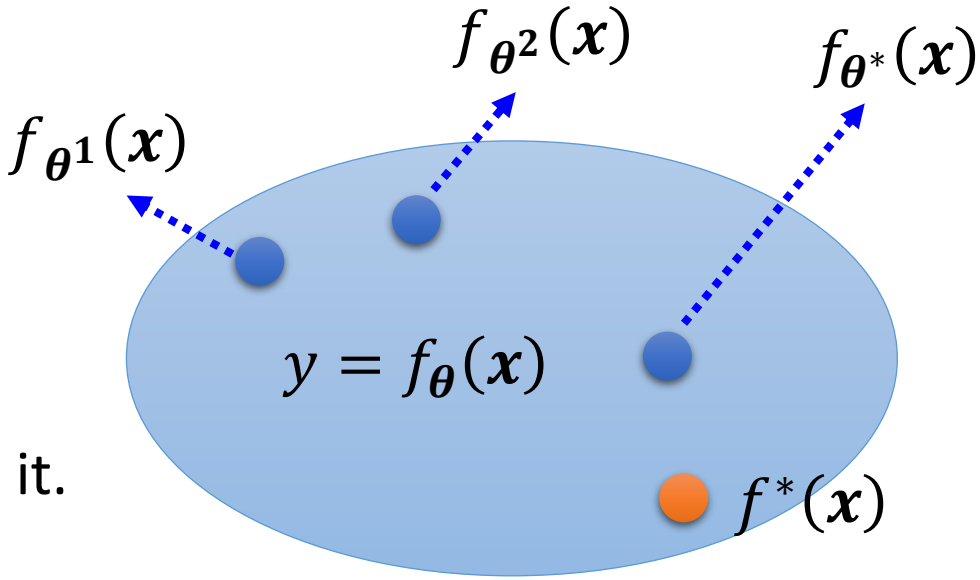


Which one???

Optimization Issue

A needle is in a haystack …

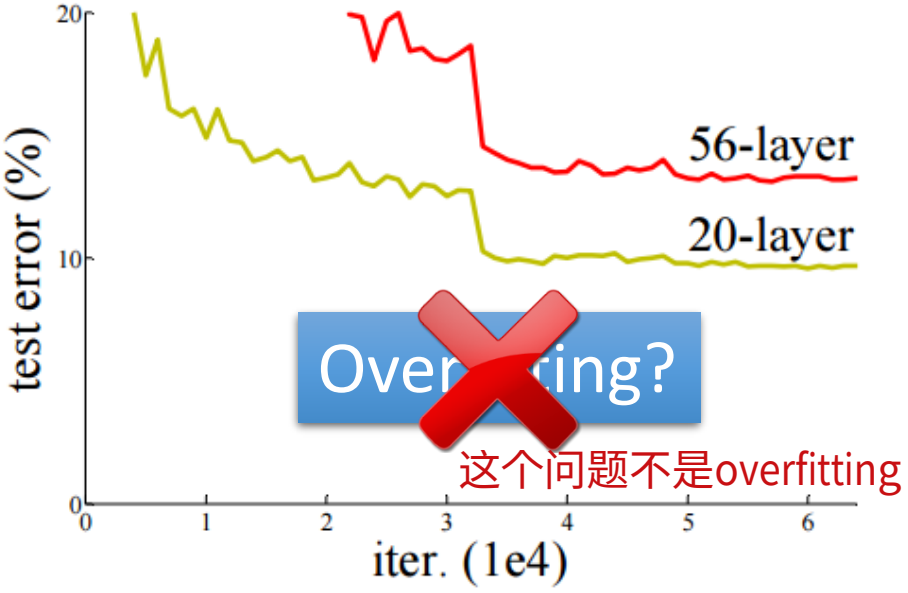
… Just cannot find it.



