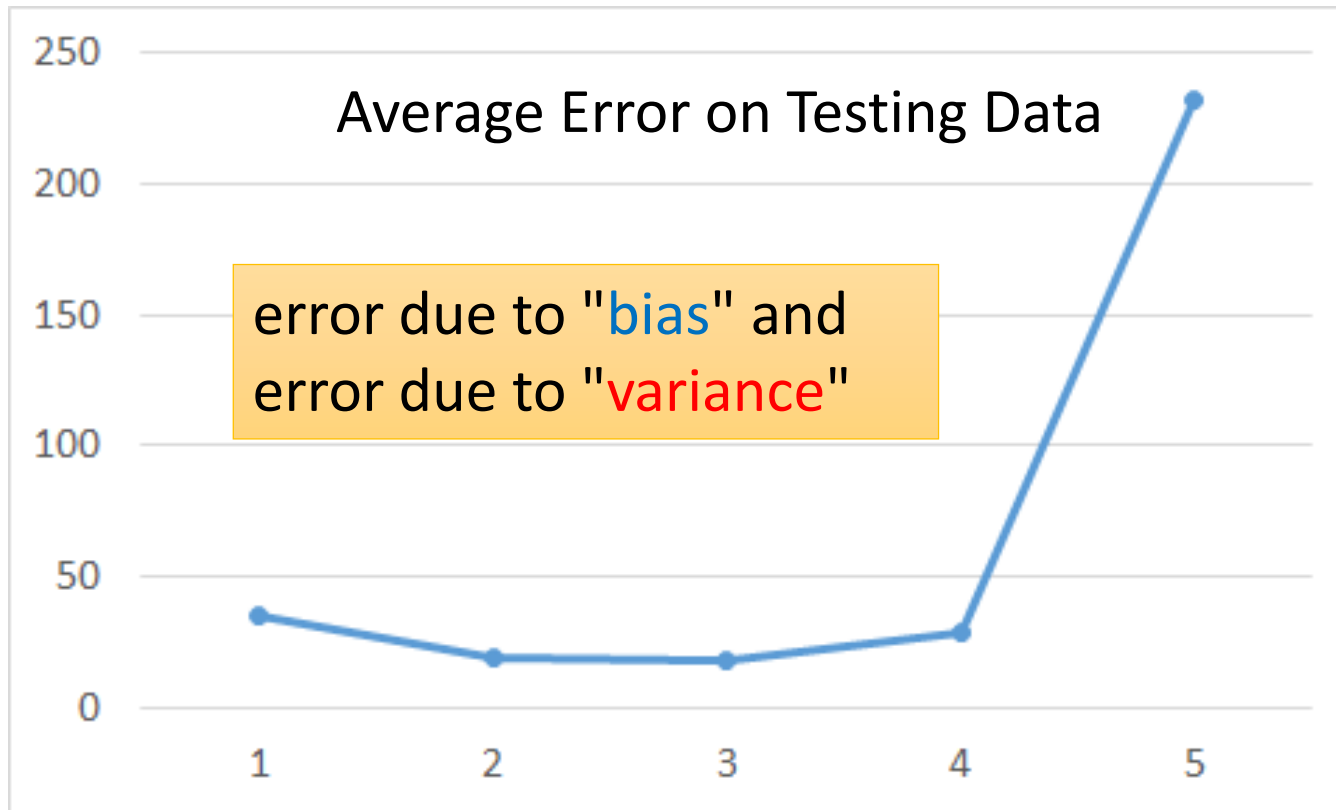


Where does the error  
come from?

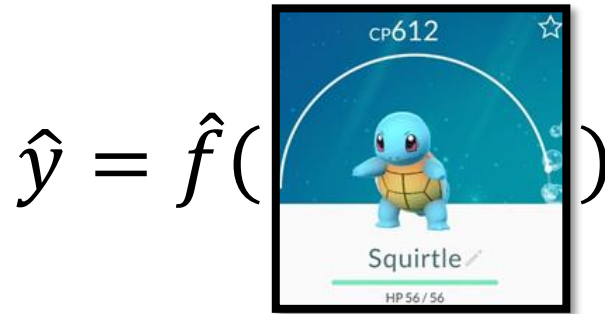
# Review

知道誤差來源才能有效地修改模型



A more complex model does not always lead to better performance on testing data.

