误差从哪里来

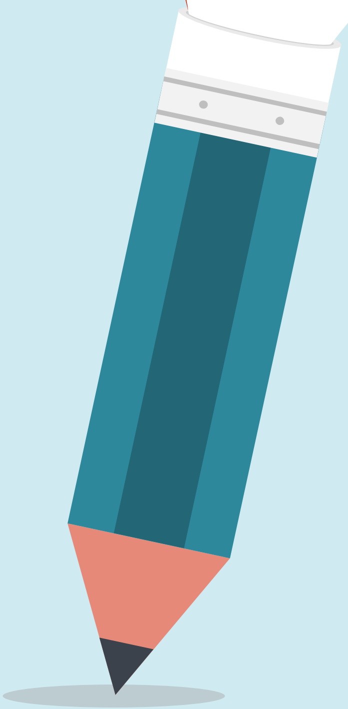
Where does the error come from ?

郝　奇

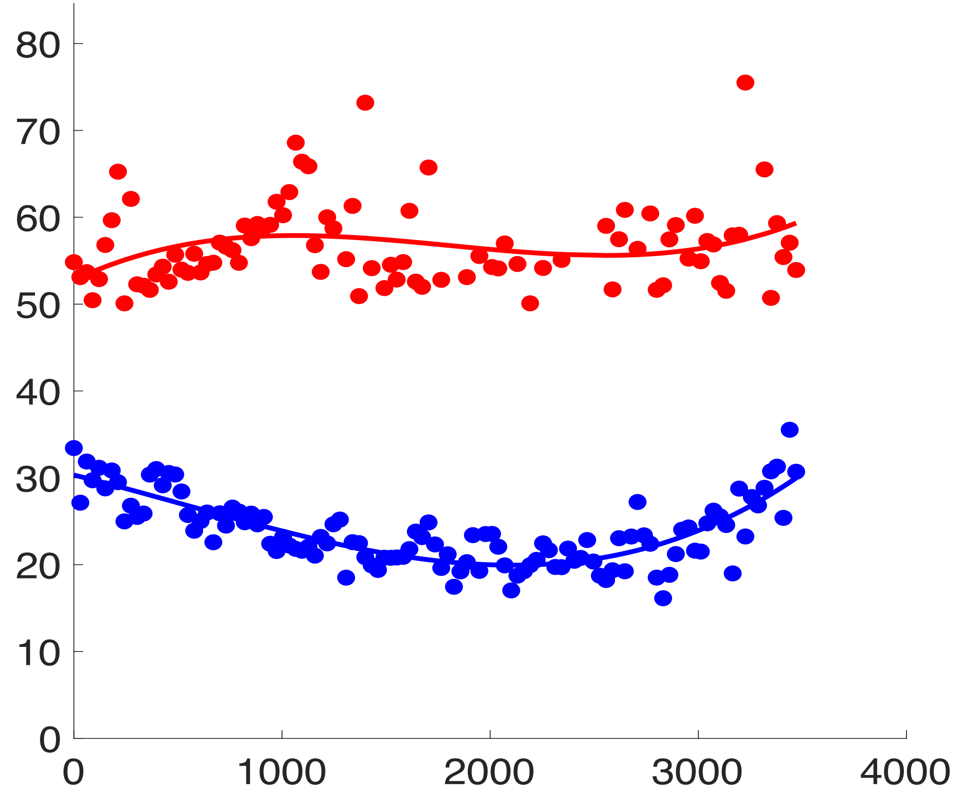
南京大学 天文与空间科学学院

Application of Machine Learning in Astronomy

机器学习在天文中的应用



Contents



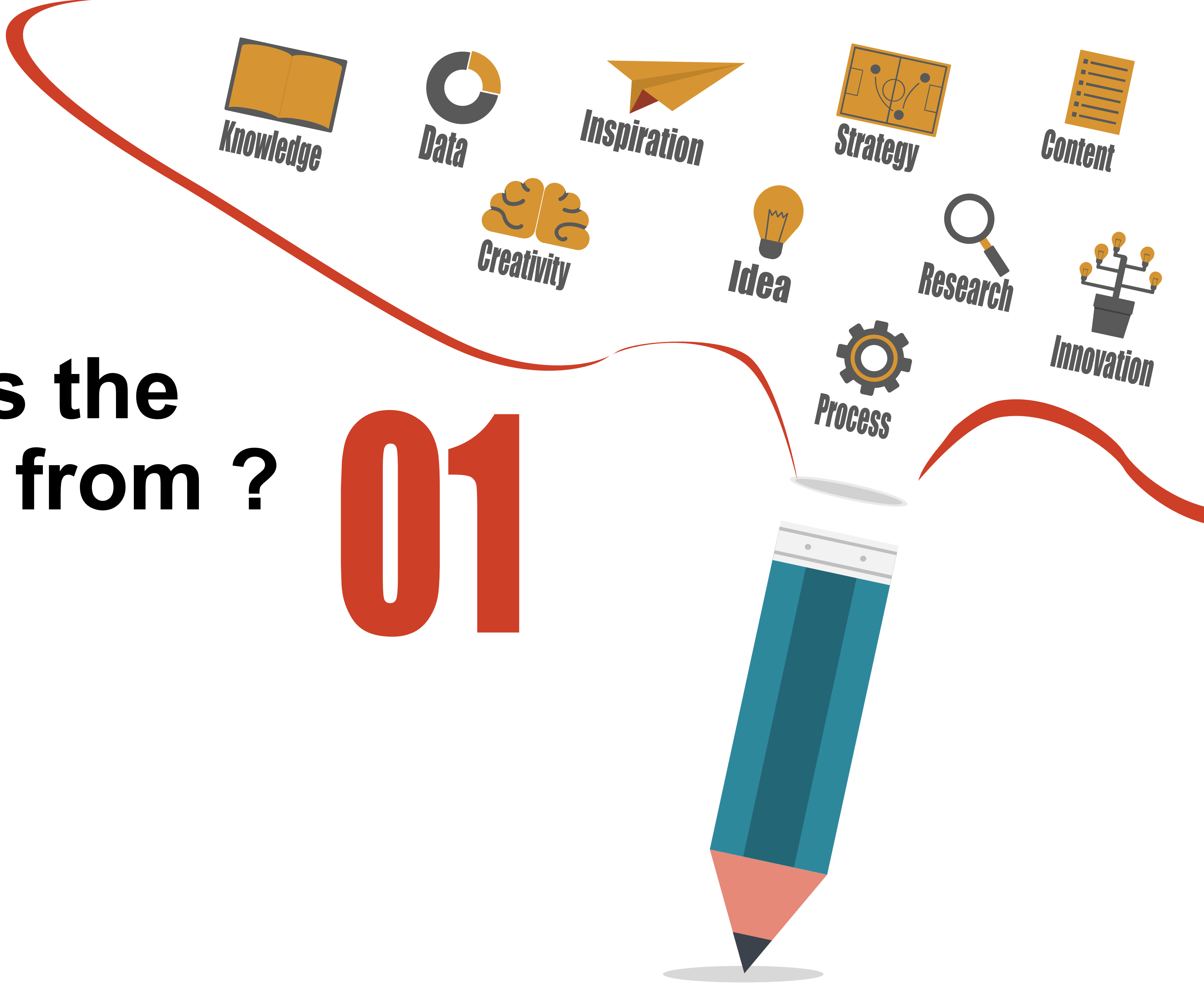
01 Where does the error come from ?

02 Bias and variance

03 Validation

Where does the error come from ?

01



1. Where does the error come from ?

$$\text{Loss} = \text{Bias} + \text{Variance} + \text{Noise}$$

1. Where does the error come from ?

Assume that the points in your training/test set are all taken from a similar distribution:

$y_i = f(x_i) + \epsilon_i$, where the noise ϵ_i satisfies $\mathbb{E}(\epsilon_i) = 0, Var(\epsilon_i) = \sigma^2$

For each example j in the test set, your prediction for $y_j = f\left(x_j\right) + \epsilon_j$ is an estimate $\hat{f}\left(x_j\right)$, then

$$
\begin{aligned}
\text{Test MSE} &= \mathbb{E}\left(\left(y - \hat{f}(x)\right)^2\right) \\
&= \mathbb{E}\left(\left(\epsilon + f(x) - \hat{f}(x)\right)^2\right) \\
&= \mathbb{E}(\epsilon^2) + \mathbb{E}\left(\left(f(x) - \hat{f}(x)\right)^2\right) \\
&= \sigma^2 + \left(\mathbb{E}\left(f(x) - \hat{f}(x)\right)\right)^2 + Var\left(f(x) - \hat{f}(x)\right) \\
&= \sigma^2 + \left(Bias\ \hat{f}(x)\right)^2 + Var\left(\hat{f}(x)\right)
\end{aligned}
$$

For simplicity, we ignore the noise term.

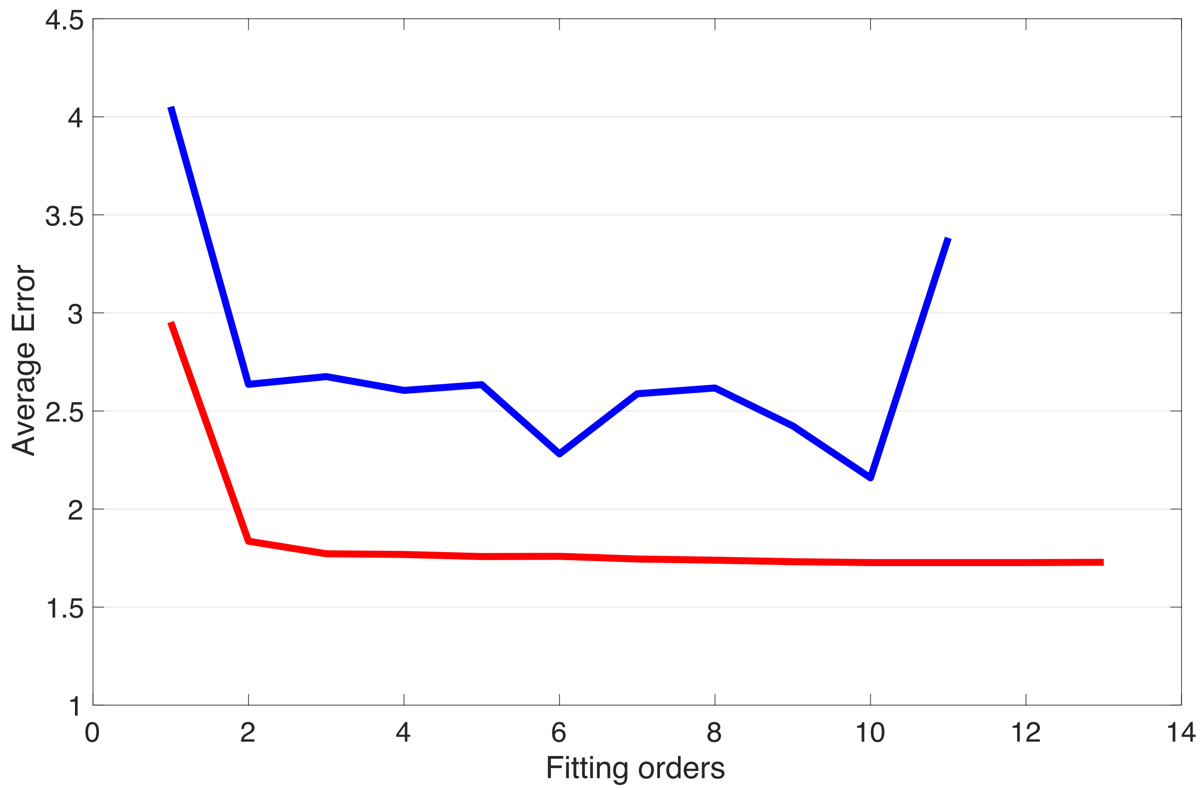
There is nothing we can do about the noise term σ^2 as we can not predict the noise ϵ by definition.