Ref: http://arxiv.org/abs/1512.03385

Model Bias v.s. Optimization Issue

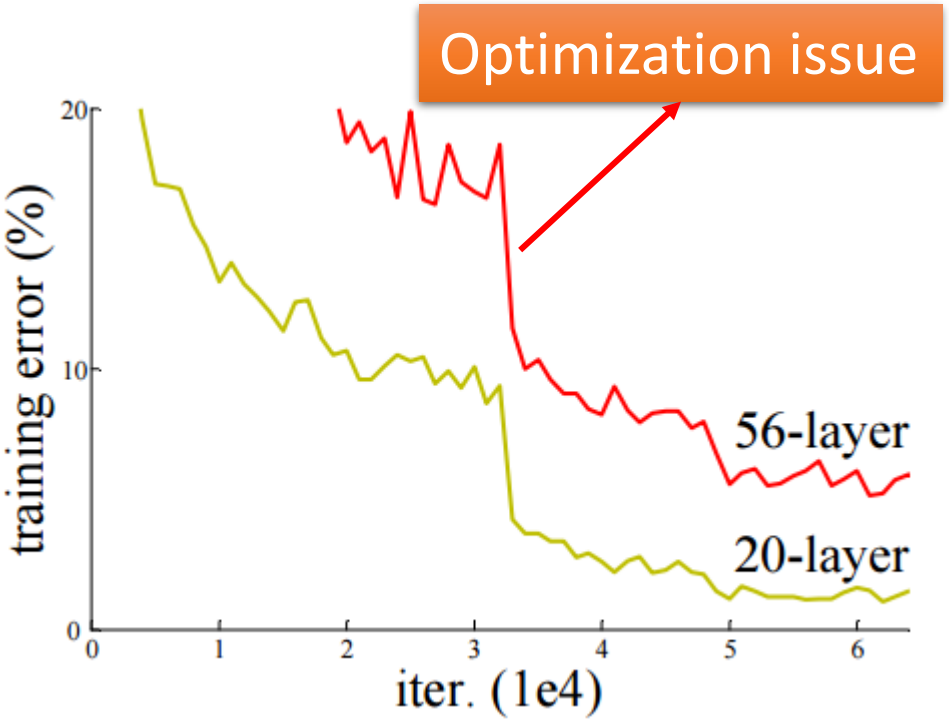
- Gaining the insights from comparison



这个问题不是overfitting

Testing Data

overfitting：training上好，testing上差



Training Data

56层的network更flexible，但是error高，说明optimization出现了问题

Ref: http://arxiv.org/abs/1512.03385

Optimization Issue

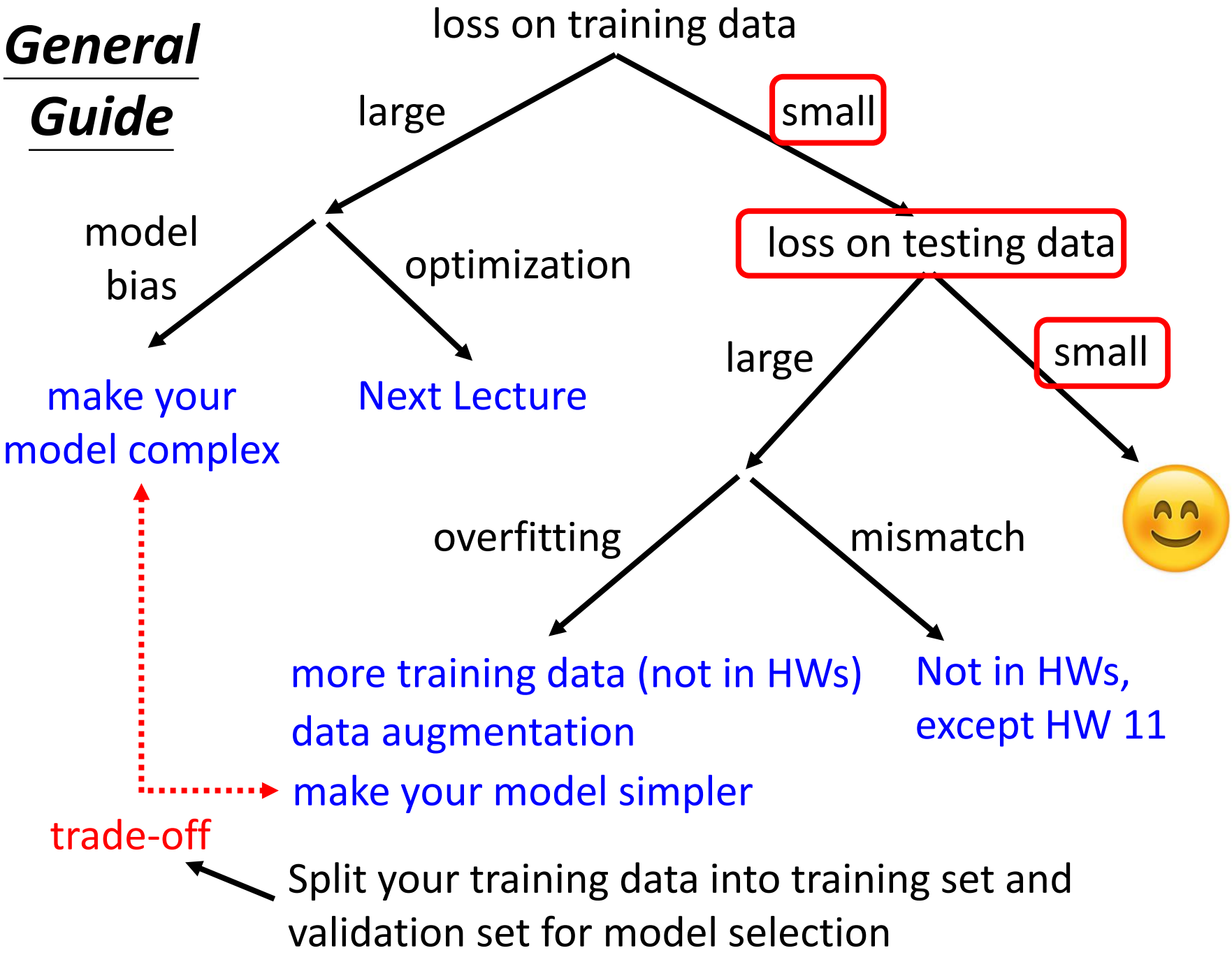
先跑比较浅的network，或者用不是deep learning的简单的模型。

- 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.

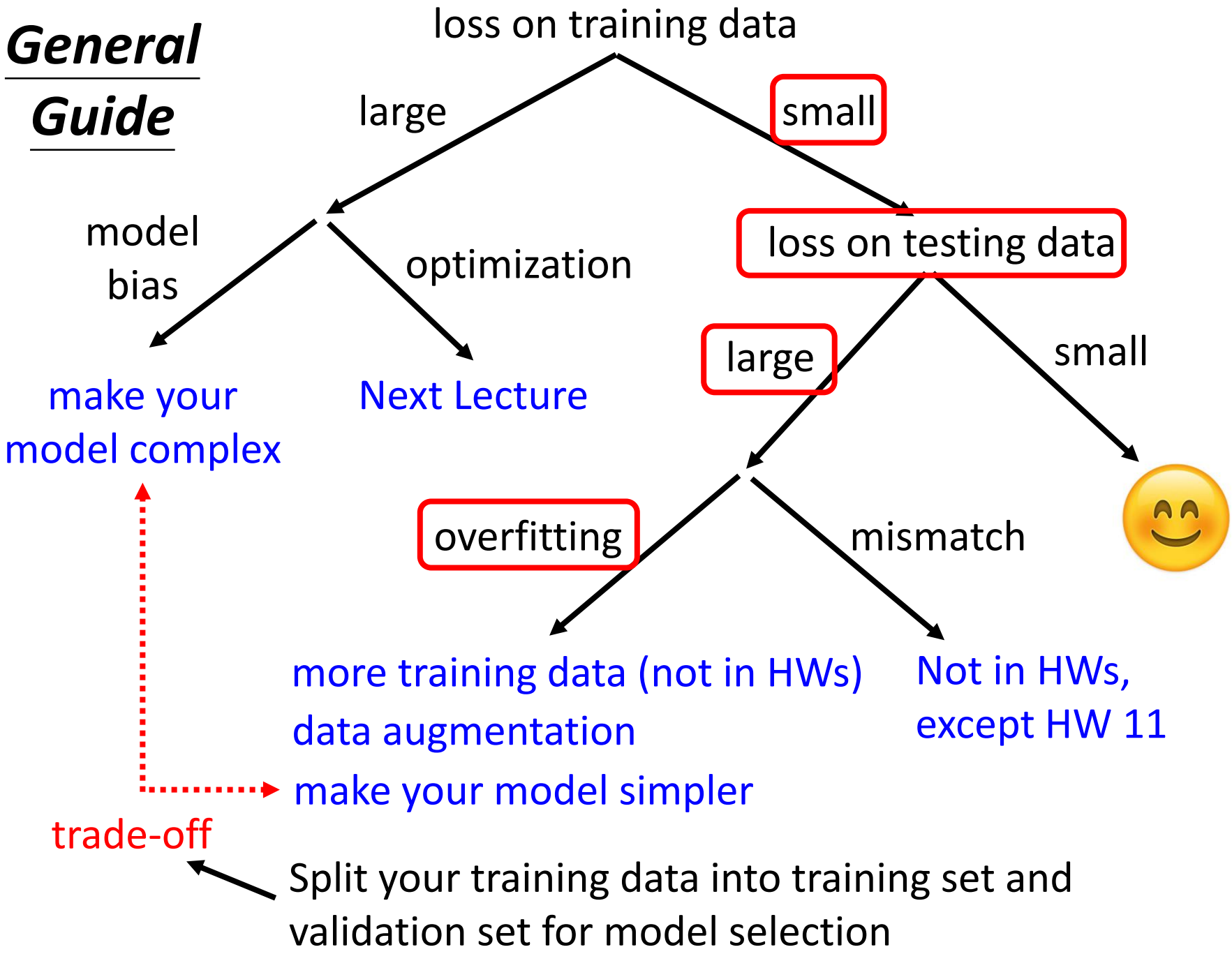
	1 layer	2 layer	3 layer	4 layer	5 layer
2017 – 2020	0.28k	0.18k	0.14k	0.10k	0.34k

- 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

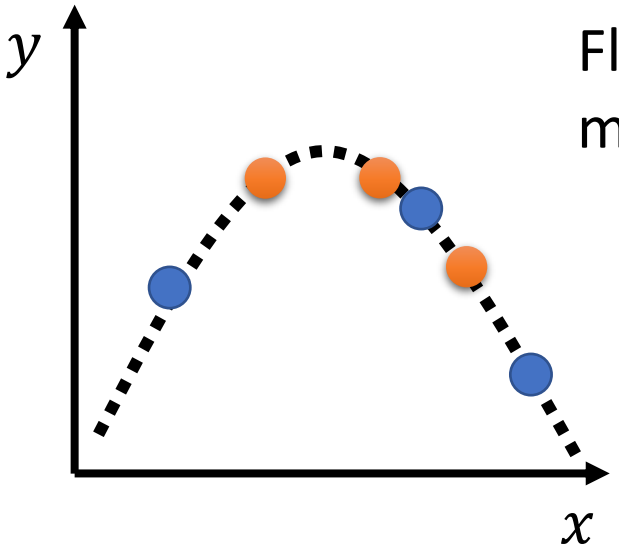
Training data: $\{(\boldsymbol{x^1}, \hat{y}^1), (\boldsymbol{x^2}, \hat{y}^2), \ldots, (\boldsymbol{x^N}, \hat{y}^N)\}$

$$f(\boldsymbol{x}) = \begin{cases} \hat{y}^i & \exists \boldsymbol{x^i} = \boldsymbol{x} \\ random & otherwise \end{cases}$$

Less than useless …

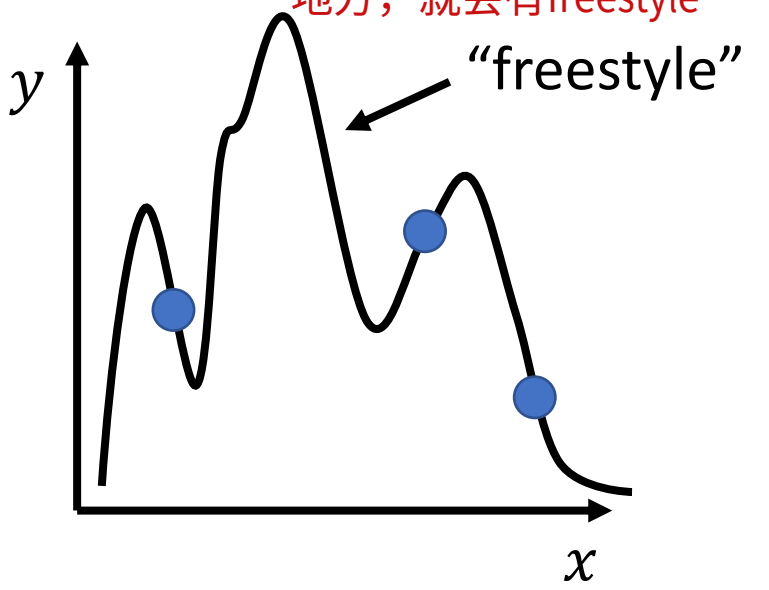
This function obtains **zero training loss**, but **large testing loss**.