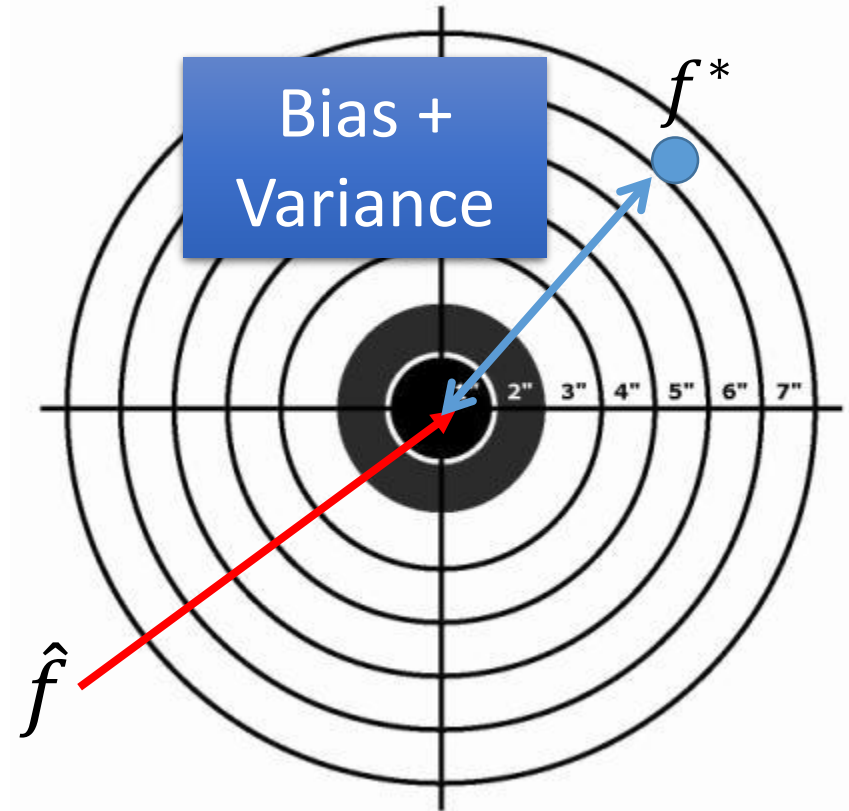
# Estimator



Only Niantic knows  $\hat{f}$

From training data,  
we find  $f^*$

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



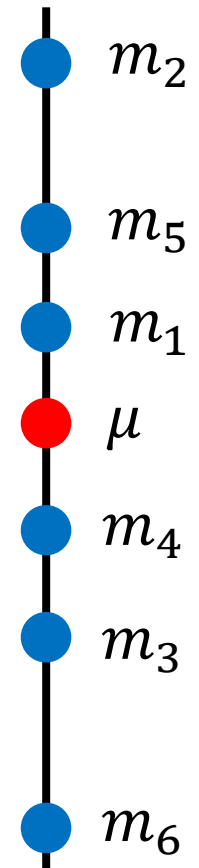
# Bias and Variance of Estimator

- Estimate the mean of a variable  $x$ 
  - assume the mean of  $x$  is  $\mu$
  - assume the variance of  $x$  is  $\sigma^2$
- Estimator of mean  $\mu$ 
  - Sample  $N$  points:  $\{x^1, x^2, \dots, x^N\}$

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

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

unbiased



# Bias and Variance of Estimator

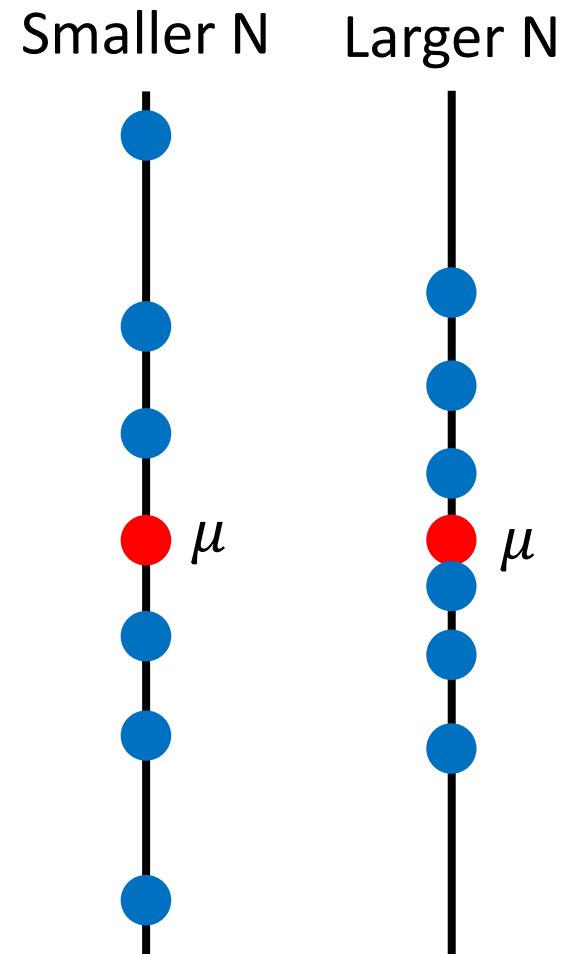
- Estimate the mean of a variable  $x$ 
  - assume the mean of  $x$  is  $\mu$
  - assume the variance of  $x$  is  $\sigma^2$
- Estimator of mean  $\mu$ 
  - Sample  $N$  points:  $\{x^1, x^2, \dots, x^N\}$

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

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

Variance depends  
on the number of  
samples

unbiased



# Bias and Variance of Estimator

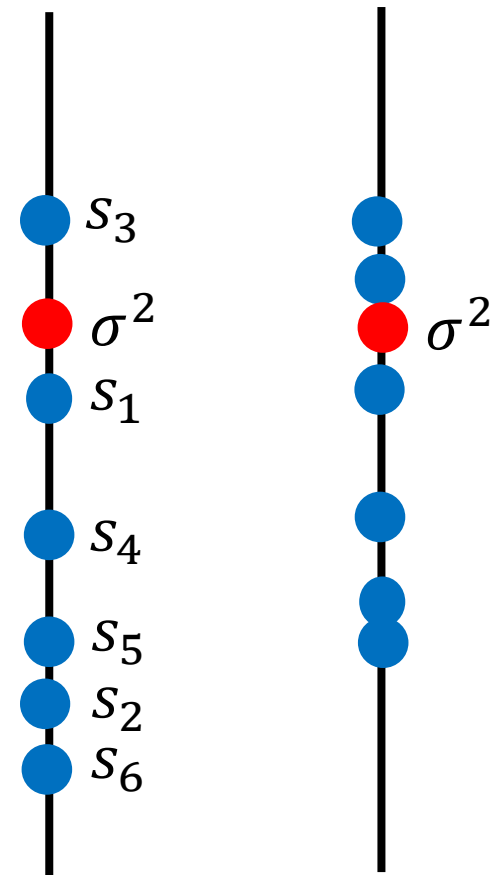
- Estimate the mean of a variable  $x$ 
  - assume the mean of  $x$  is  $\mu$
  - assume the variance of  $x$  is  $\sigma^2$
- Estimator of variance  $\sigma^2$ 
  - Sample  $N$  points:  $\{x^1, x^2, \dots, x^N\}$