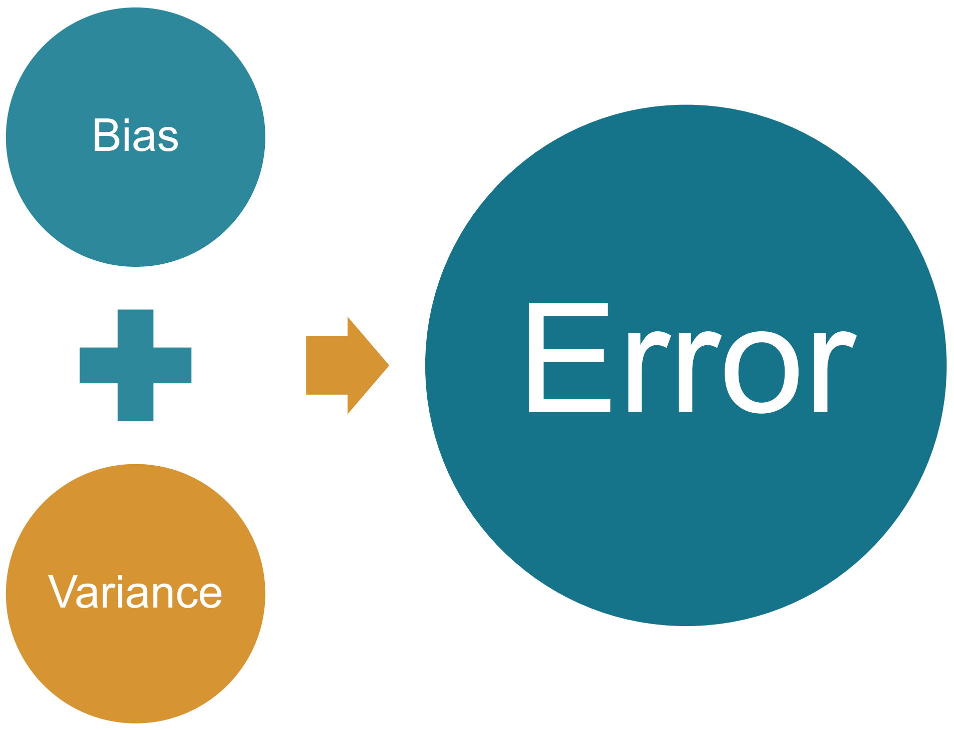
Bias and Variance

02

2. Bias and Variance

Average error on training/testing data





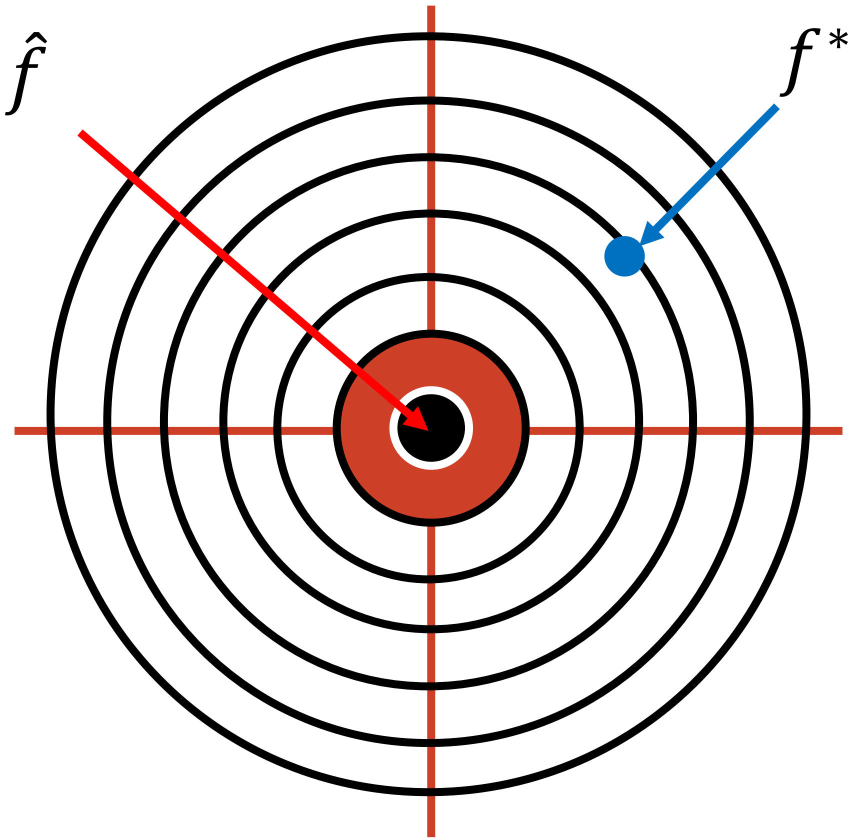
2. Bias and Variance

2.1 Estimator

Assume the real $\hat{y} = \hat{f}(x)$

From training data, we find f^*

f^* is an estimator of $\hat{f}$



Cite: Hung-yi Lee, Machine Learning

2. Bias and Variance

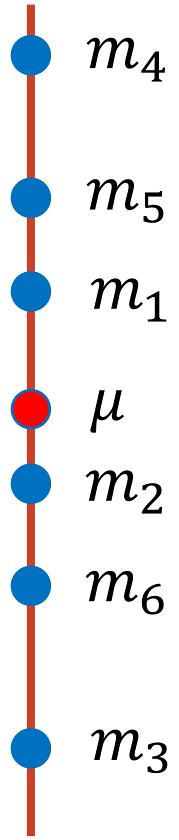
2.1 Estimator

- Estimate the mean of a variable x
  - Assume the mean of x is μ
  - Assume the variance of x is σ^2
- Estimator of the mean μ
  - Sample n points: $\{x^1, x^2, x^3, \ldots, x^n\}$

$$m = \frac{1}{n}\sum_n x^n \neq \mu$$

$$E[m] = E\left[\frac{1}{n}\sum_n x^n\right] = \frac{1}{n}\sum_n E[x^n] = \mu$$

unbiased



Cite: Hung-yi Lee, Machine Learning

2. Bias and Variance

2.1 Estimator

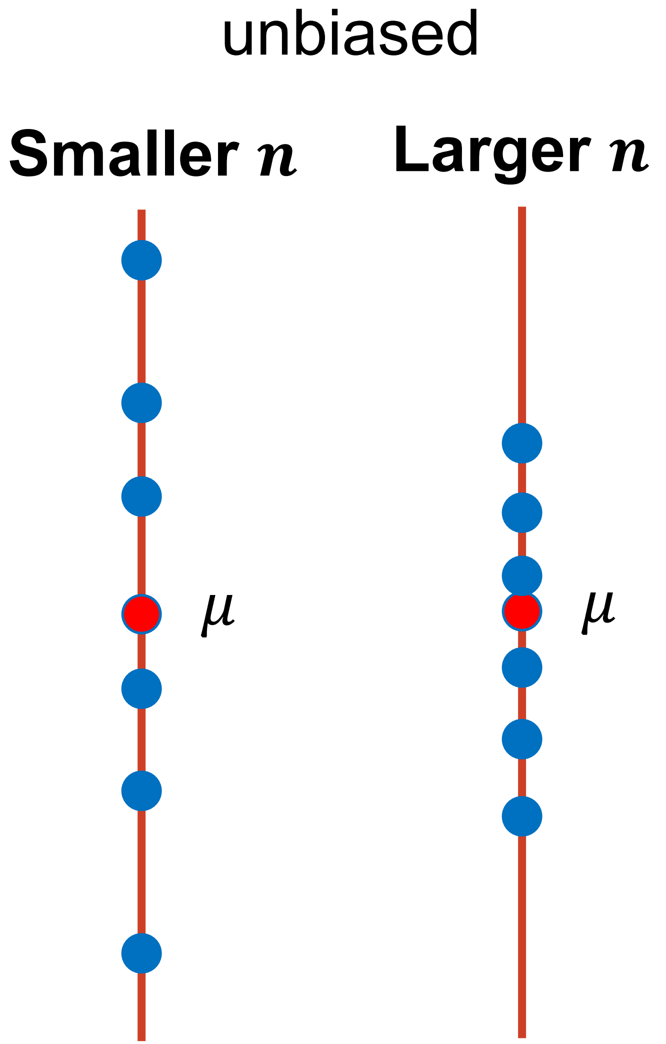
- Estimate the mean of a variable x
  - Assume the mean of x is μ
  - Assume the variance of x is σ^2

- Estimator of the mean μ
  - Sample n points: $\{x^1, x^2, x^3, \ldots, x^n\}$

$$m = \frac{1}{n}\sum_n x^n \neq \mu$$

$$Var[m] = \frac{\sigma^2}{n}$$

Variance depends on the number n of samples



Cite: Hung-yi Lee, Machine Learning

2. Bias and Variance

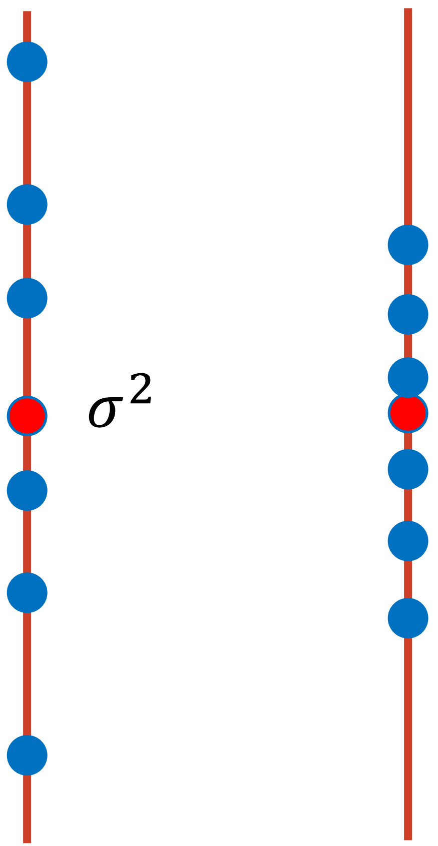
2.1 Estimator

- Estimate the mean of a variable x
  - Assume the mean of x is μ
  - Assume the variance of x is σ^2
- Estimator of the variance σ^2
  - Sample n points: $\{x^1, x^2, x^3, \ldots, x^n\}$

$$m = \frac{1}{n}\sum_n x^n \neq \mu \qquad s^2 = \frac{1}{n}\sum_n (x^n - m)^2$$