Overfitting



Flexible model

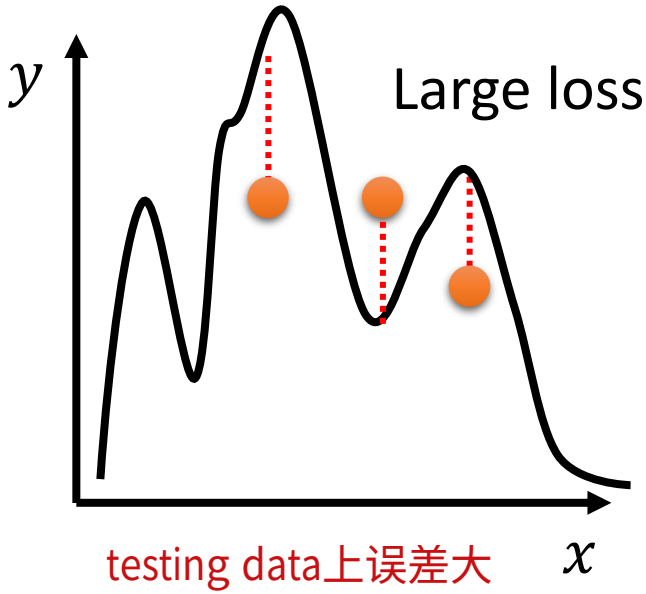
在没有training data限制的地方，就会有freestyle



Real data distribution (not observable)

Training data

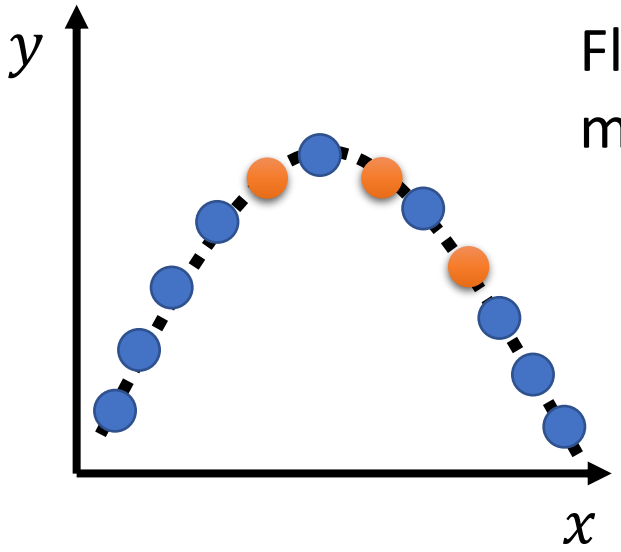
Testing data



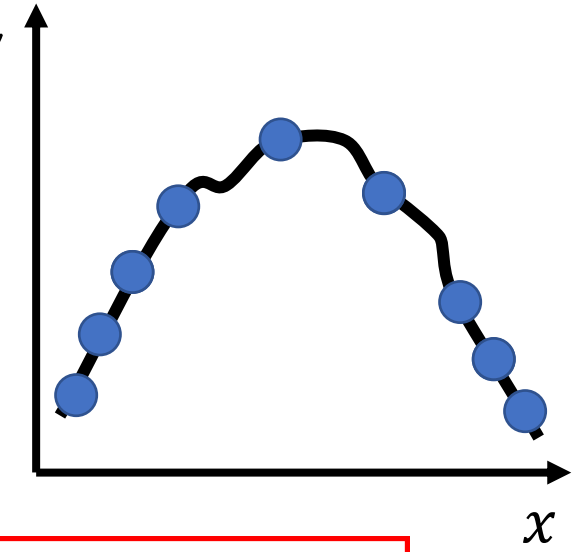
testing data上误差大

如果坚持想用flexible model的话，
只能在数据上做文章了。
1. 更多数据
2. 数据增强

Overfitting



Flexible model



1

More training data

(cannot do it in HWs)

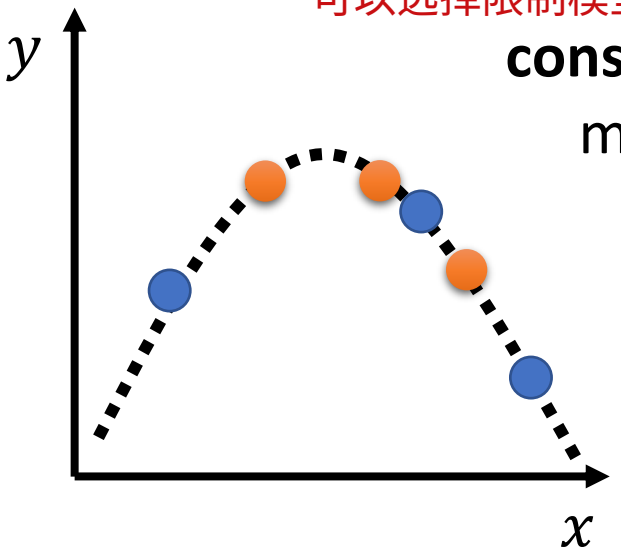
2 数据增强

Data augmentation (you can do that in HWs)




data augmentation要合适合理

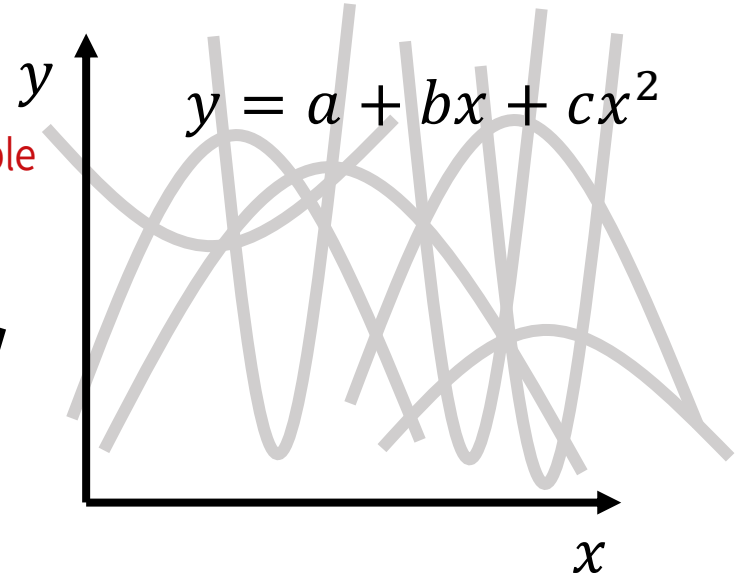
Overfitting



可以选择限制模型，降低flexible

constrained

model

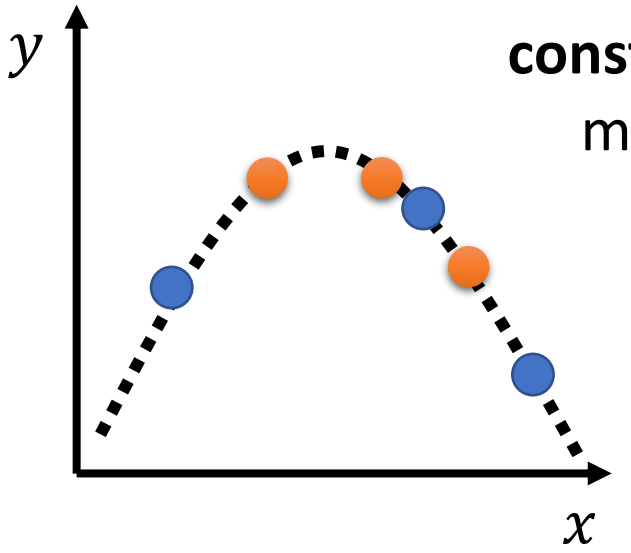


▪▪▪▪ Real data distribution (not observable)