$$m = \frac{1}{N} \sum_n x^n \quad s = \frac{1}{N} \sum_n (x^n - m)^2$$

Biased estimator

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

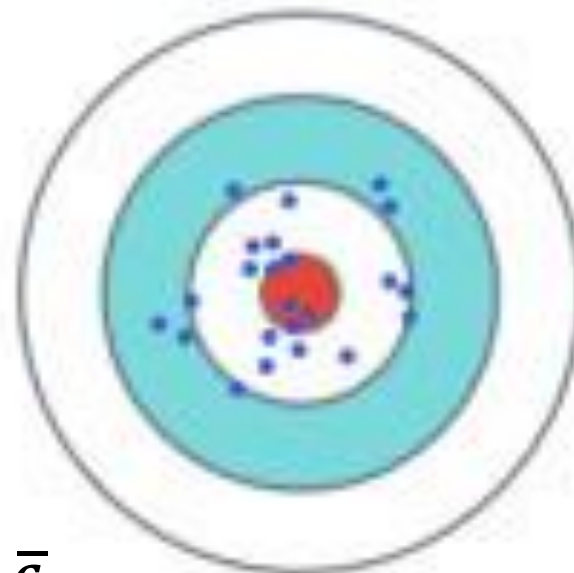
Increase  $N$



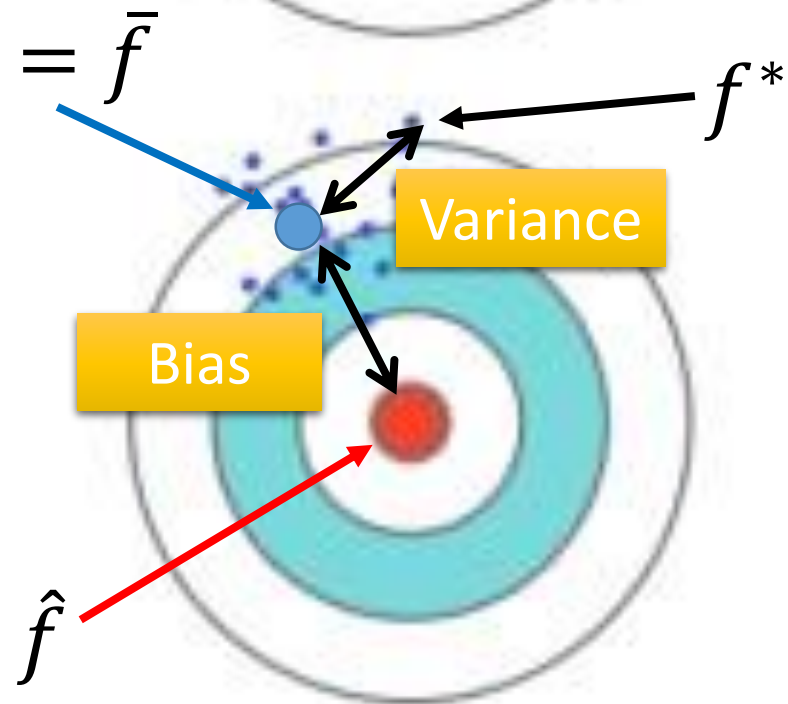
Low Variance

High Variance

Low Bias



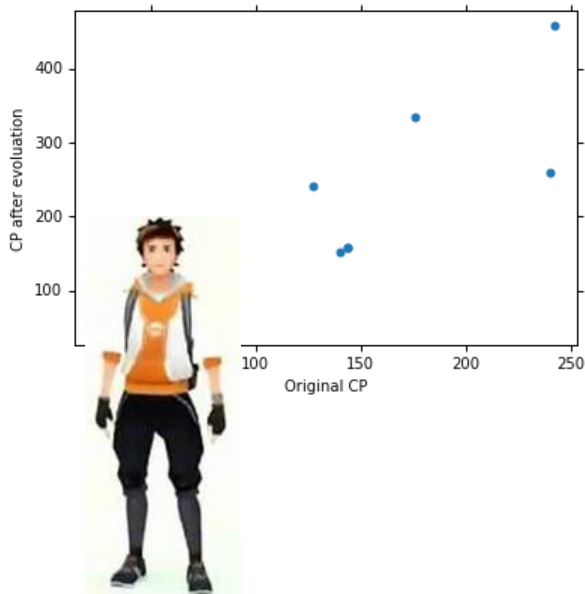
High Bias



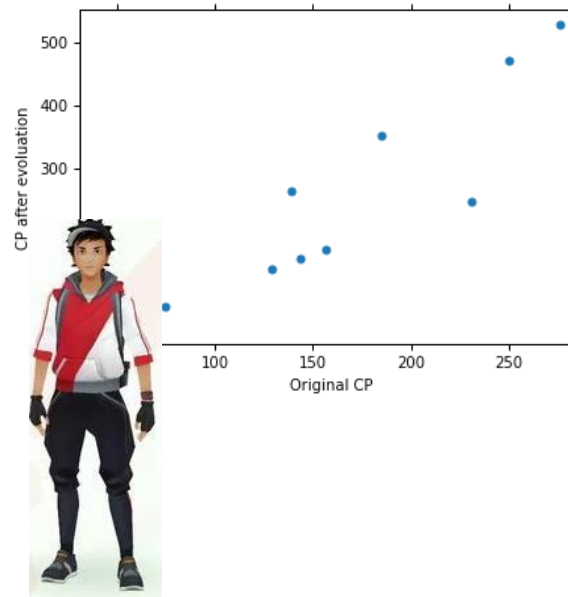
# Parallel Universes

- In all the universes, we are collecting (catching) 10 Pokémon as training data to find  $f^*$