Biased estimator $E[s^2] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$

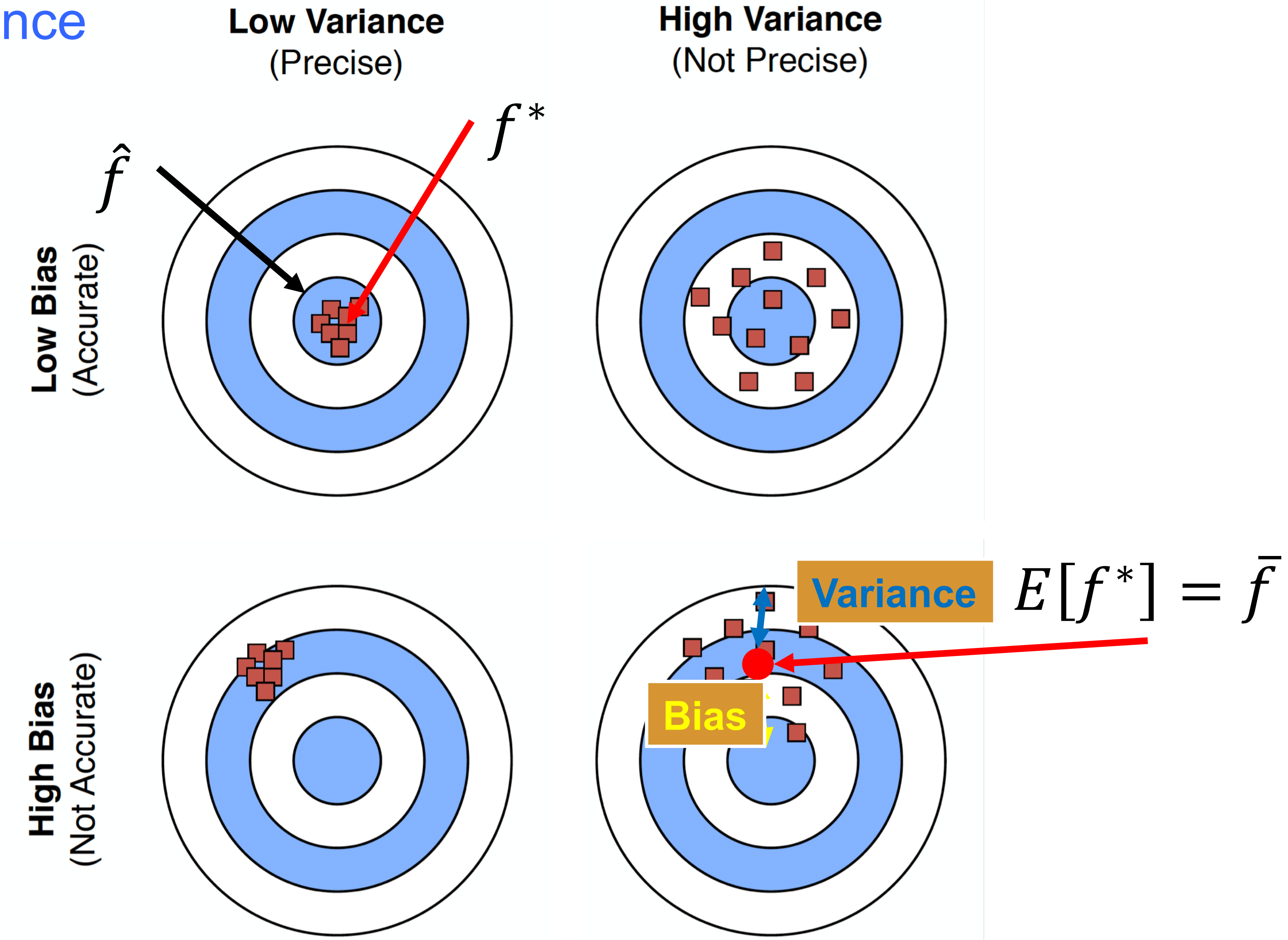
Increase n



Cite: Hung-yi Lee, Machine Learning

2. Bias and Variance

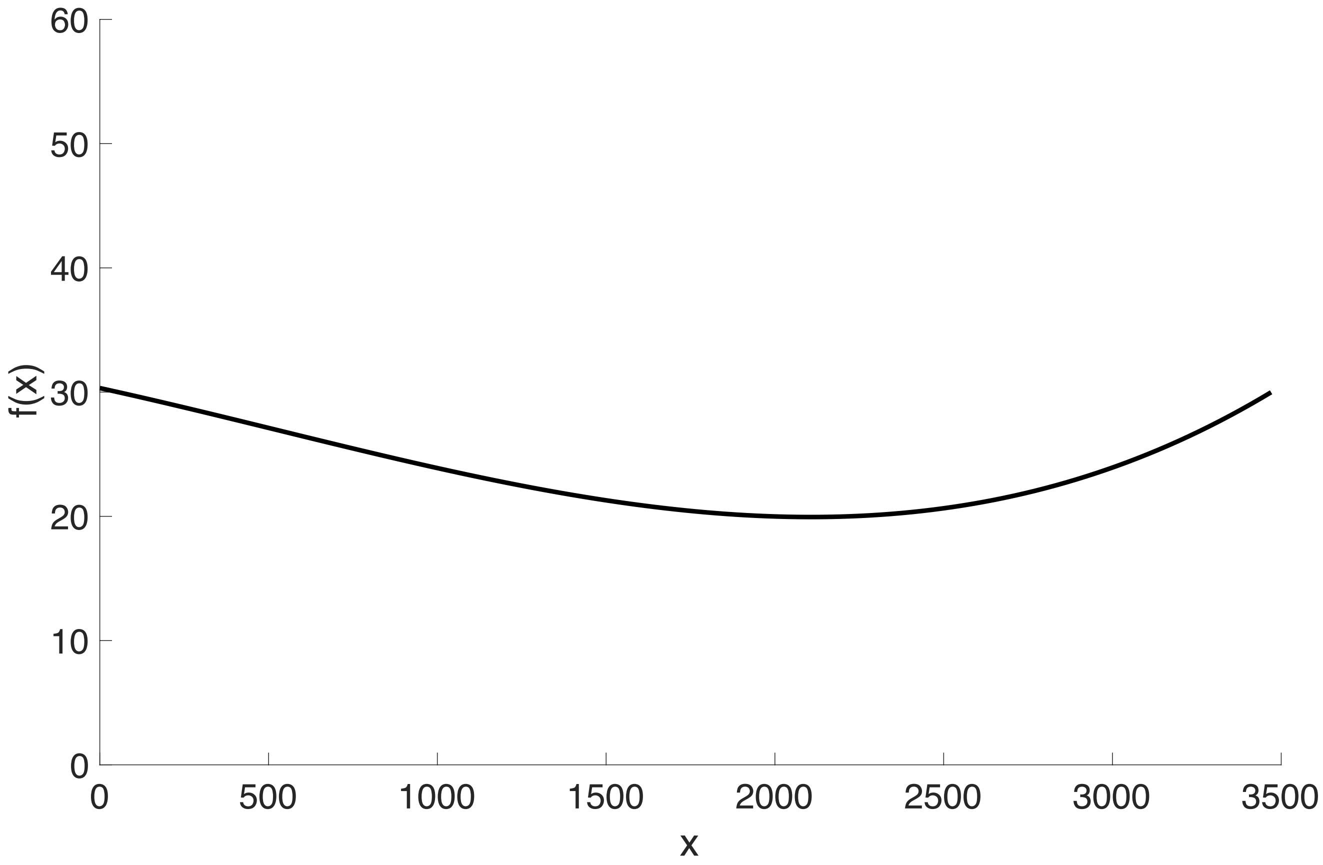
2.2 Bias V.S. Variance



Cite: Sebastian Raschka, Machine Learning

2. Bias and Variance

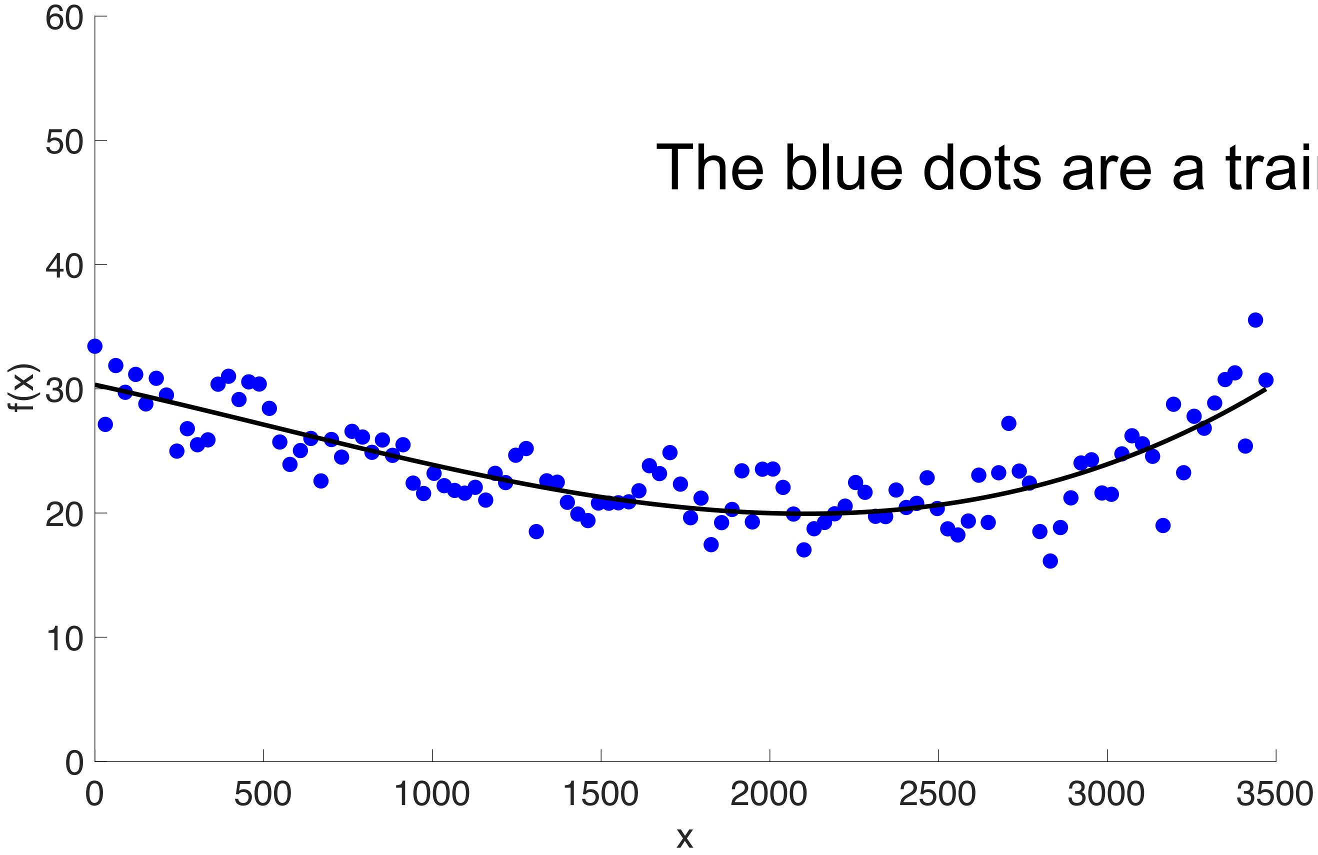
2.2 Bias V.S. Variance



Assume f(x) is some true (target) function

2. Bias and Variance

2.2 Bias V.S. Variance



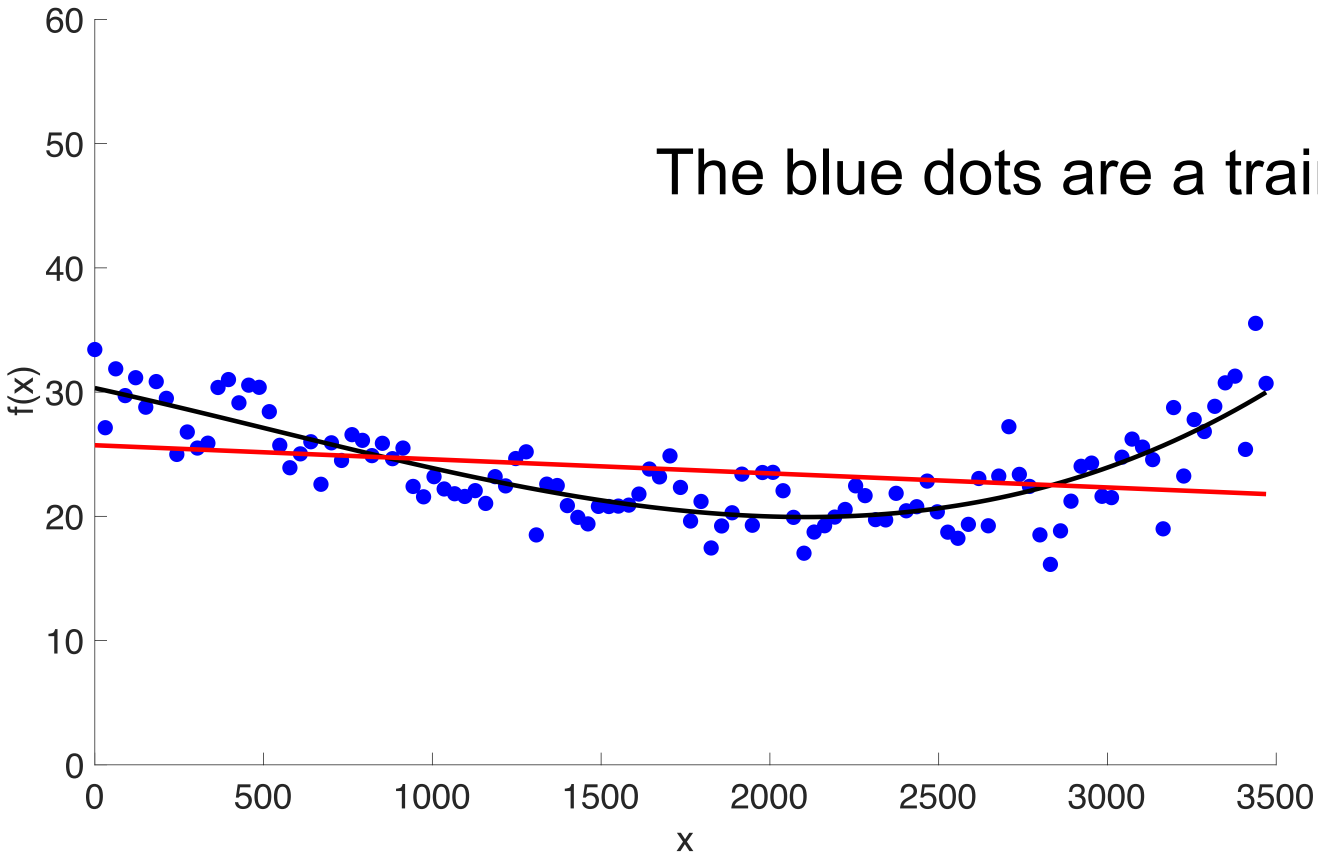
The blue dots are a training dataset

Assume f(x) is some true (target) function

2. Bias and Variance

2.2 Bias V.S. Variance

High Bias



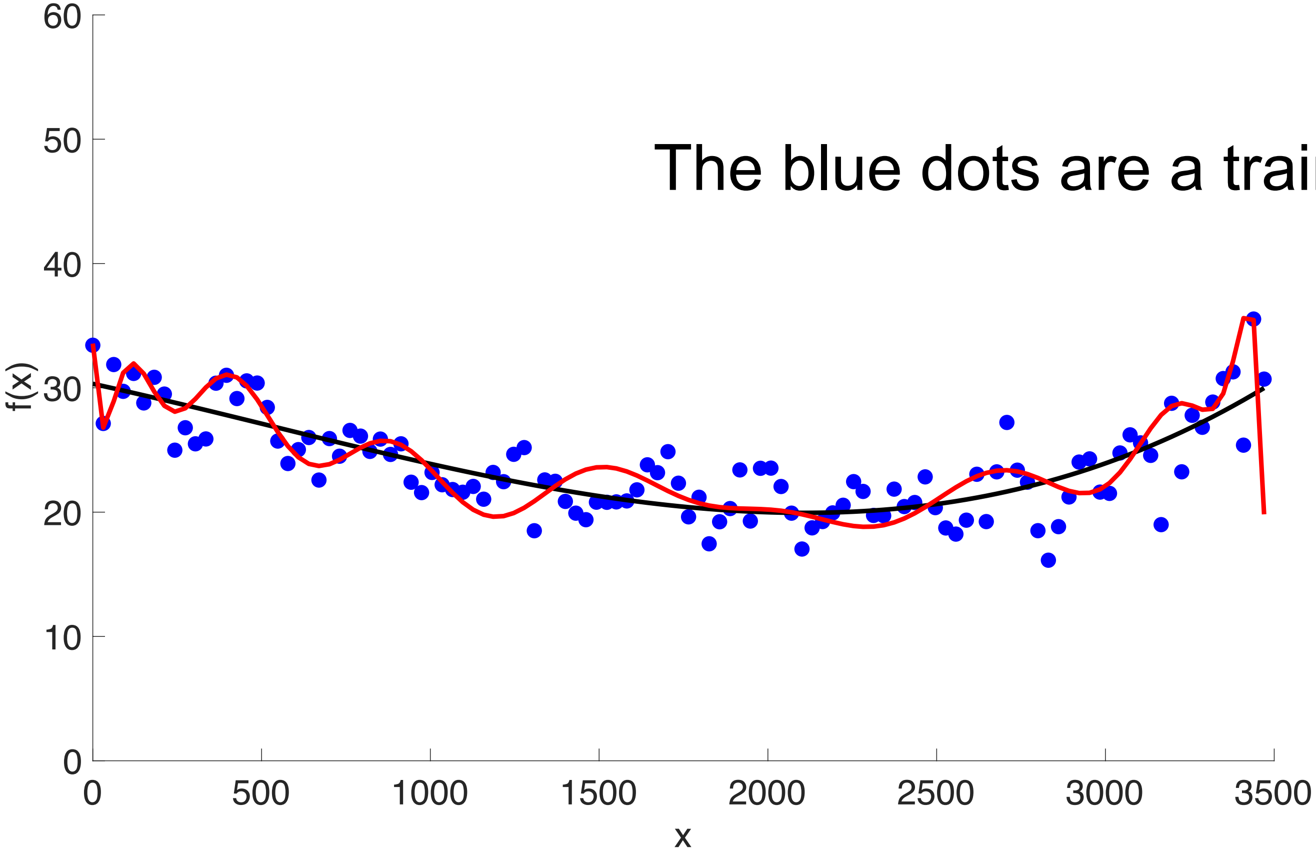
The blue dots are a training dataset

Assume f(x) is some true (target) function

2. Bias and Variance

2.2 Bias V.S. Variance