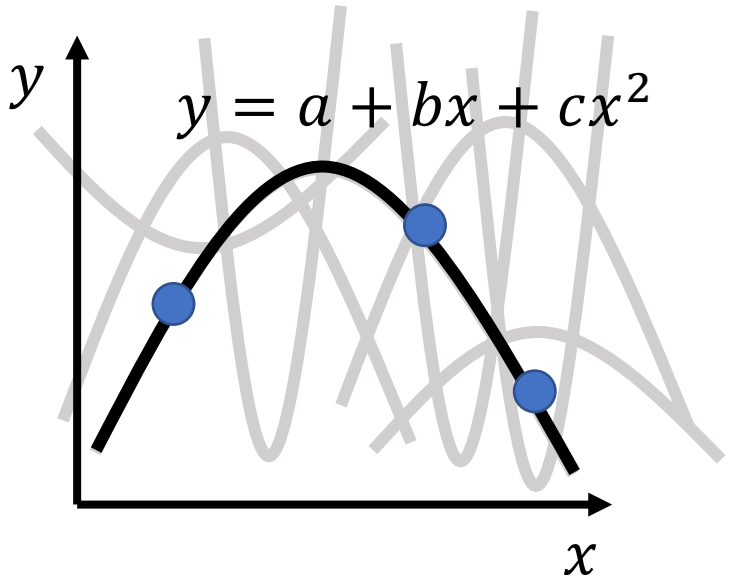
● Training data

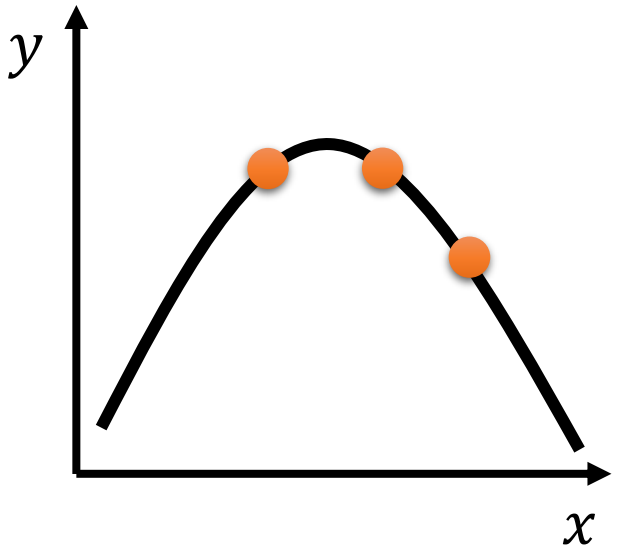
● Testing data

Overfitting



constrained model



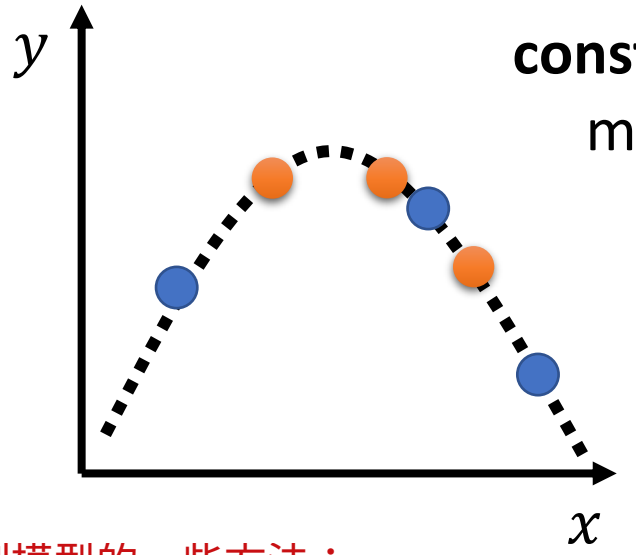


Real data distribution (not observable)

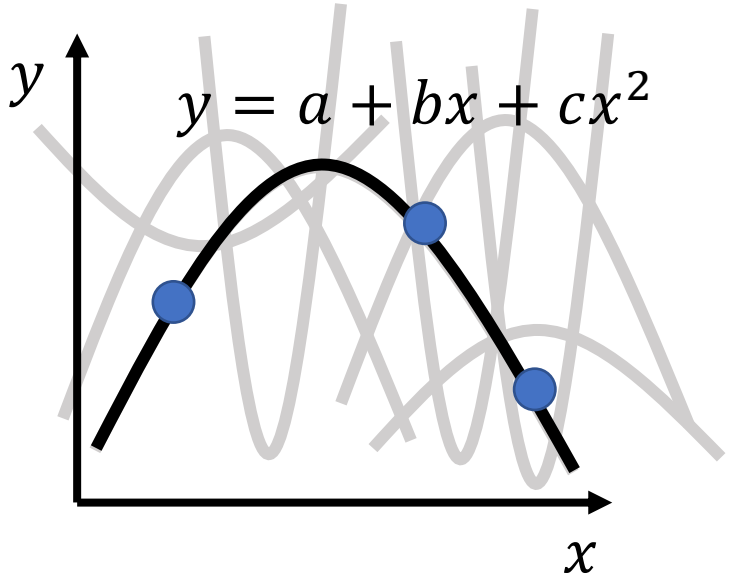
Training data

Testing data

Overfitting

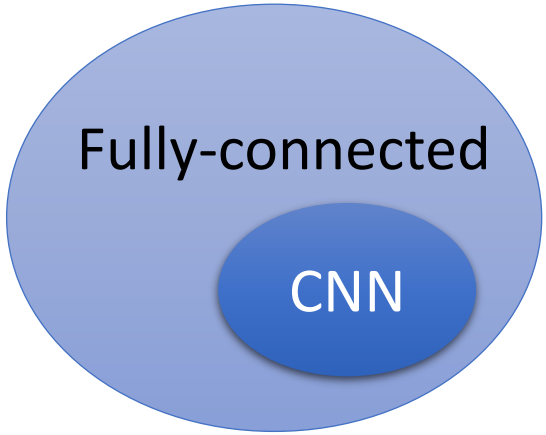


constrained
model



限制模型的一些方法：

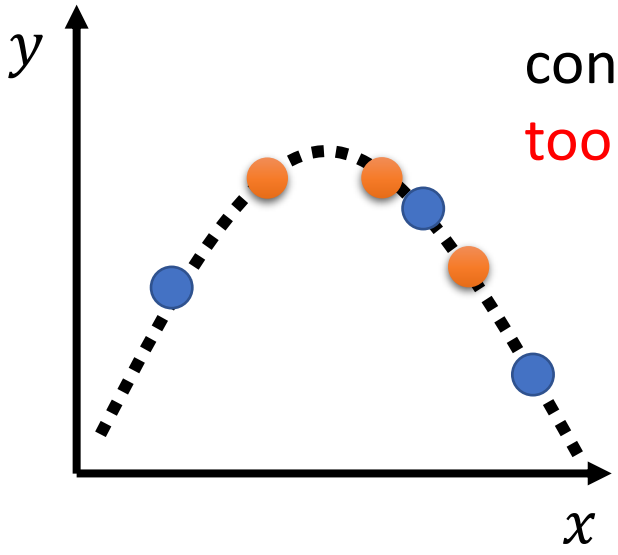
- Less parameters, sharing parameters 共享参数
- Less features eg. 降维
- Early stopping
- Regularization
- Dropout