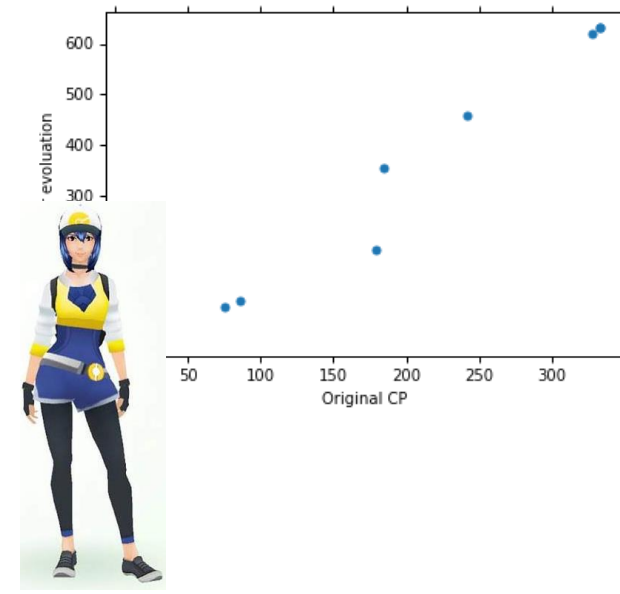
Universe 1



Universe 2



Universe 3

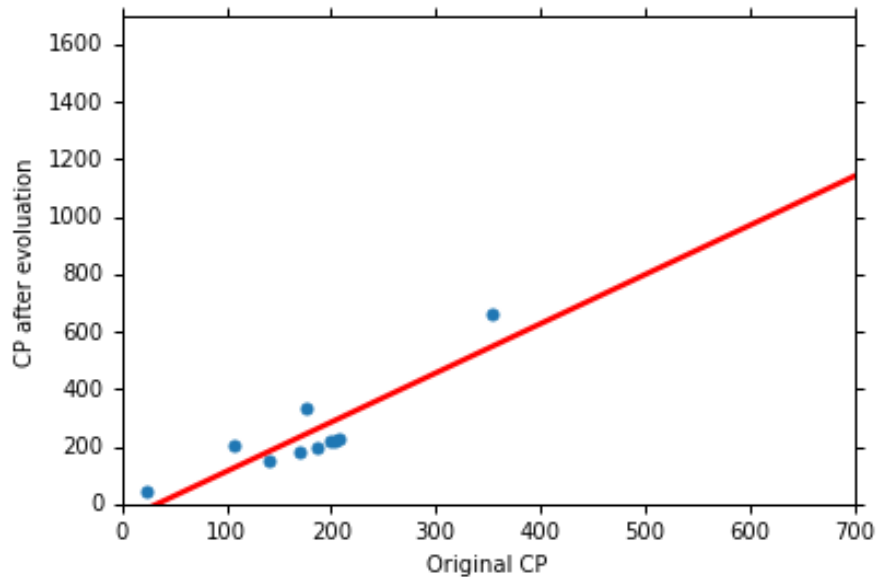




# Parallel Universes

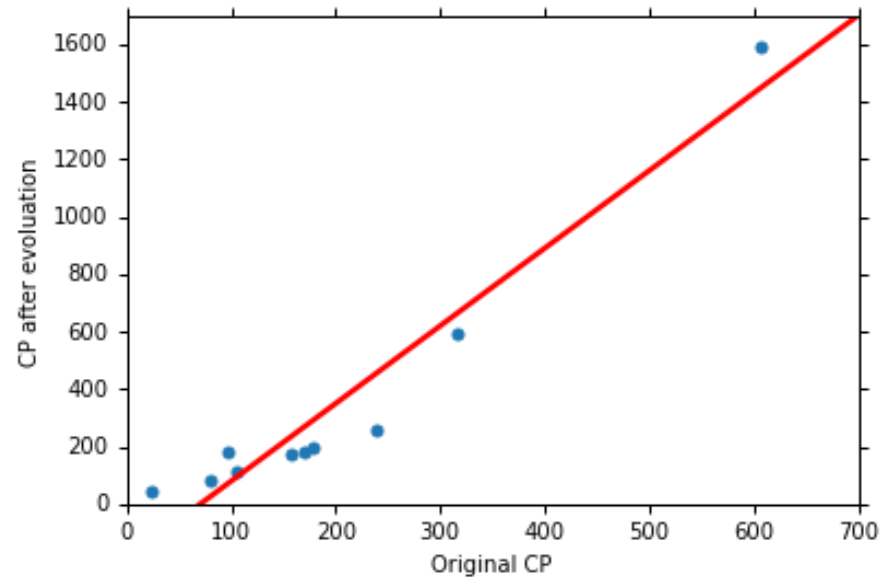
- In different universes, we use the same model, but obtain different  $f^*$