High Variance

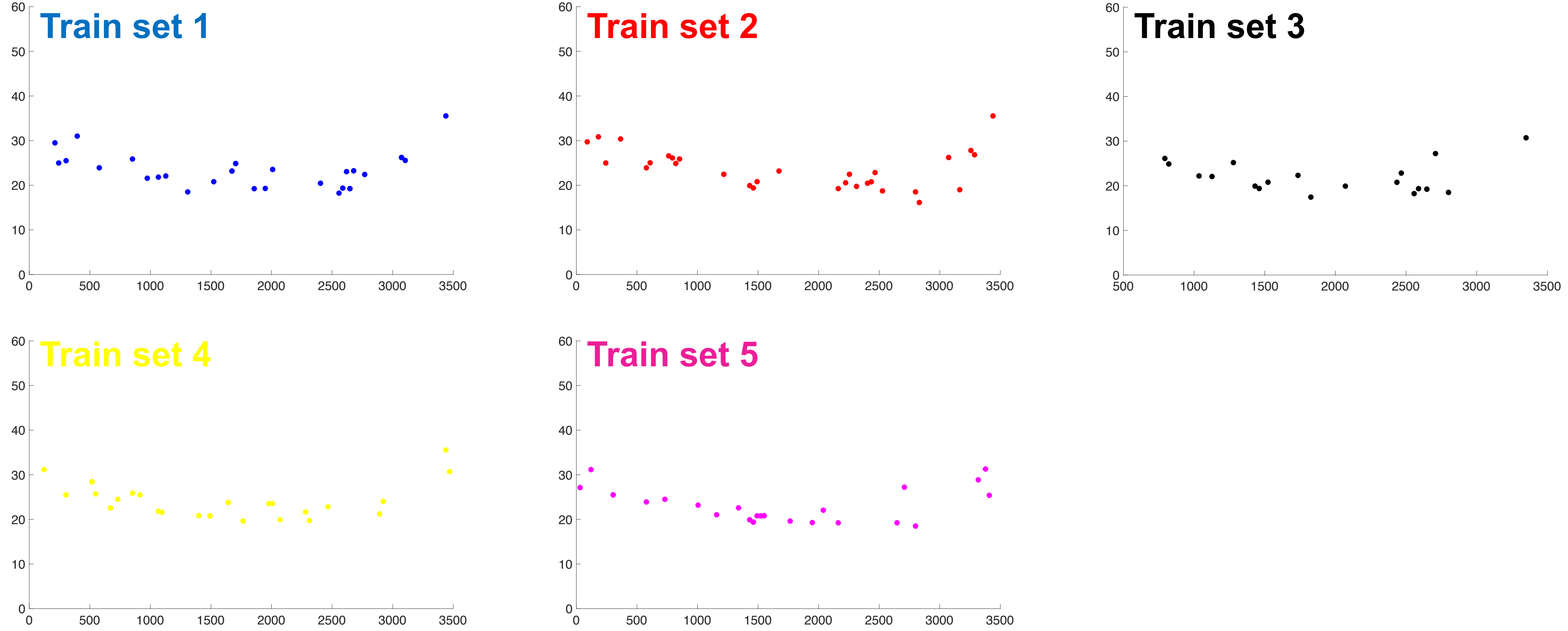


Assume f(x) is some true (target) function

2. Bias and Variance

2.2 Bias V.S. Variance

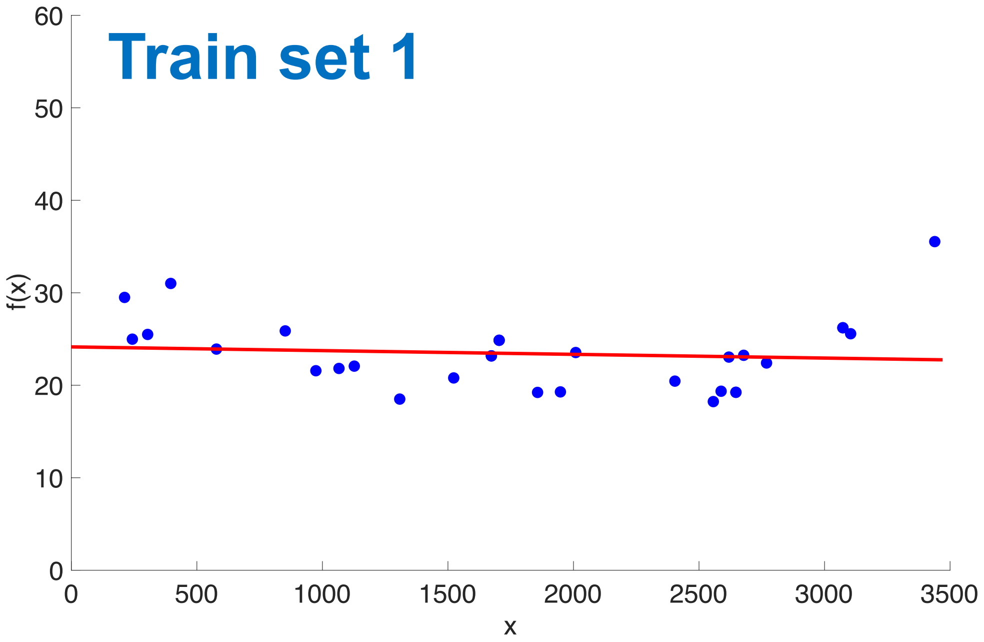
Suppose we have multiple training sets

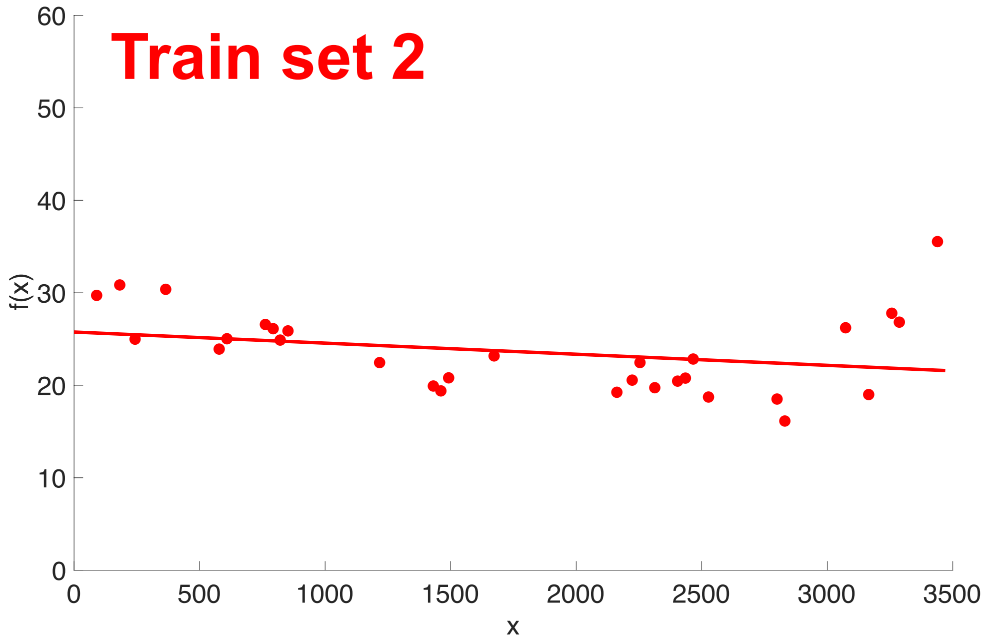


2. Bias and Variance

2.2 Bias V.S. Variance

In different training dataset , we use the same model, but obtain different f^*





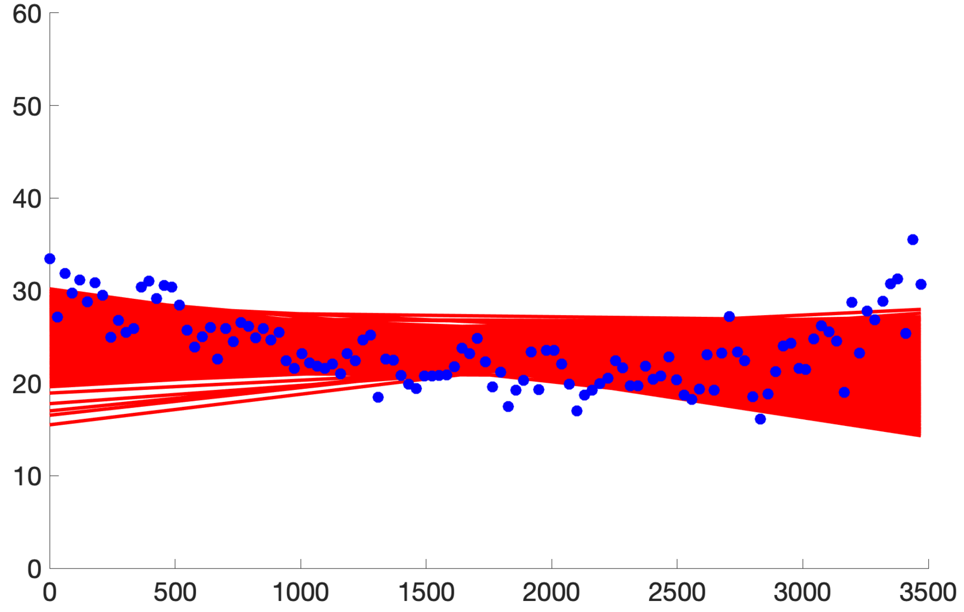
$$y = b + w \cdot x_{date}$$

$$y = b + w \cdot x_{date}$$

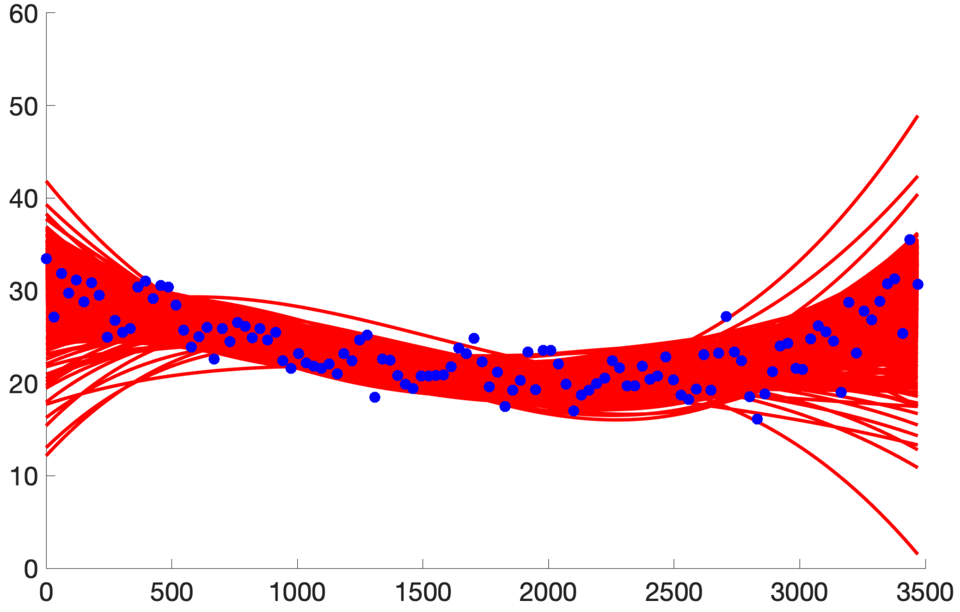
2. Bias and Variance

2.2 Bias V.S. Variance

f^* in 1000 training sets



$$y = b + w \cdot x_{date}$$

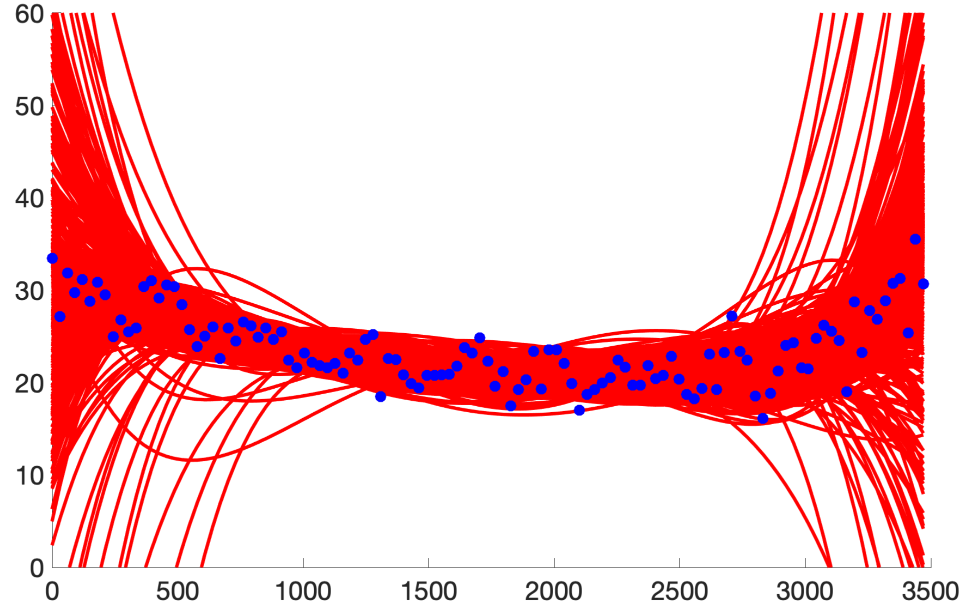


$$y = b + w_1 \cdot x_{date} + w_2 \cdot (x_{date})^2 + w_3 \cdot (x_{date})^3$$

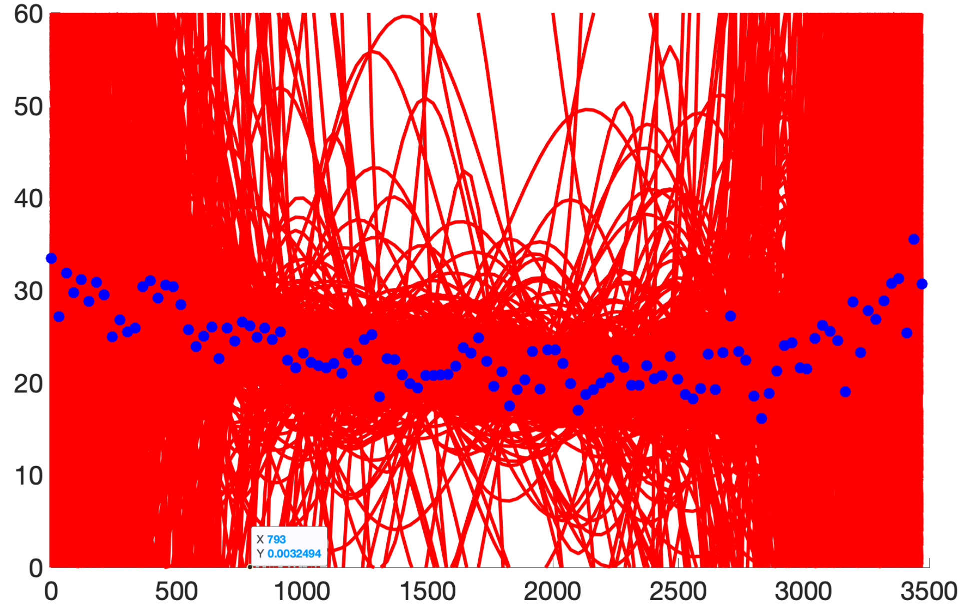