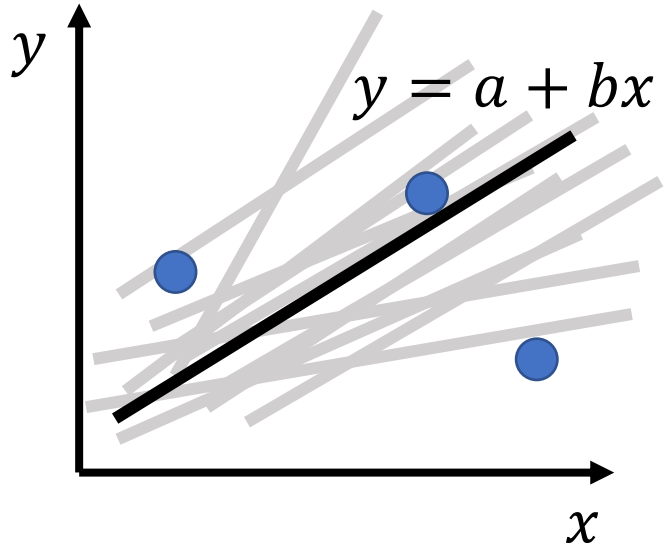
CNN：针对图像的特性，来限制模型的弹性

如果模型的限制太大

Overfitting



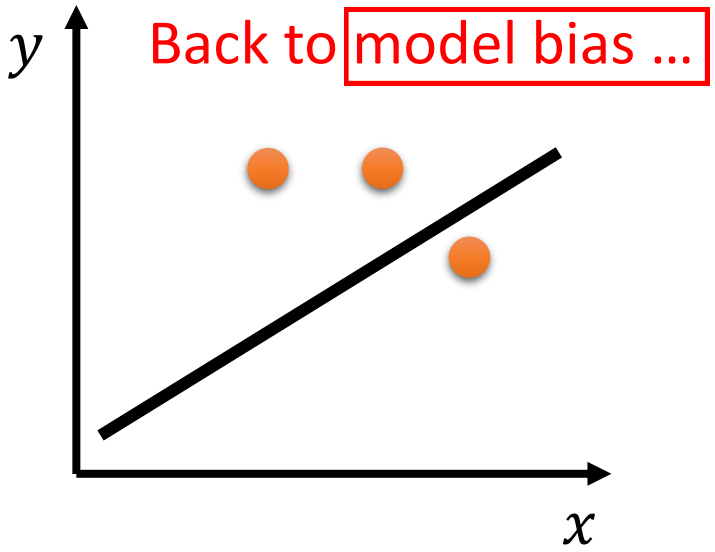
constrain too much



Real data distribution (not observable)