Universe 123



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

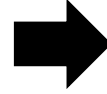
Universe 345



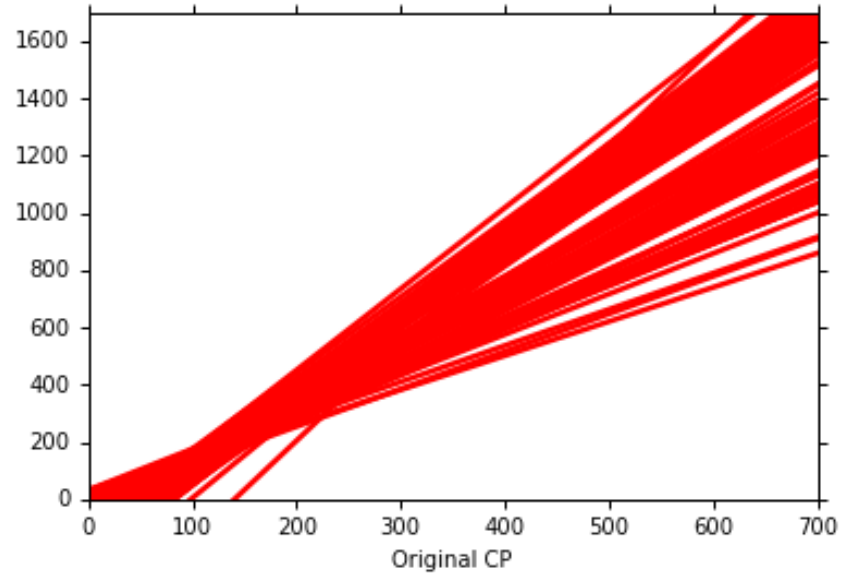
$$y = b + w \cdot x_{cp}$$

# $f^*$ in 100 Universes

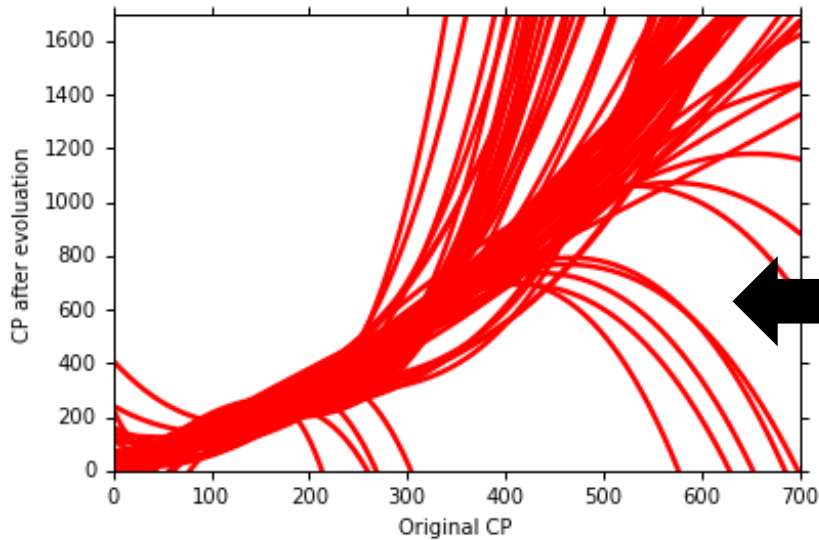
$$y = b + w \cdot x_{cp}$$



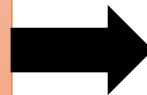
CP after evolution



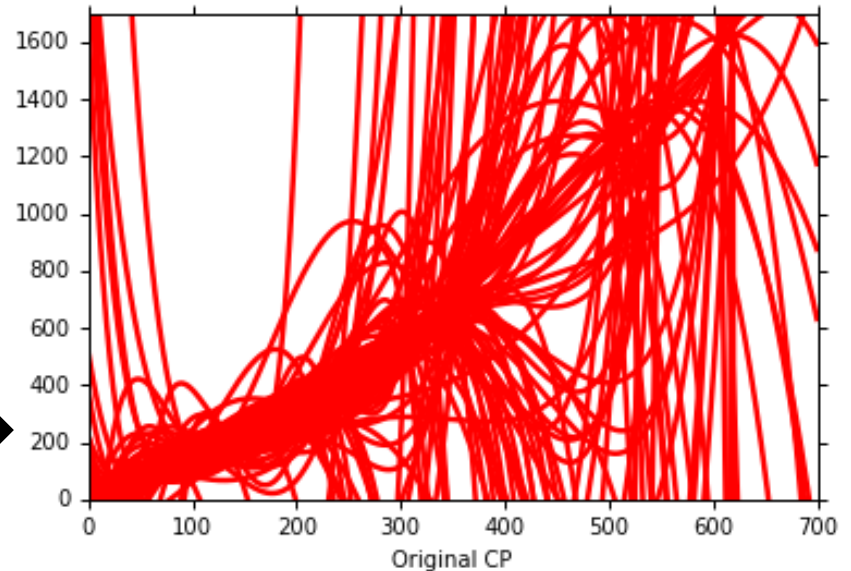
$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3$$



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

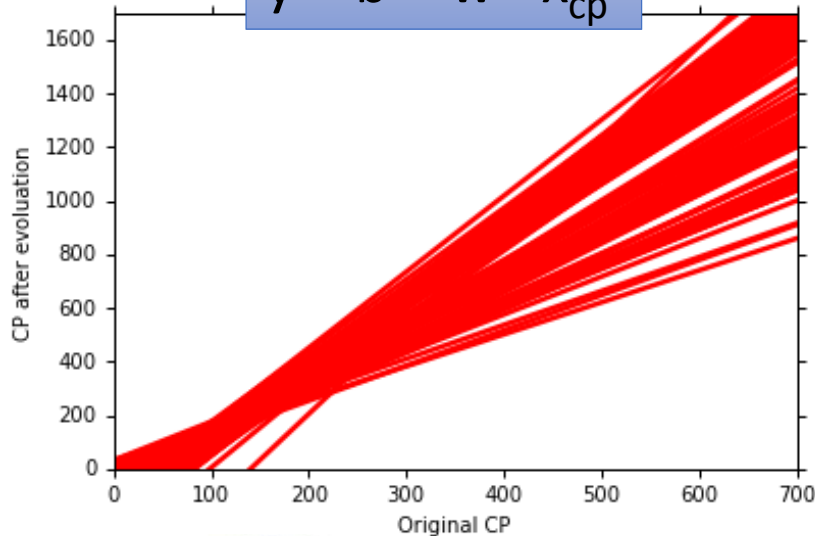


CP after evolution



# Variance

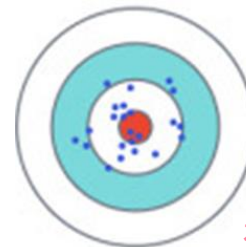
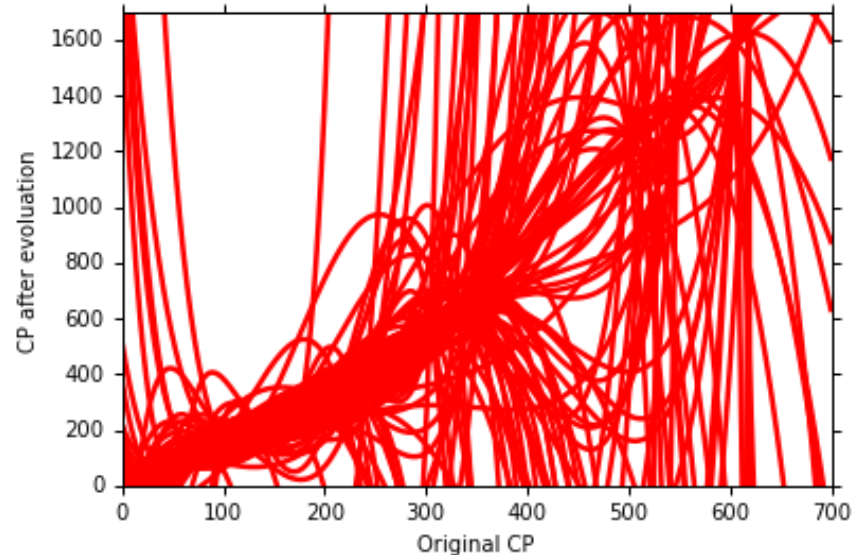
$$y = b + w \cdot x_{cp}$$