2. Bias and Variance

2.2 Bias V.S. Variance

f^* in 1000 training sets



$$y = b + w_1 \cdot x_{date} + w_2 \cdot (x_{date})^2 + w_3 \cdot (x_{date})^3 + w_4 \cdot (x_{date})^4 + w_5 \cdot (x_{date})^5$$

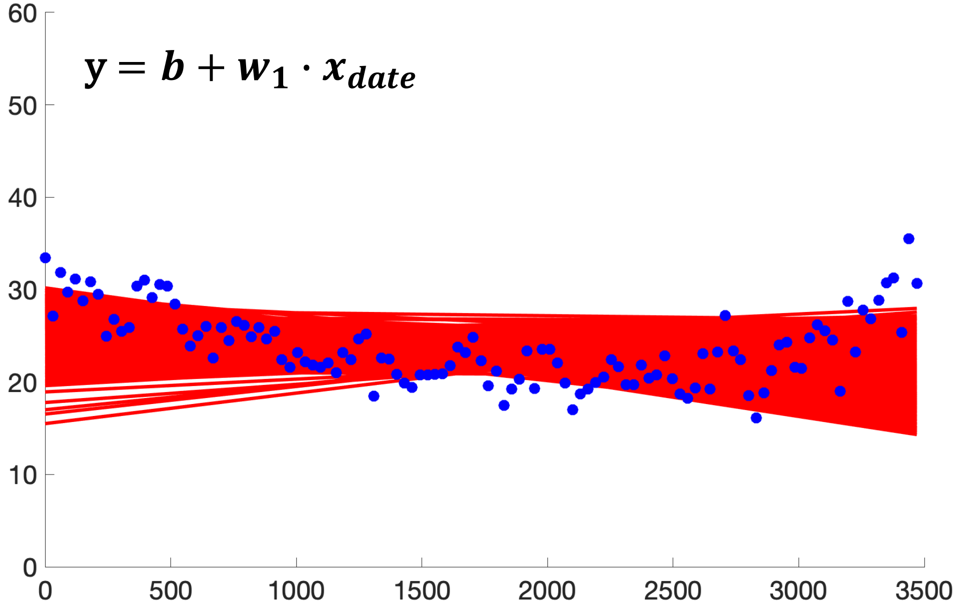


$$y = b + w_1 \cdot x_{date} + w_2 \cdot (x_{date})^2 + \cdots + w_{10} \cdot (x_{date})^{10} + w_{11} \cdot (x_{date})^{11}$$

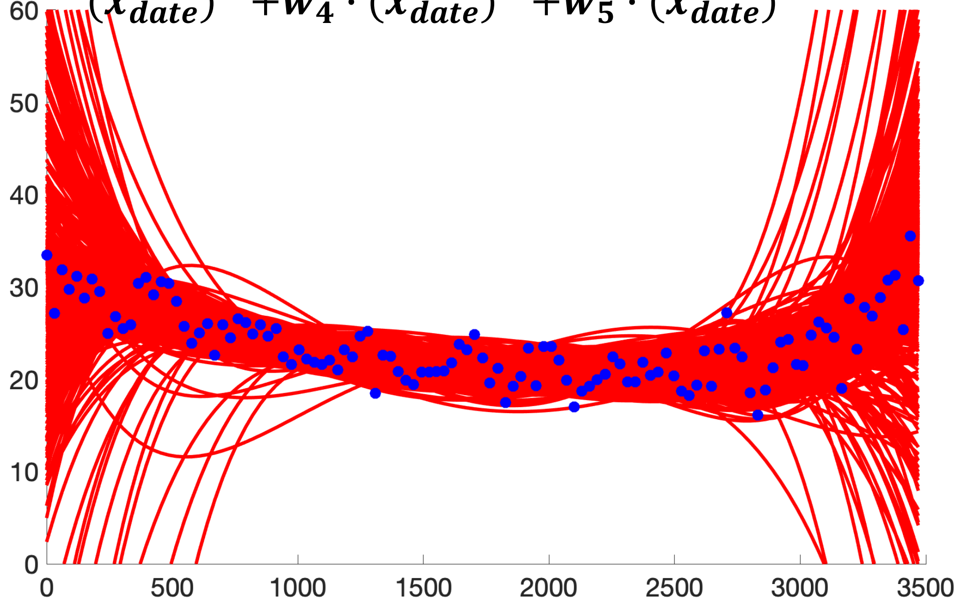
2. Bias and Variance

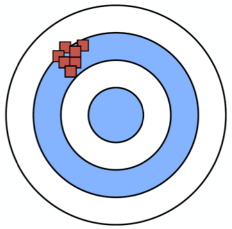
2.2 Bias V.S. Variance

Variance

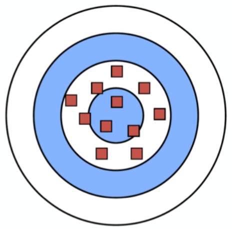


$y = b + w_1 \cdot x_{date} + w_2 \cdot (x_{date})^2 + w_3 \cdot (x_{date})^3 + w_4 \cdot (x_{date})^4 + w_5 \cdot (x_{date})^5$





Small Variance



Large Variance

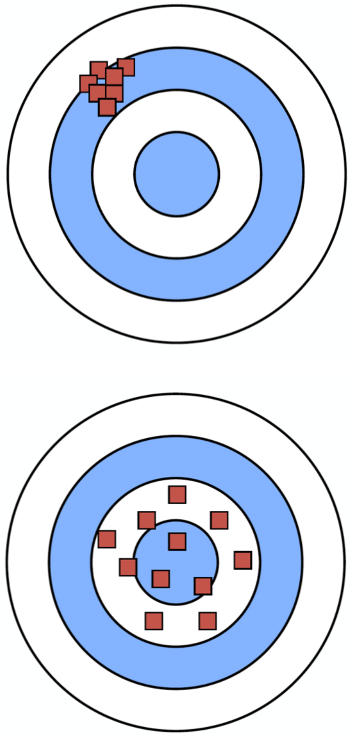
Simpler model is less influenced by the sampled data