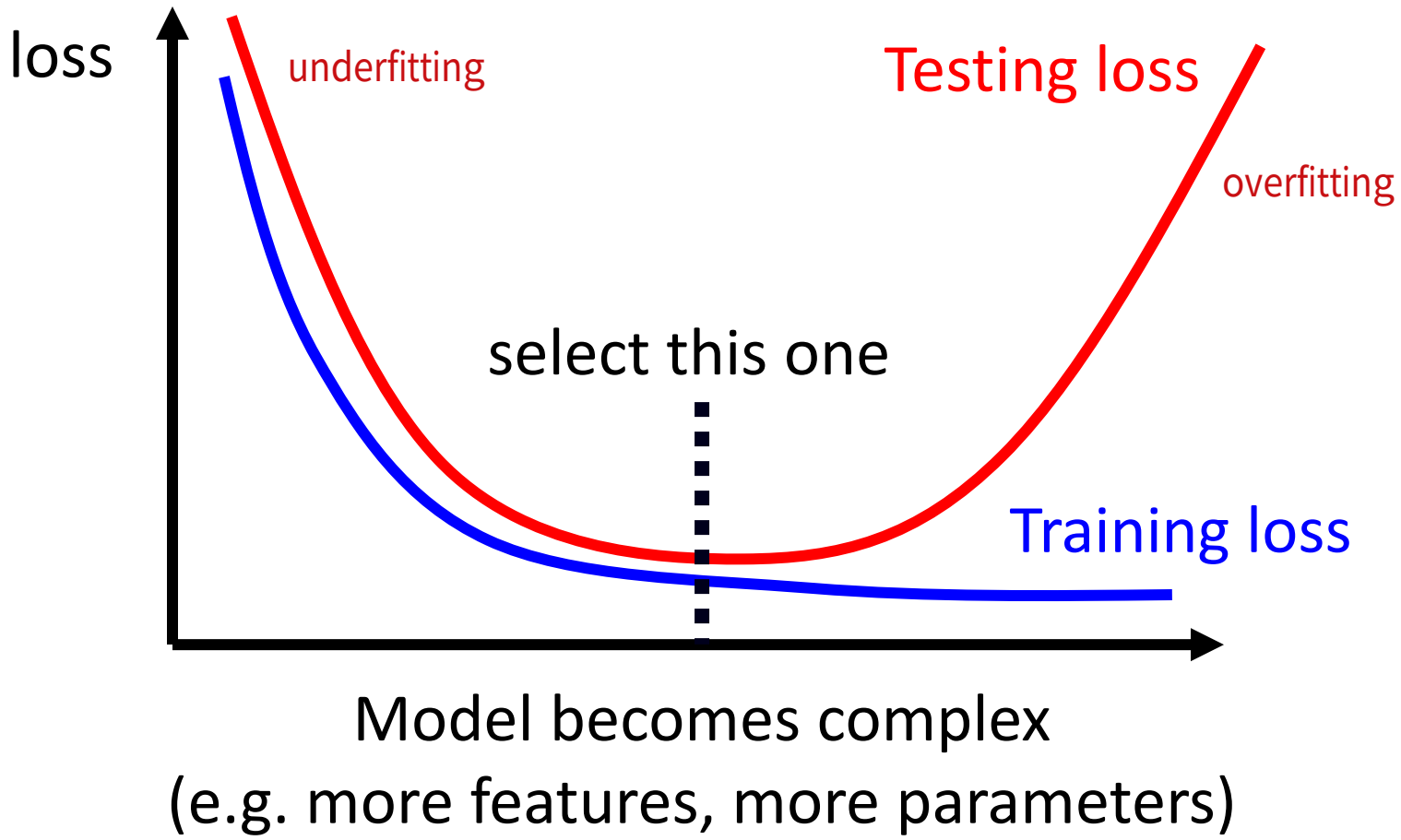
Training data

Testing data

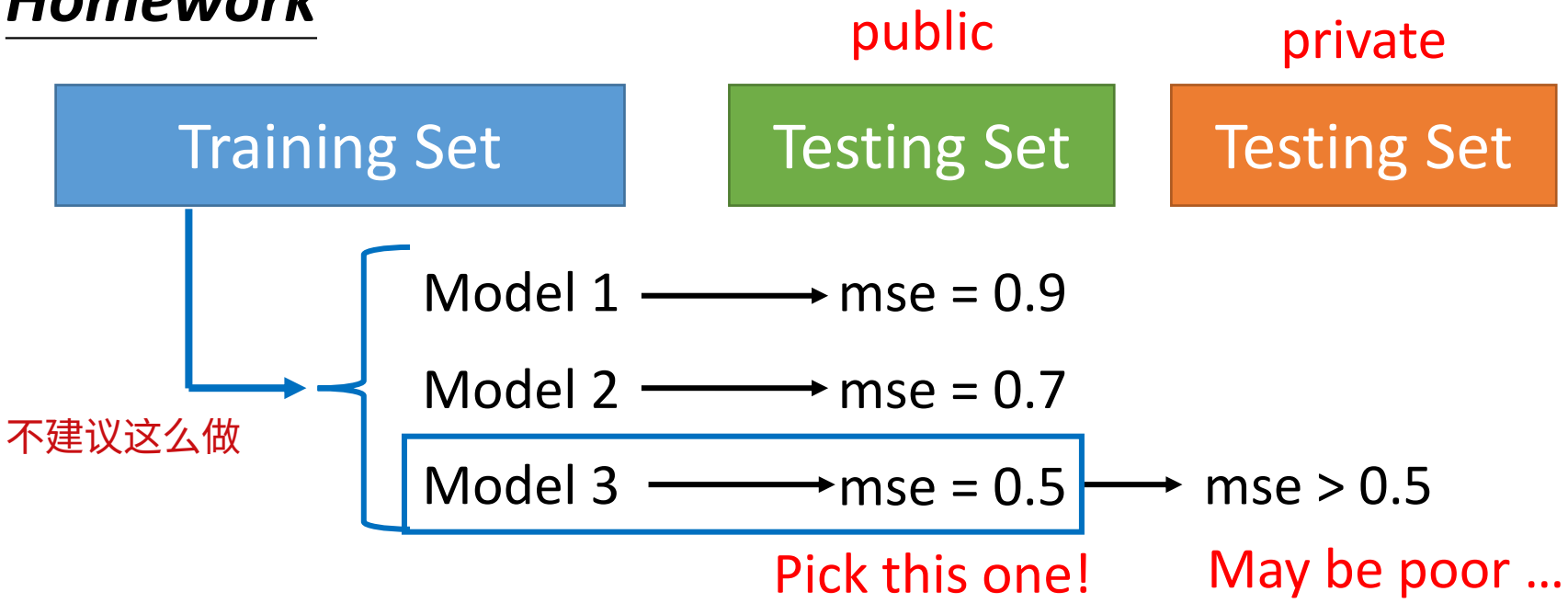


Back to model bias …

Bias-Complexity Trade-off



Homework



The extreme example again

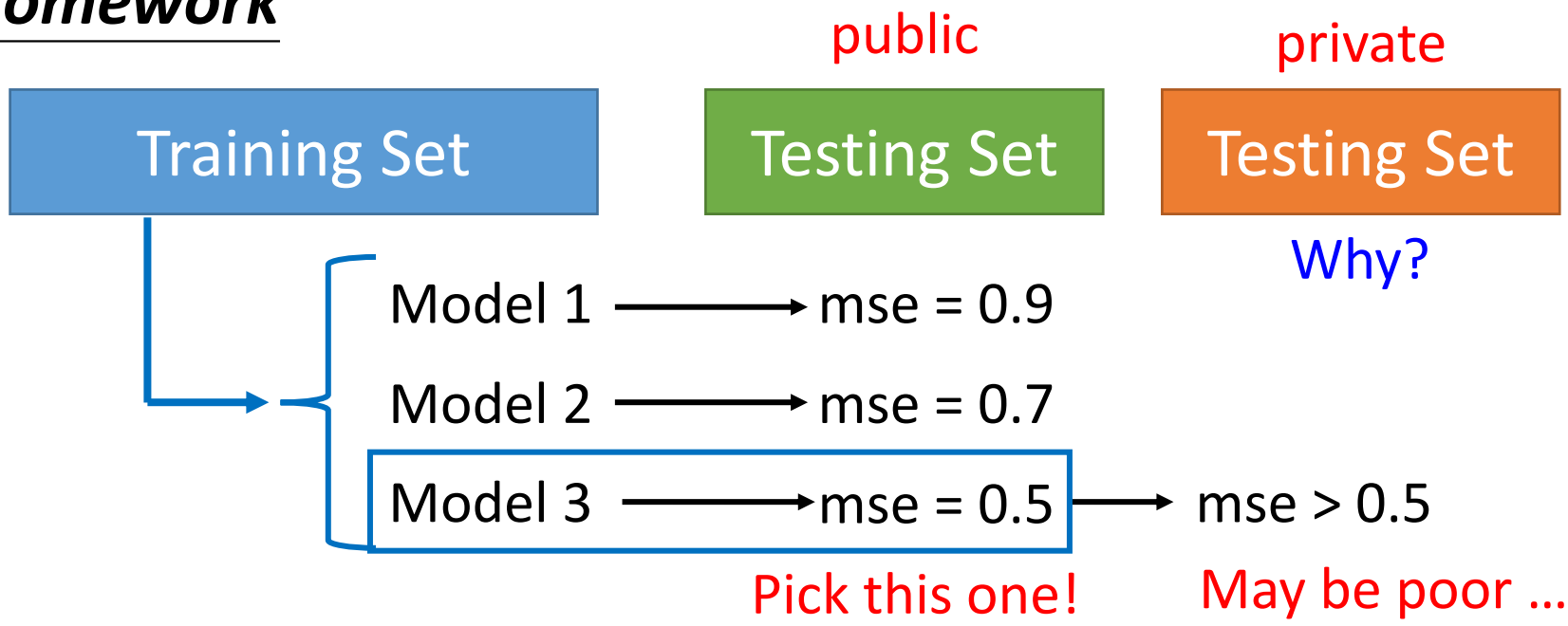
$$f_k(\boldsymbol{x}) = \begin{cases} \hat{y}^i & \exists \boldsymbol{x}^i = \boldsymbol{x} \\ random & otherwise \end{cases}$$

k: 1 - 10000000000000000000

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

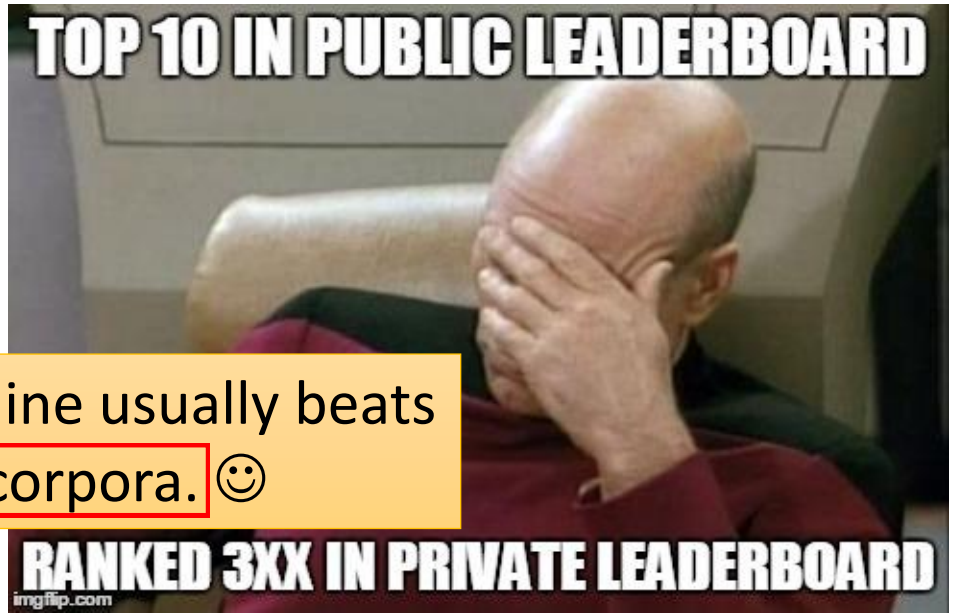
So you select $f_{56789}(\boldsymbol{x})$ …… Random on private testing set

Homework



What will happen?

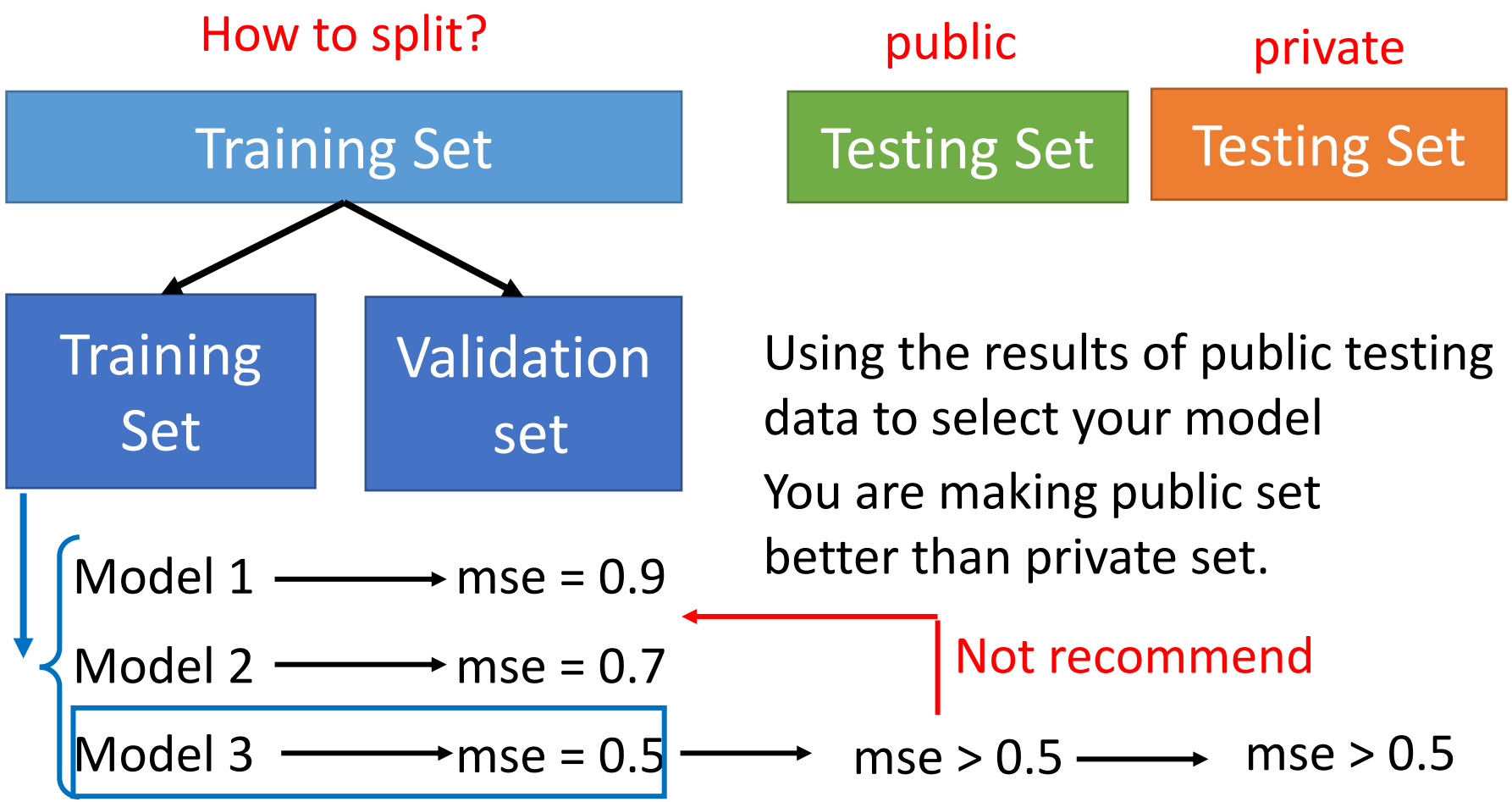
http://www.chioka.in/how-to-select-your-final-models-in-a-kaggle-competitio/