Small  
Variance

簡單的模型變異散佈較緊  
因受data影響較少

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



Large  
Variance

複雜的模型變異散佈較開

Simpler model is less influenced by the sampled data

Consider the extreme case  $f(x) = 5$

# Bias

$$E[f^*] = \bar{f}$$

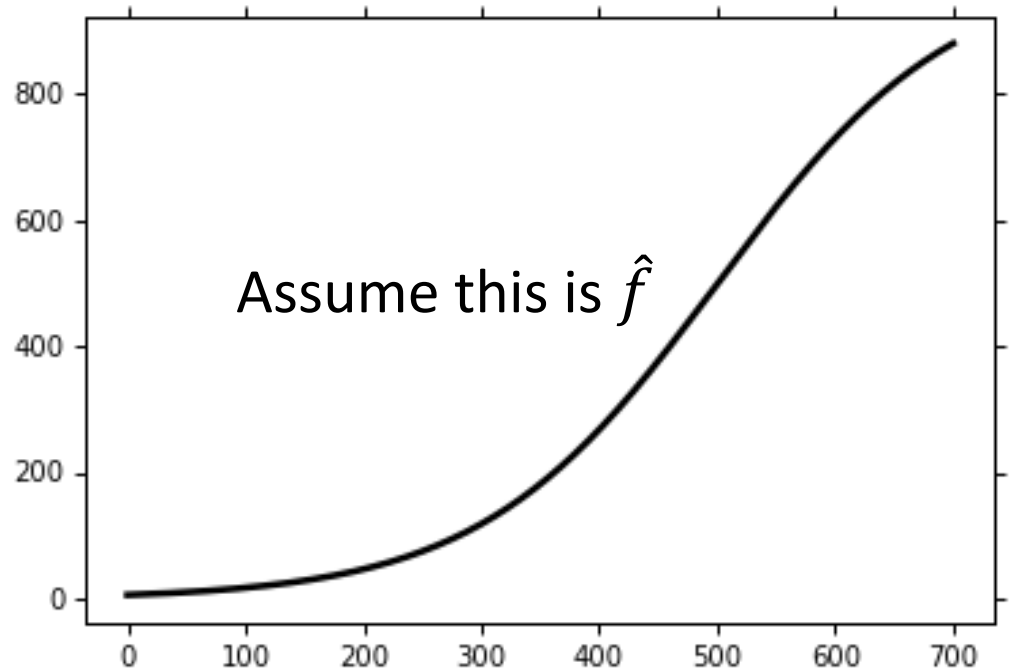
- Bias: If we average all the  $f^*$ , is it close to  $\hat{f}$  ?