2. Bias and Variance

2.2 Bias V.S. Variance

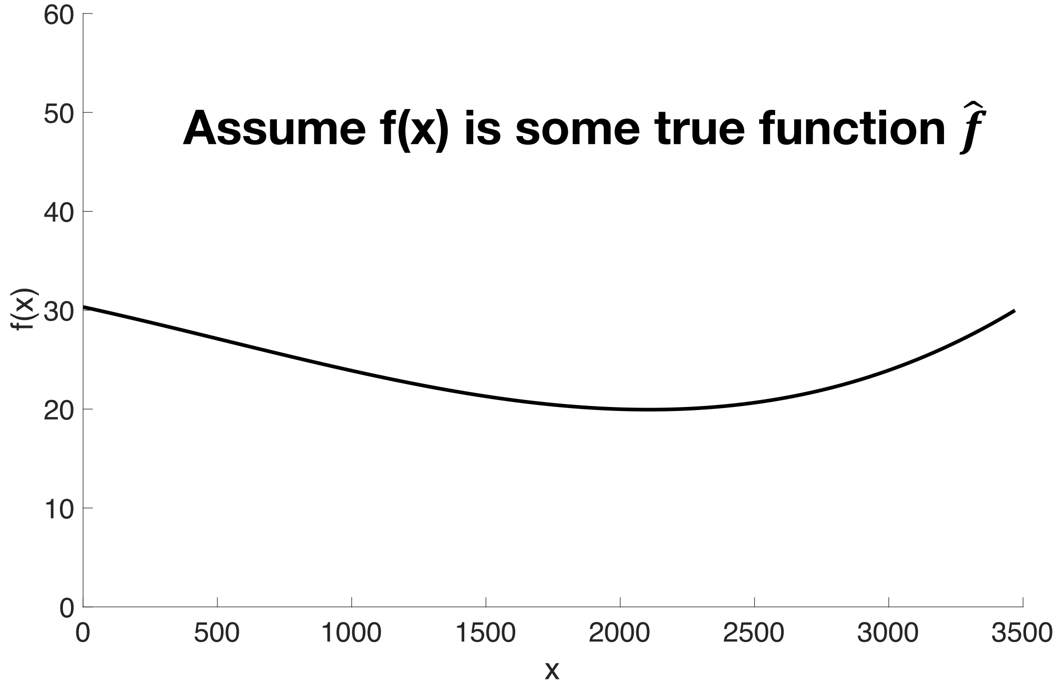
Bias

$E[f^*] = \bar{f}$ **If we average all the f^*, is it close to $\hat{f}$?**



Large Bias

Small Bias

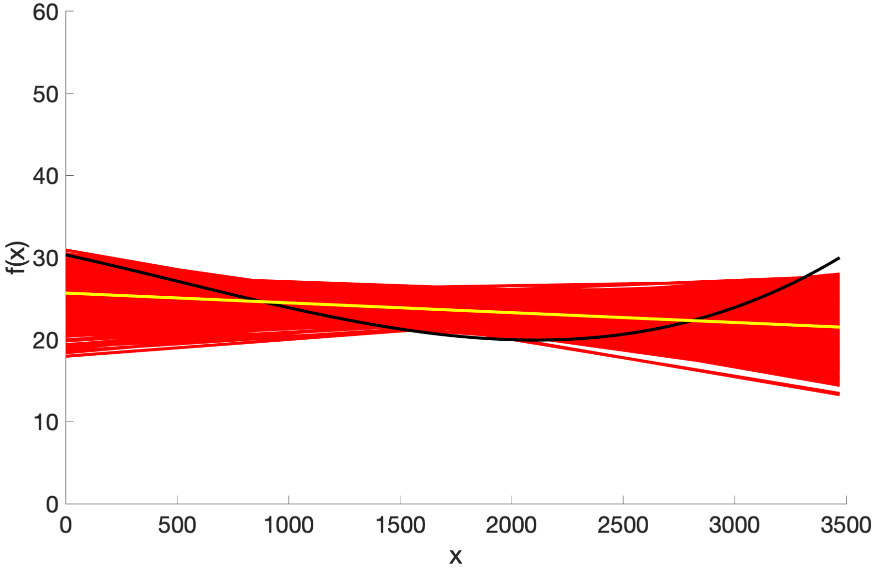


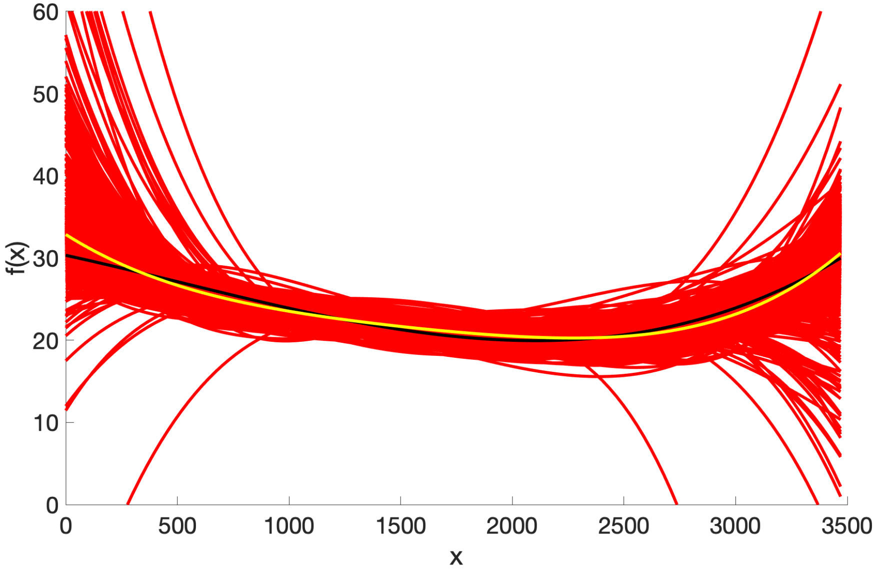
2. Bias and Variance

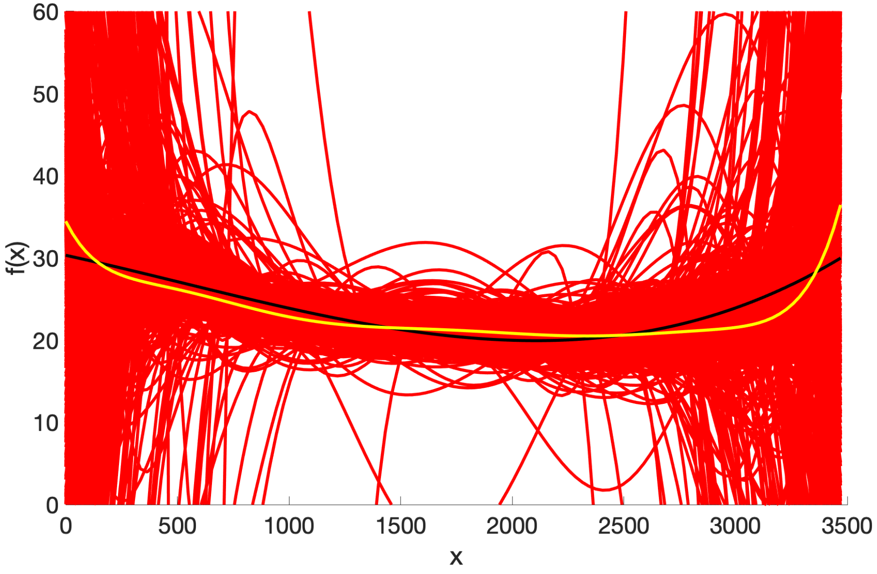
2.2 Bias V.S. Variance

Bias

- **Black** curve: the true function $\hat{f}$
- **Red** curves: f^* in 1000 training sets
- **Yellow** curve: the average of 1000 $f^* = \bar{f}$



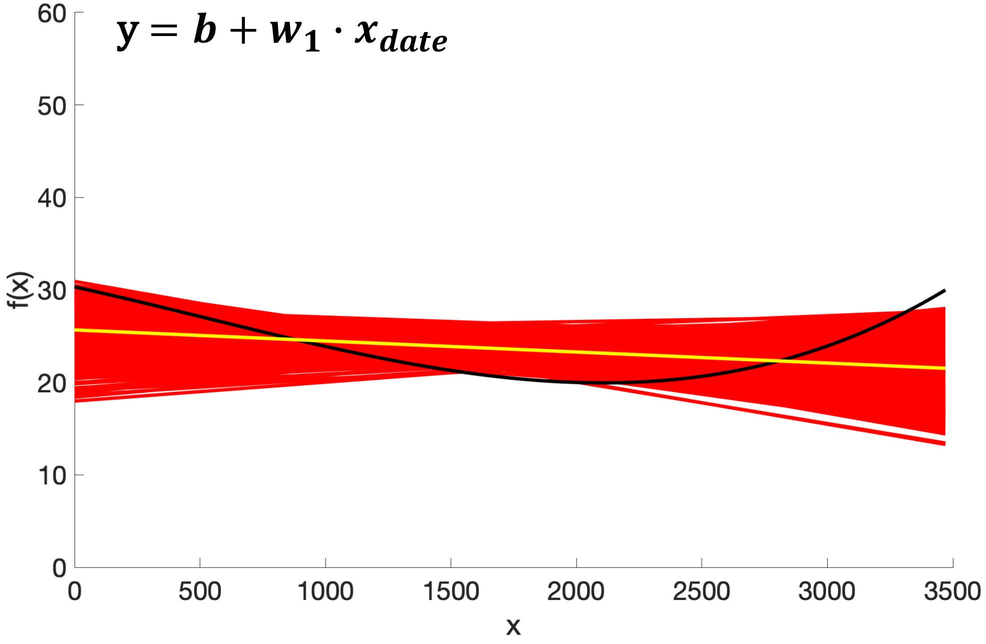




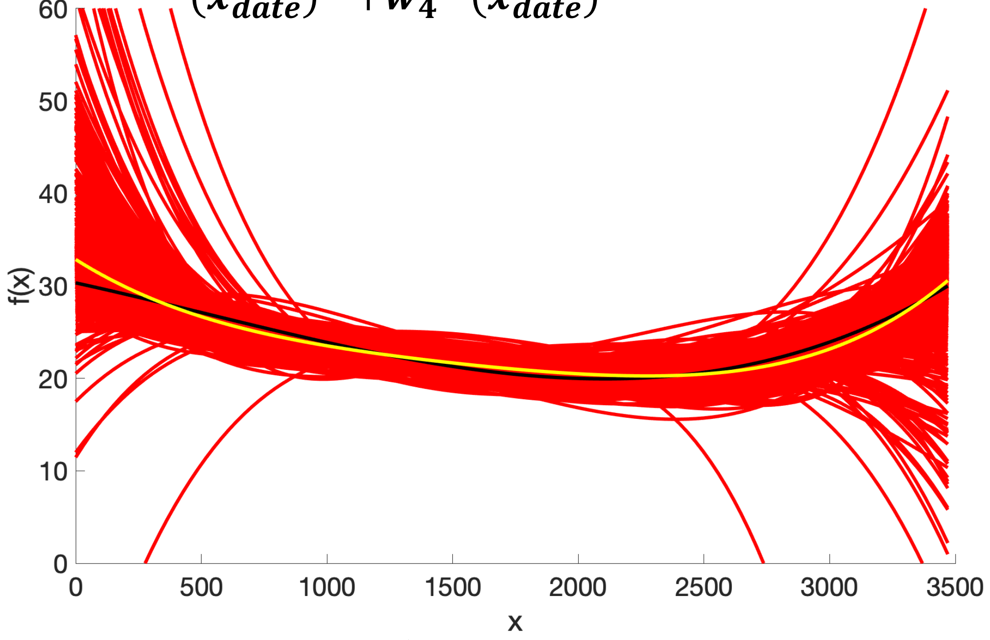
2. Bias and Variance

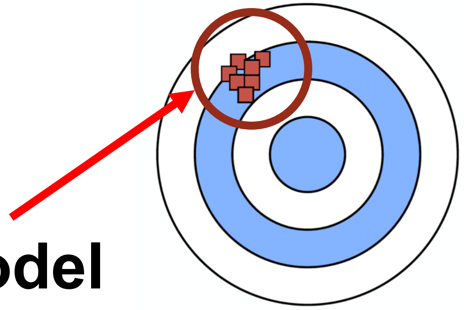
2.2 Bias V.S. Variance

Bias



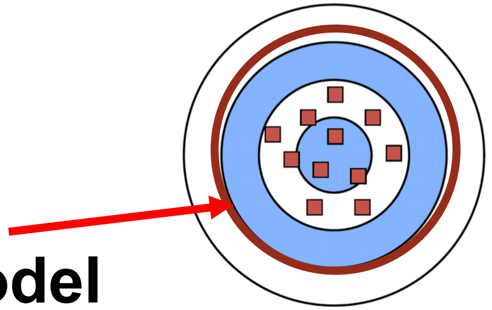
$y = b + w_1 \cdot x_{date} + w_2 \cdot (x_{date})^2 + w_3 \cdot (x_{date})^3 + w_4 \cdot (x_{date})^4$





Large Bias

model

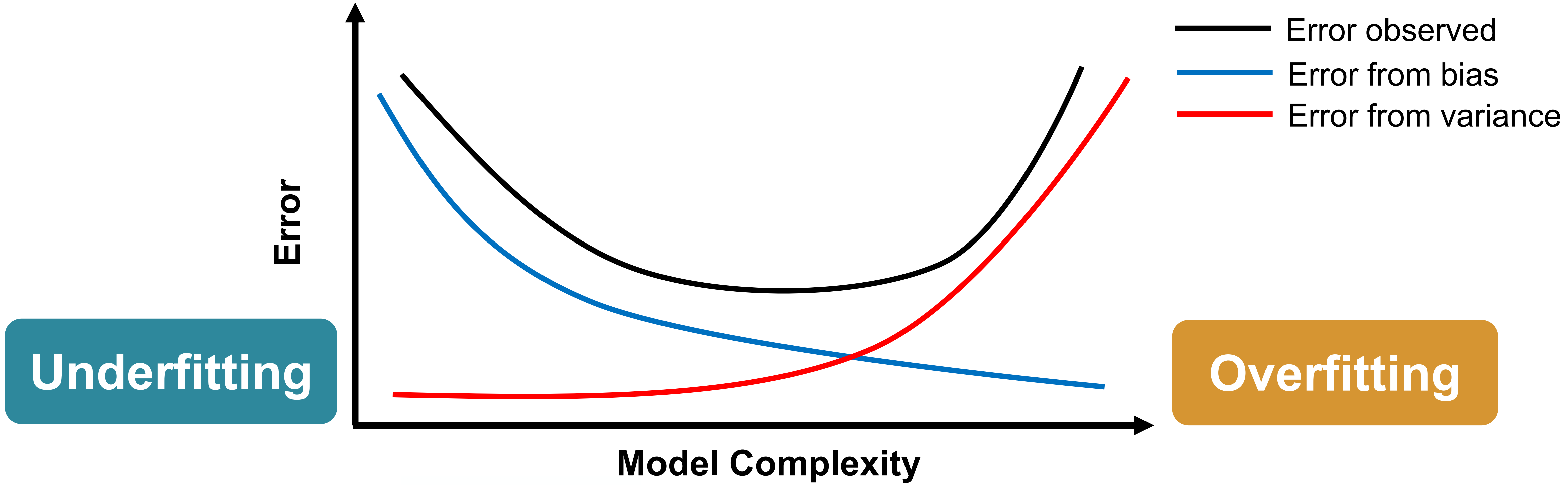


model

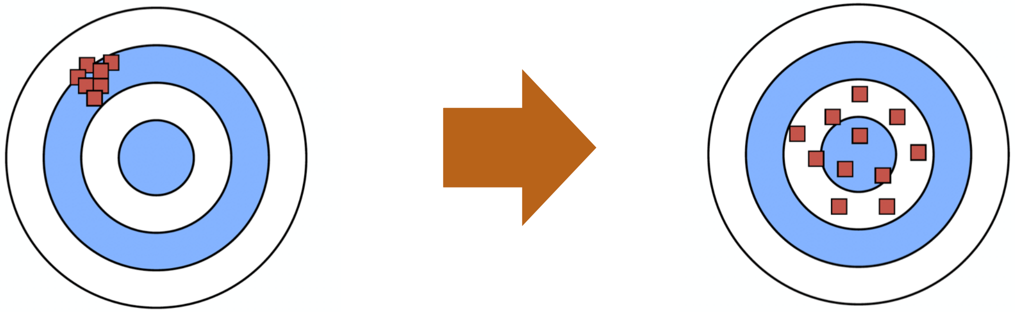
Small Bias

2. Bias and Variance

2.2 Bias V.S. Variance



Large Bias
Small Variance



Small Bias
Large Variance

2. Bias and Variance

2.2 Bias V.S. Variance

- If your model cannot even fit the training samples, then you may have large bias;
- If your model can fit the training data, but large error on testing data, then you probably have large variance.