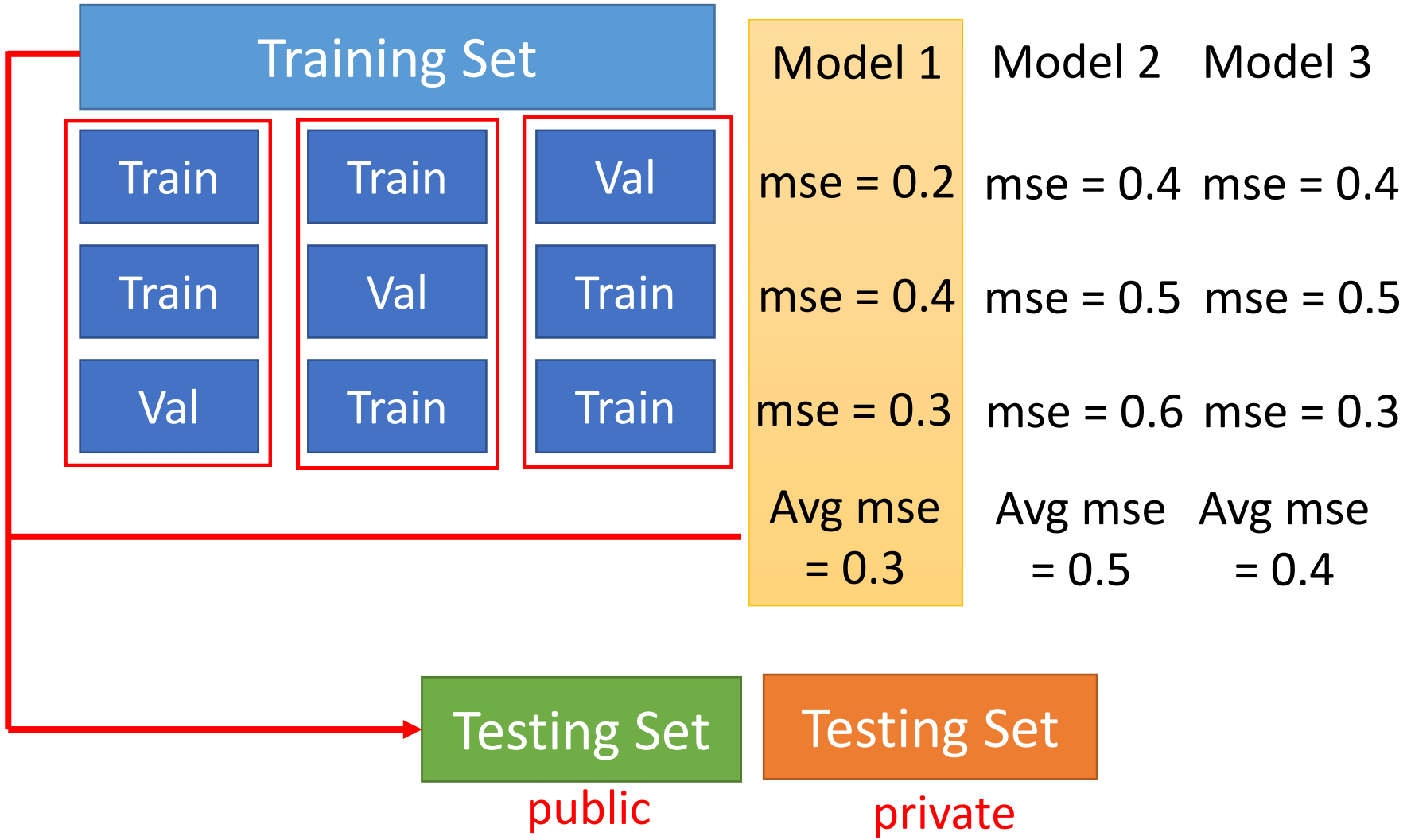
This explains why machine usually beats human on benchmark corpora. ☺