Large  
Bias



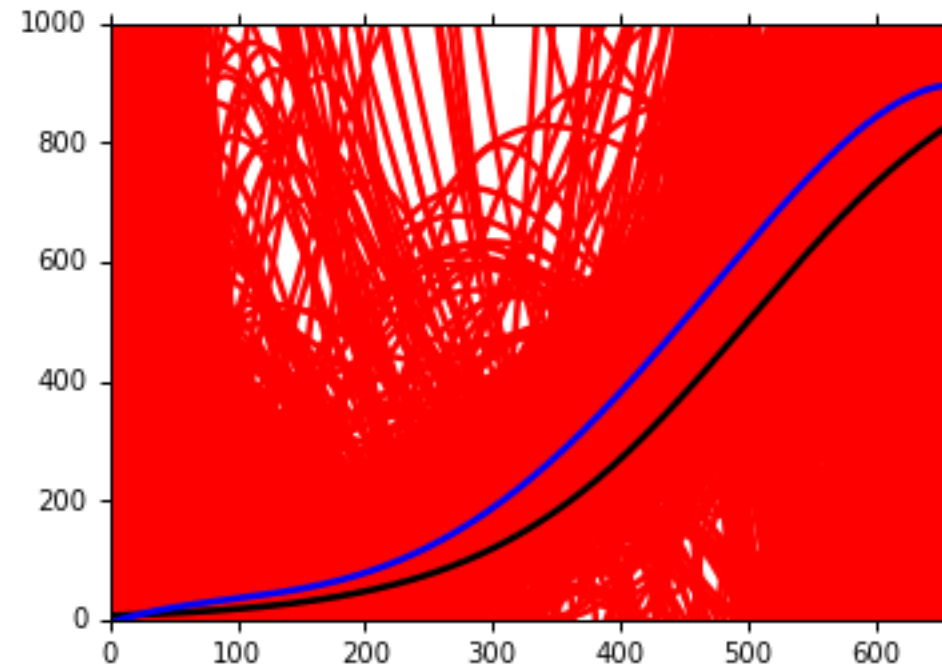
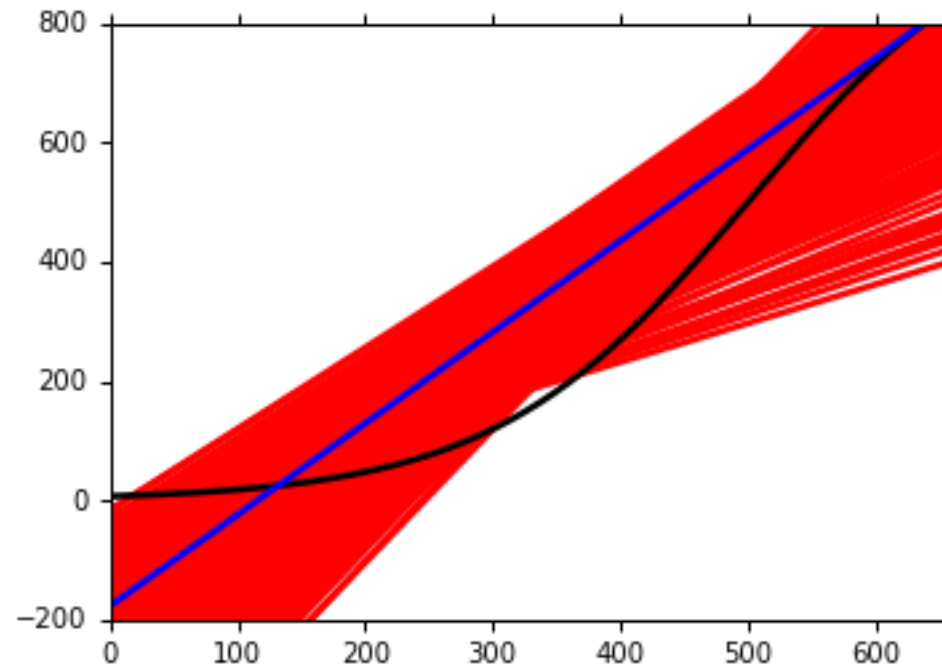
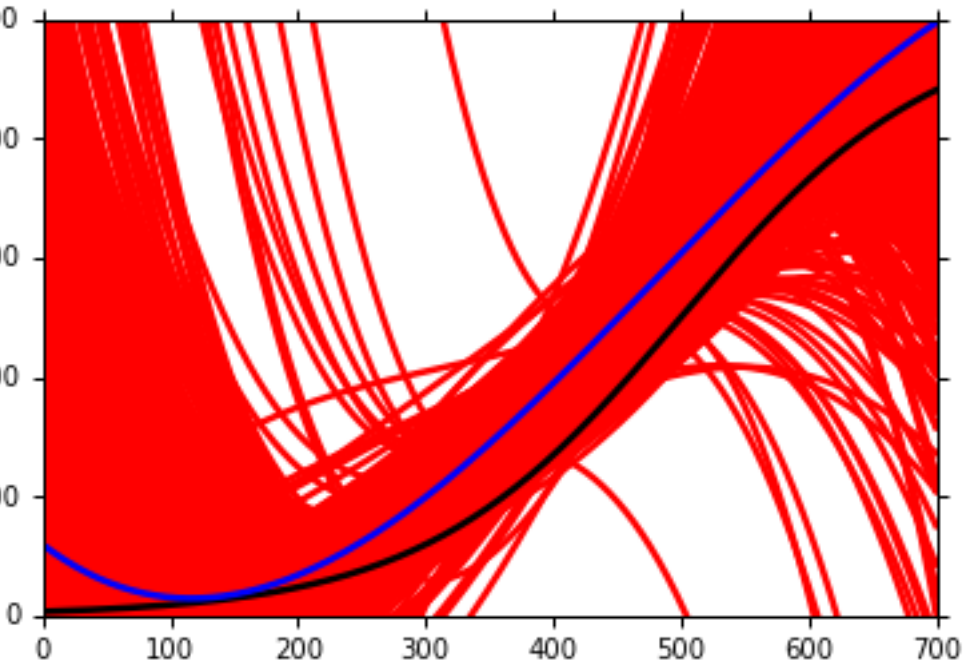
Small  
Bias