High Bias $\longleftrightarrow$ Underfitting

High Variance $\longleftrightarrow$ Overfitting

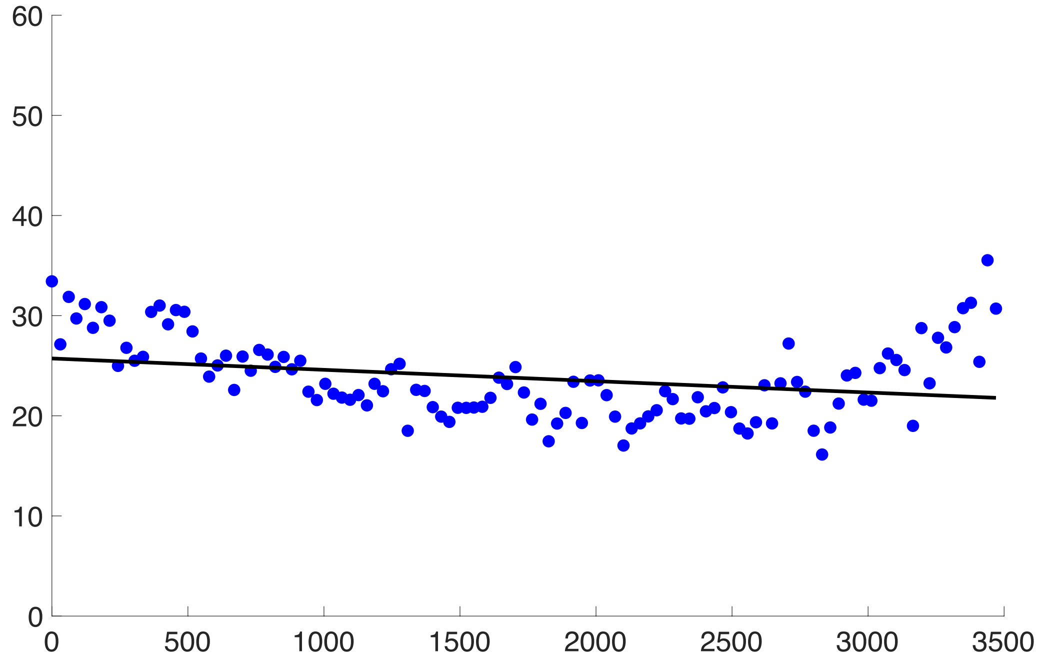
Large σ^2 $\longleftrightarrow$ Noise data

2. Bias and Variance

2.2 Bias V.S. Variance

What to do with large bias?

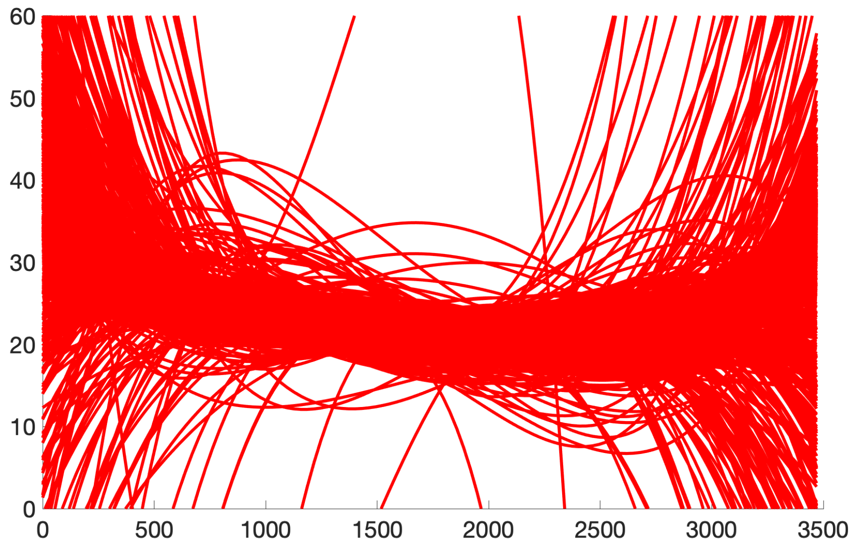
- Add more features as input
- Choose a more complex model



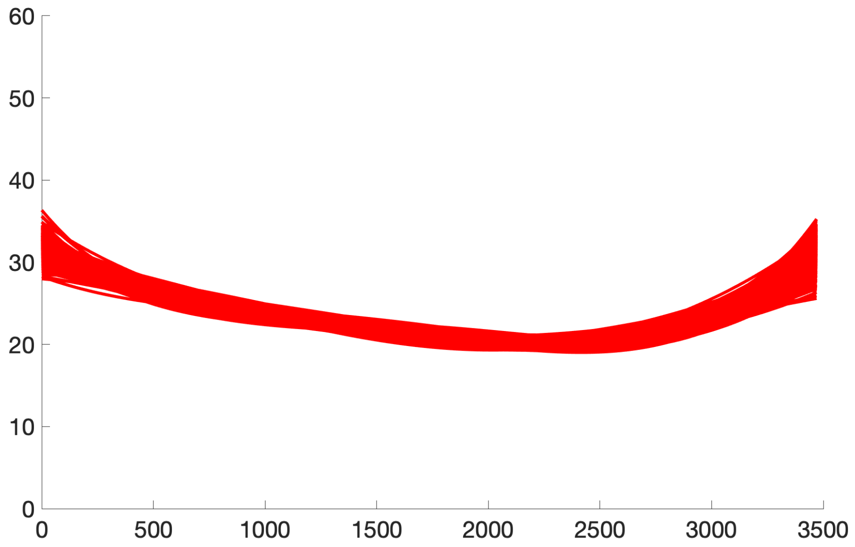
2.2 Bias V.S. Variance

What to do with large variance?

- Add more data
  - Effective, but not always practical



10 samples



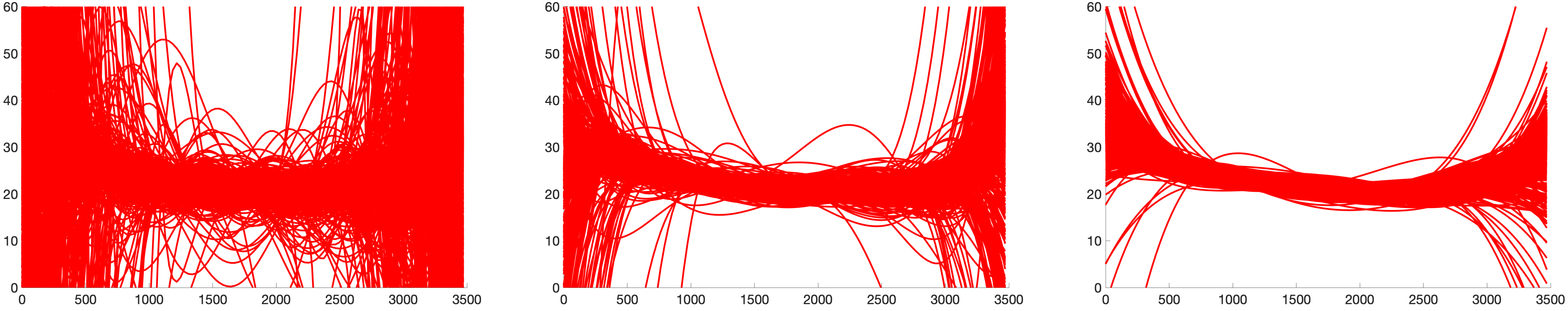
100 samples

2. Bias and Variance

2.2 Bias V.S. Variance

What to do with large variance?