因为这可以算作是public set

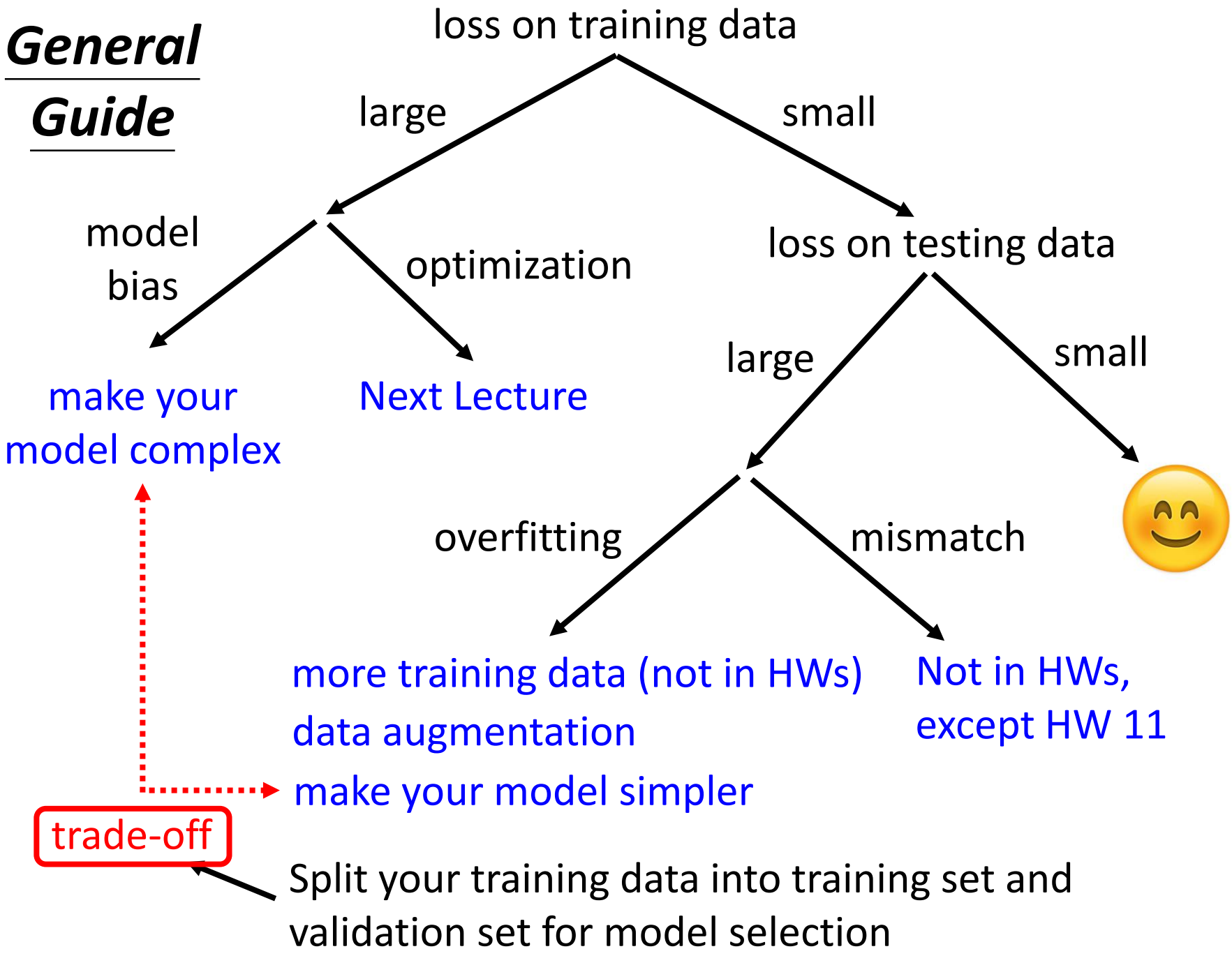
Cross Validation （CV，交叉验证）



N-fold Cross Validation K折交叉验证

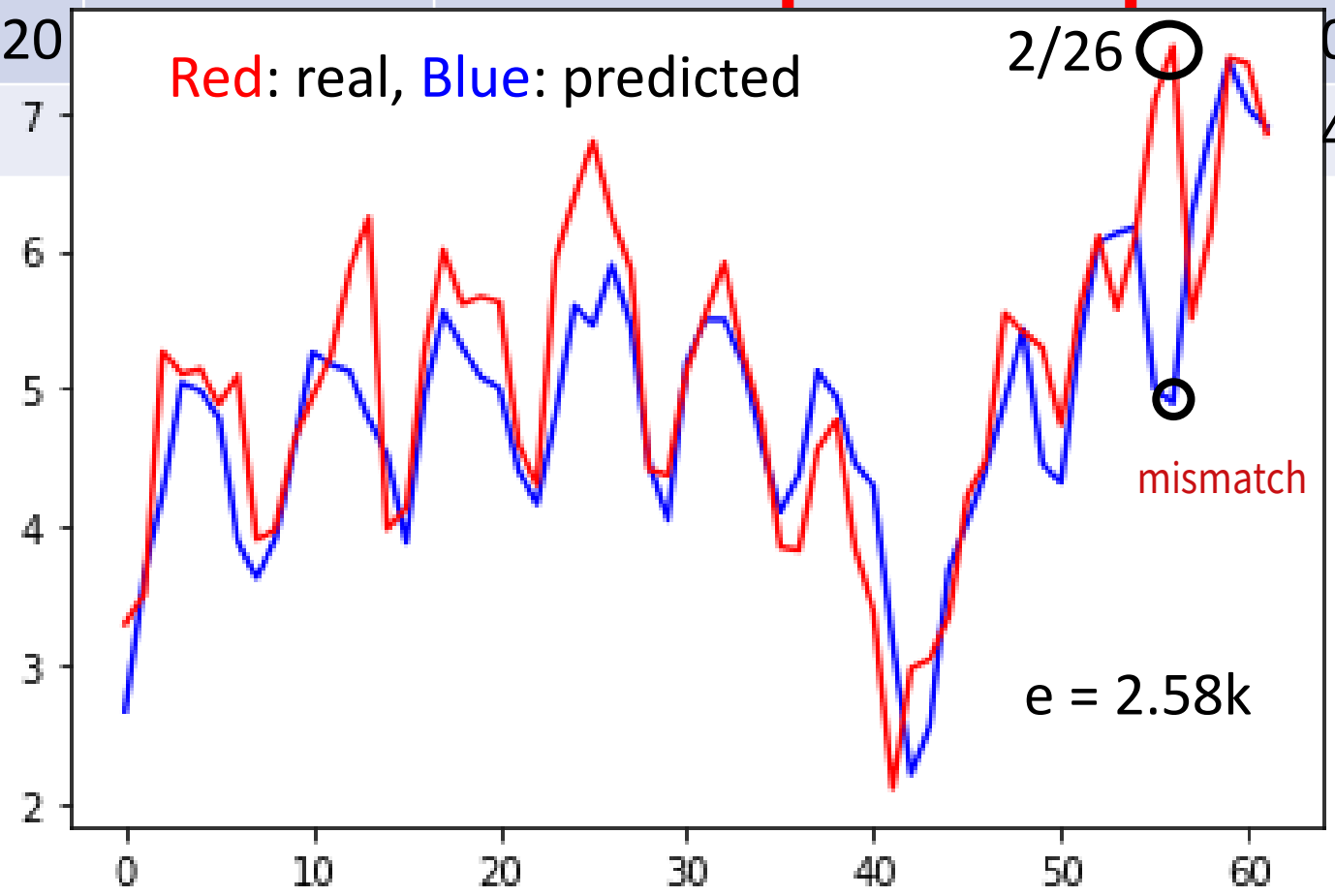


General Guide

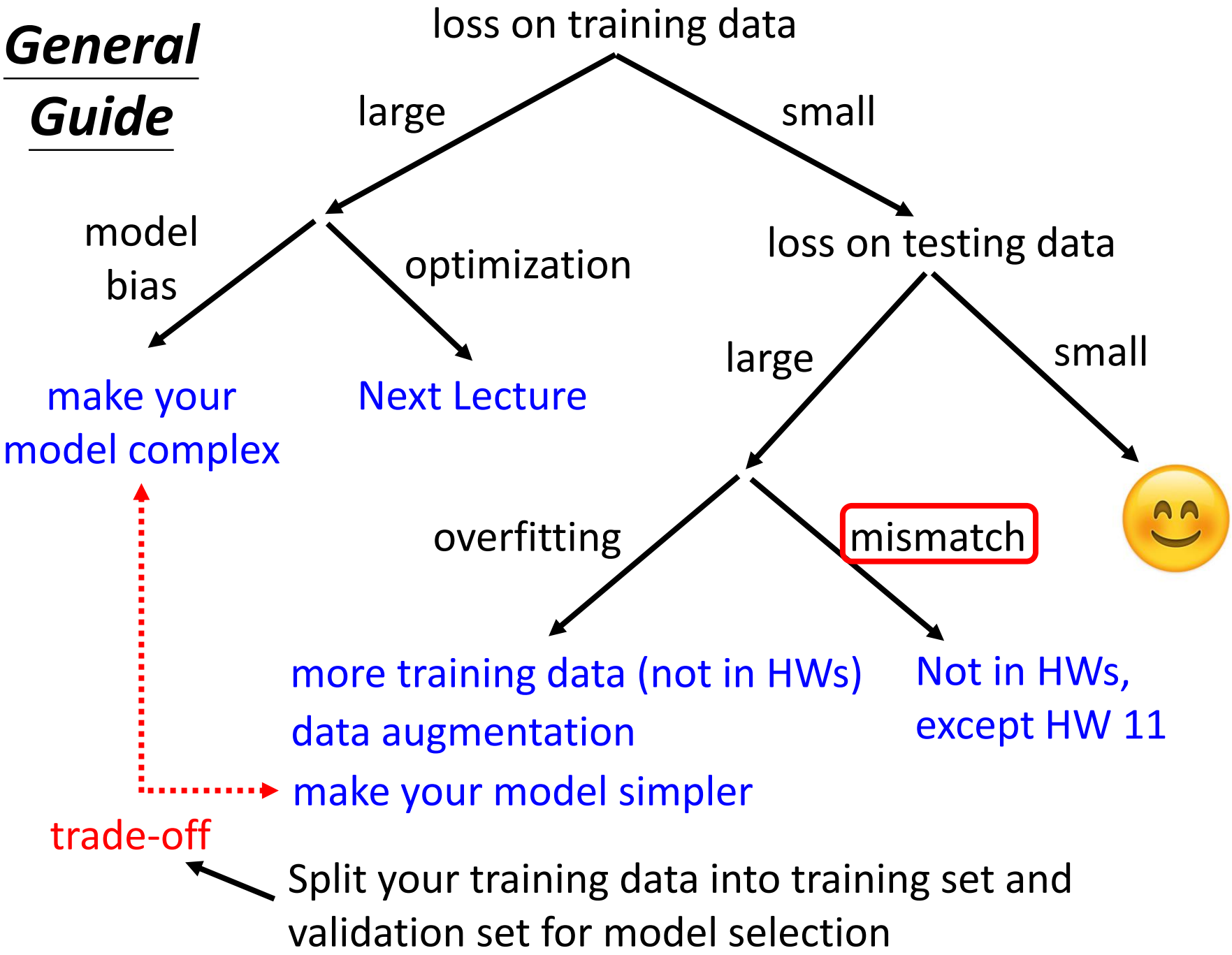


Let’s predict no. of views of 2/26!

	1 layer	2 layer	3 layer	4 layer
2017 – 2020				0k
2021				4k



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

Simply increasing the training data will not help.

Testing Data

General Guide