- Regularization

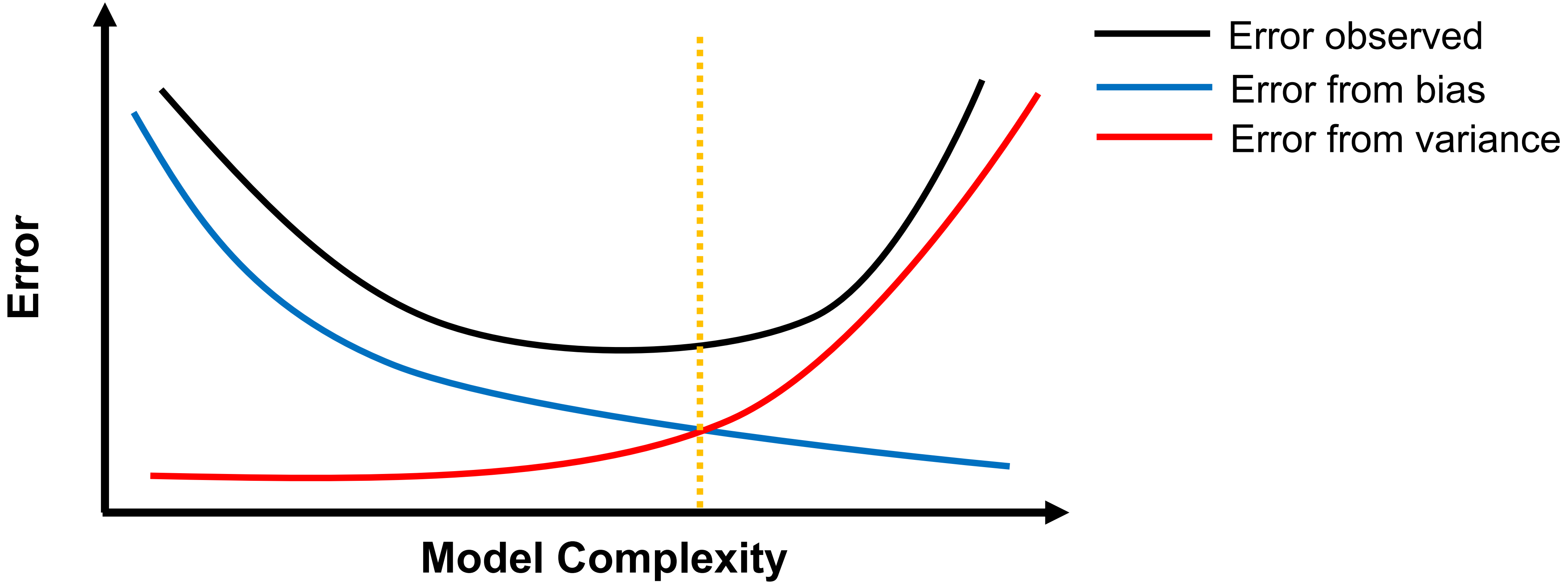


2. Bias and Variance

2.2 Bias V.S. Variance

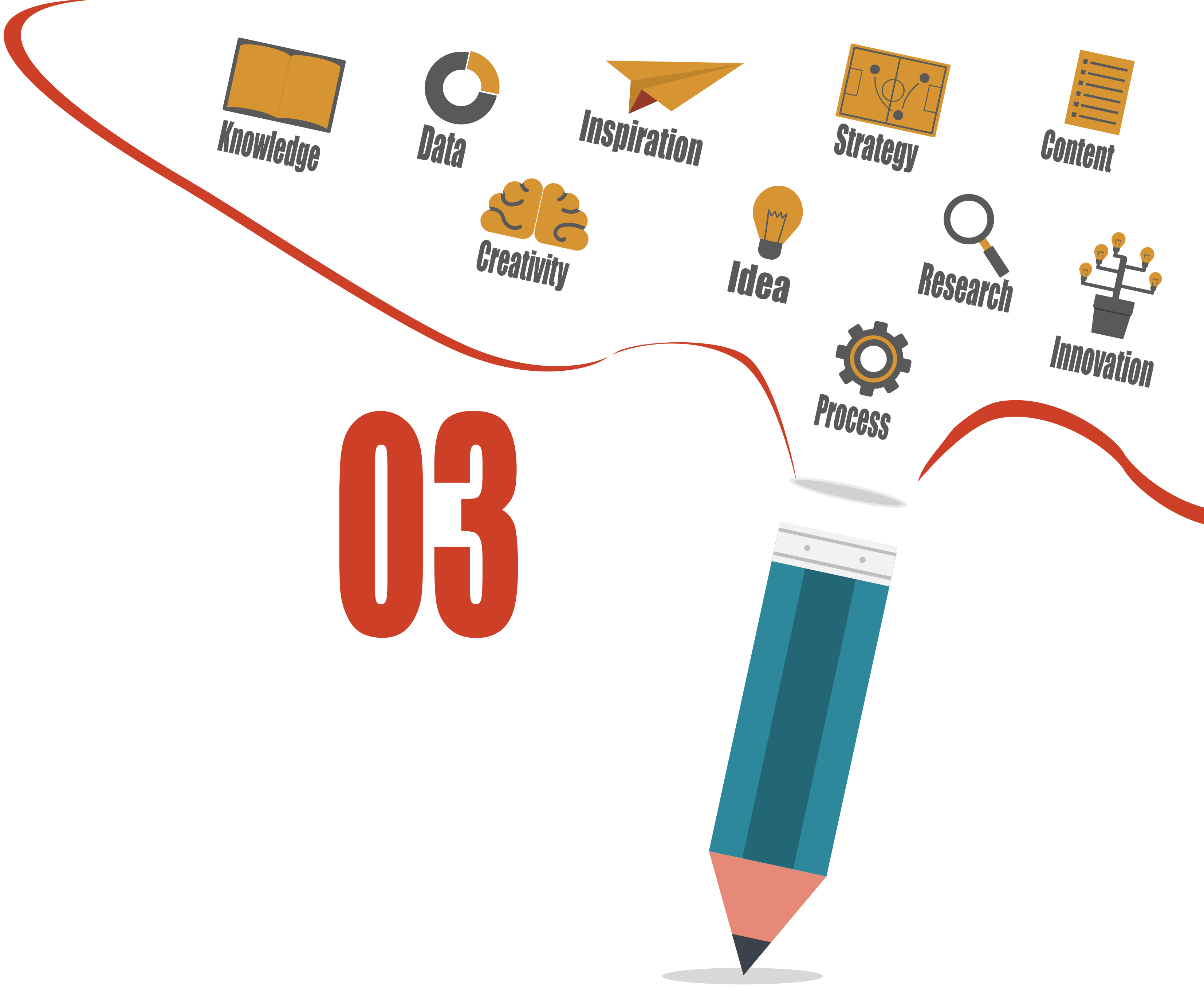
Model Selection

- There is a trade-off between bias and variance
- Select a model that balances two kinds of error to minimize total error



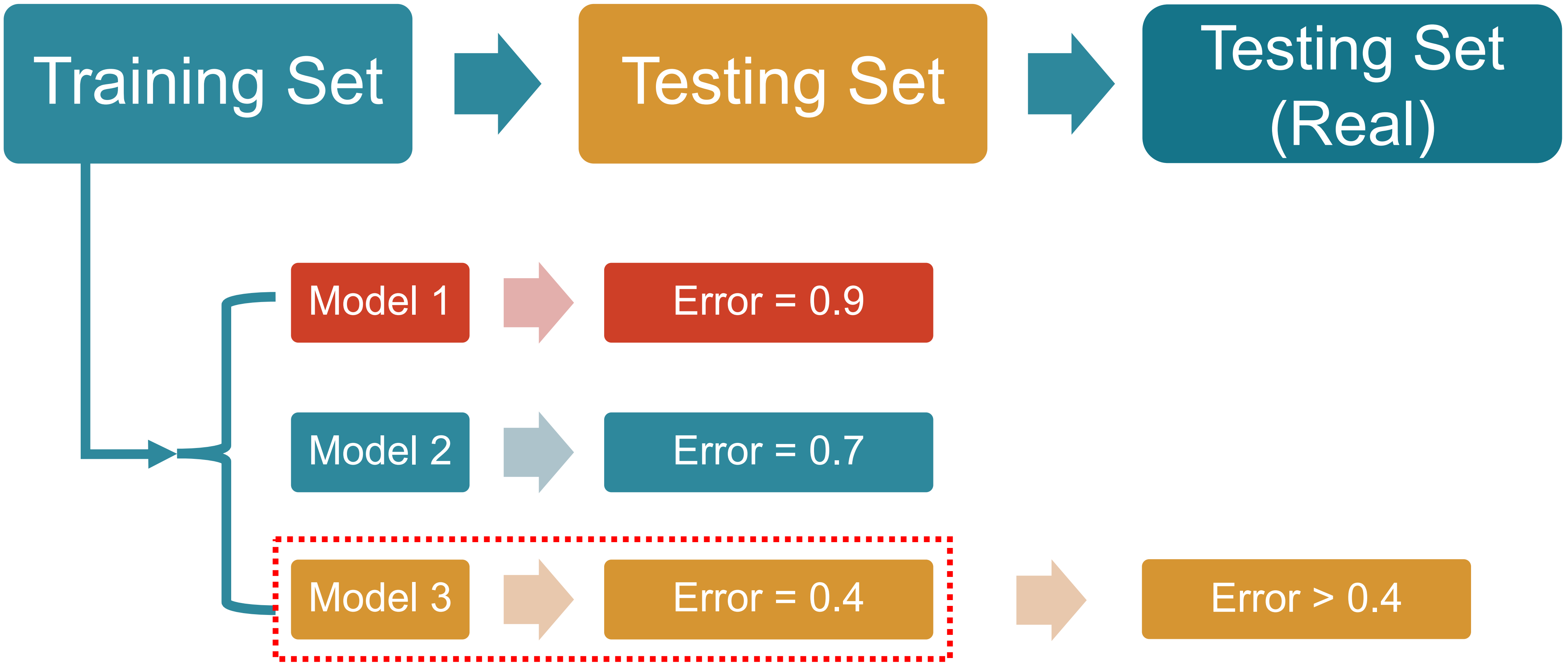
Validation

03



3. Validation

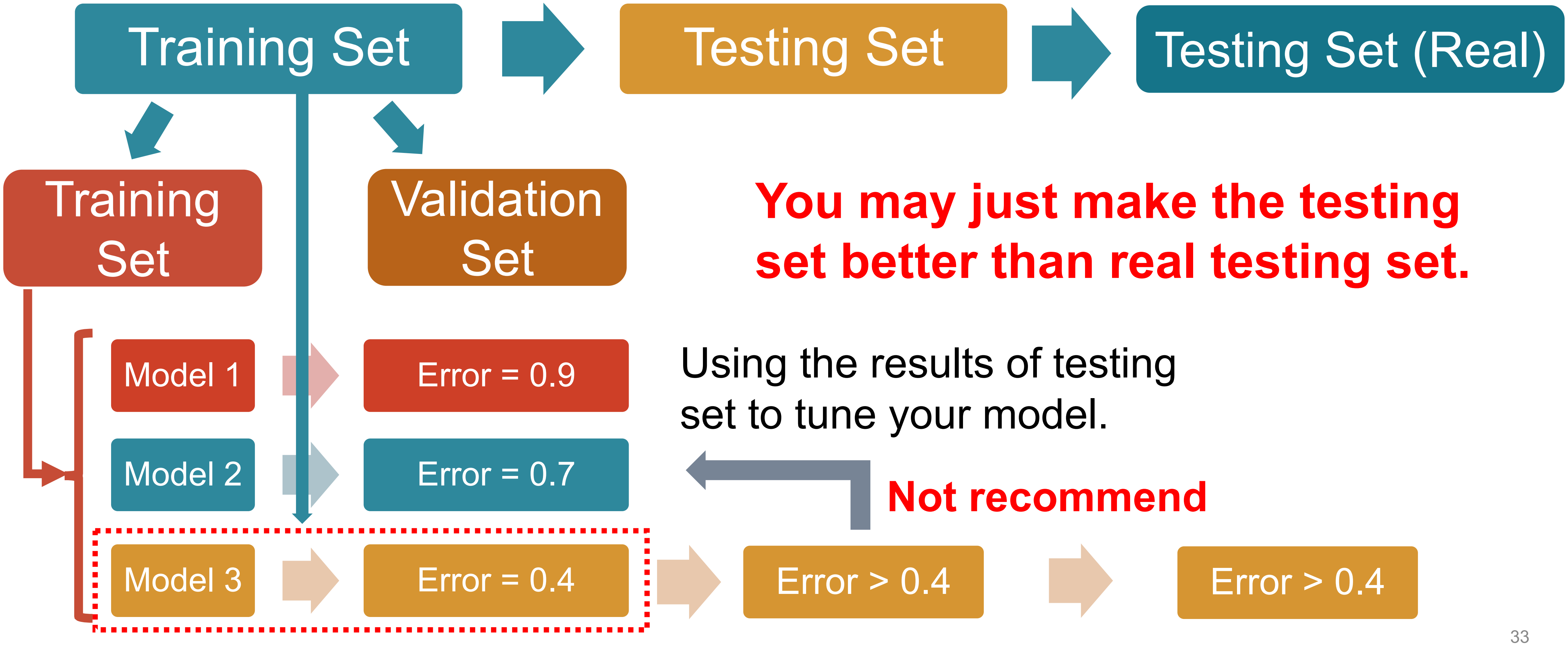
3.1 A problem when you select models



Cite: Hung-yi Lee, Machine Learning

3. Validation

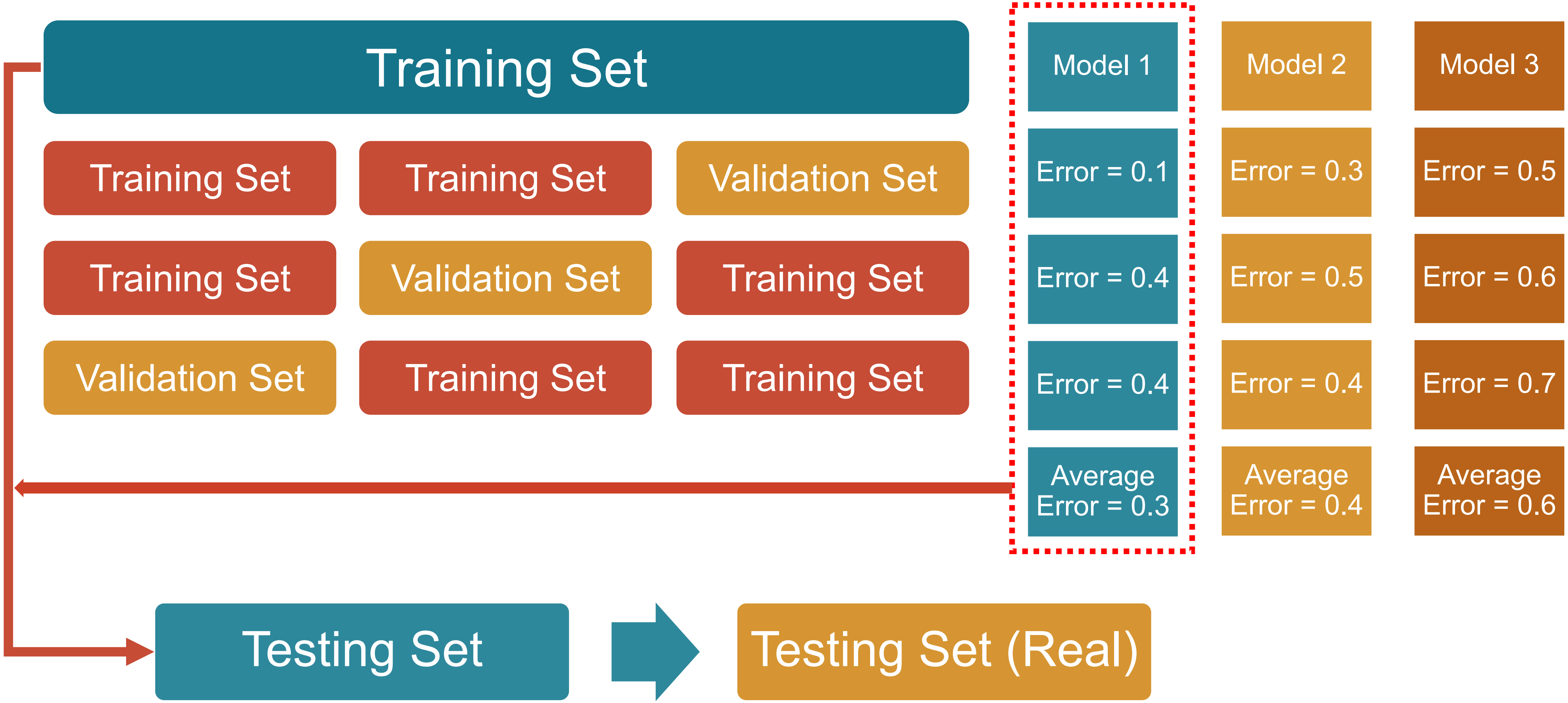
3.2 Cross validation



Cite: Hung-yi Lee, Machine Learning

3. Validation

3.2 K-fold cross validation



Cite: Hung-yi Lee, Machine Learning

Where does the error come from ?

Hao, Qi

School of Astronomy and Space Science

THANKS