Black curve: the true function  $\hat{f}$

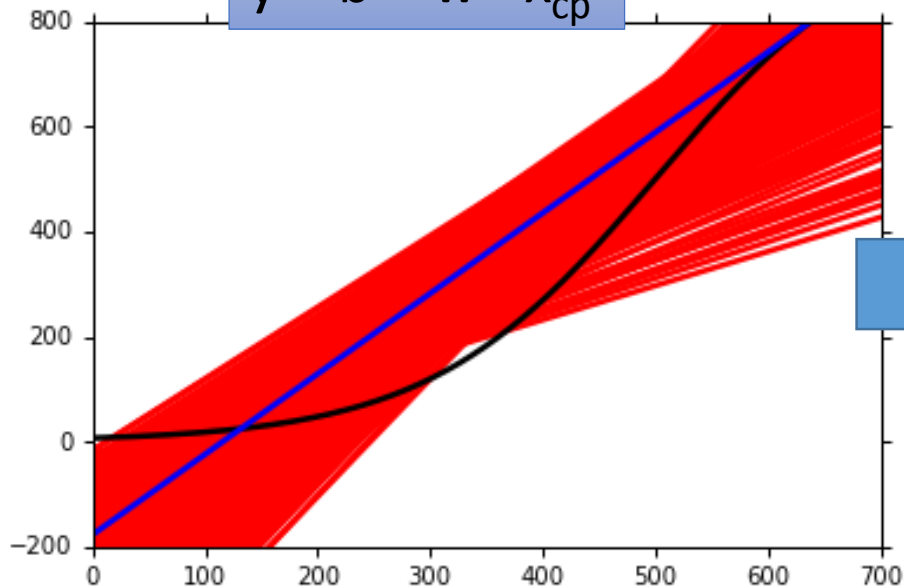
Red curves: 5000  $f^*$

Blue curve: the average of 5000  $f^*$   
 $= \bar{f}$

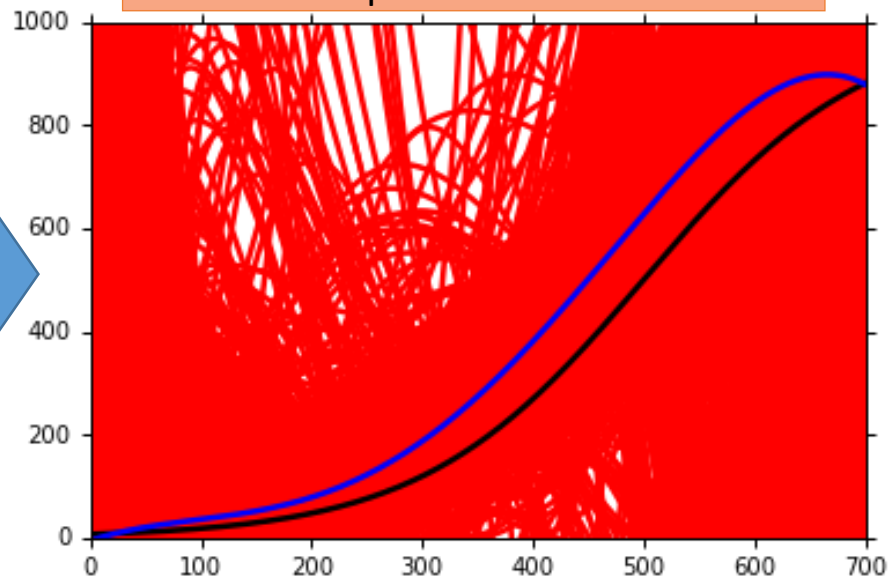


# Bias

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



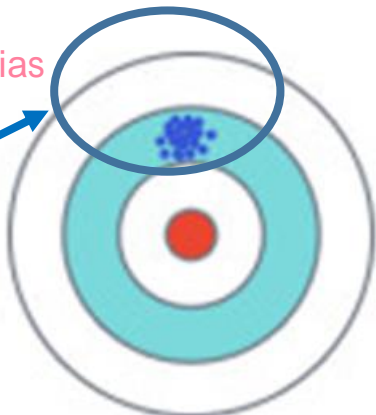
$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4 + w_5 \cdot (x_{cp})^5$$



簡單的模型有大bias

model

function set 的 space 比較小，可能未包含 target

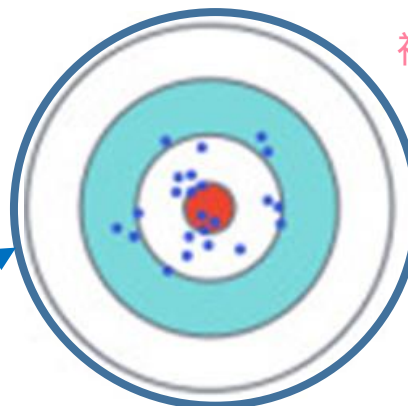


Large Bias

複雜的模型有小bias

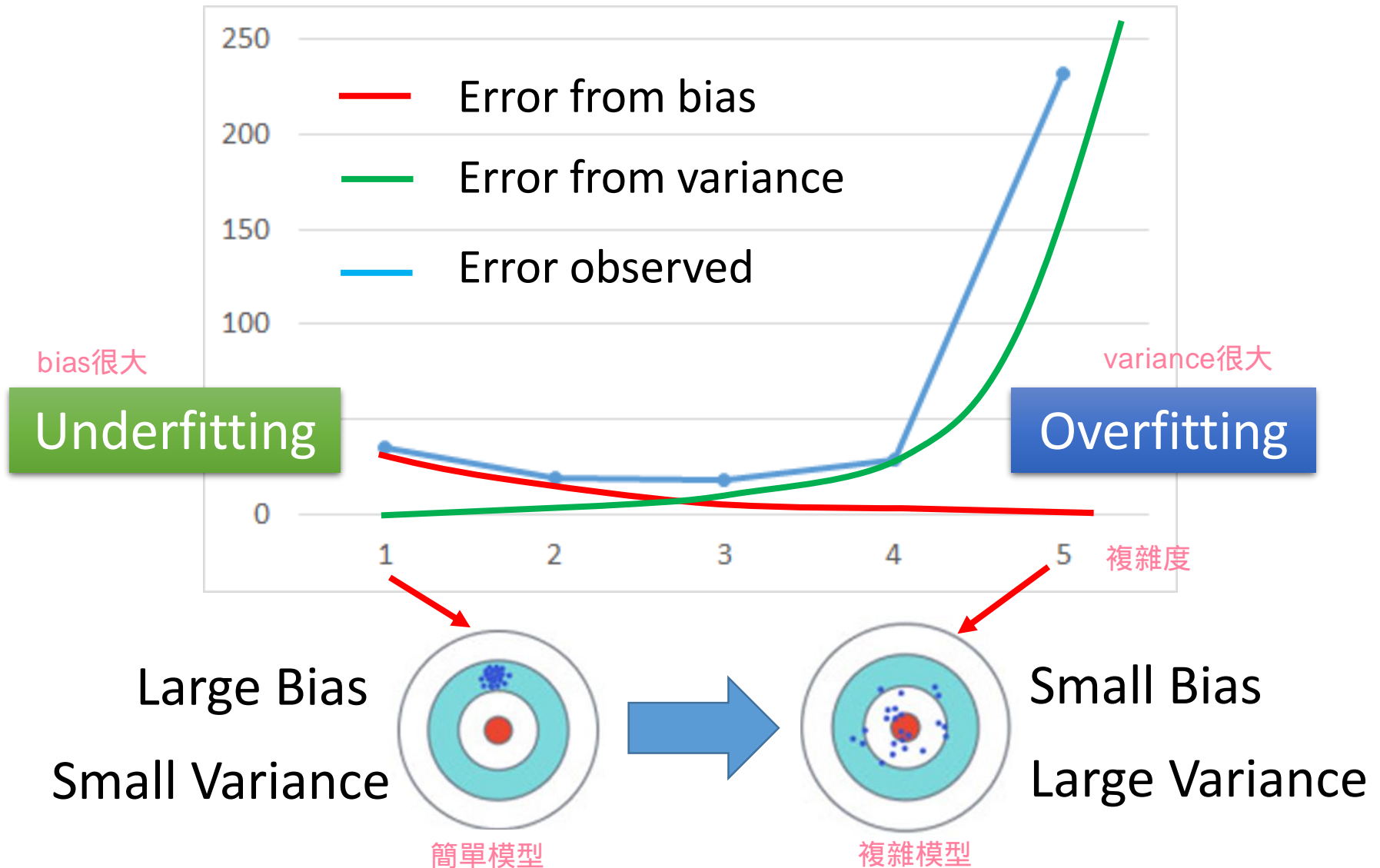
model

function set 的 space 比較大，可能包含 target



Small Bias

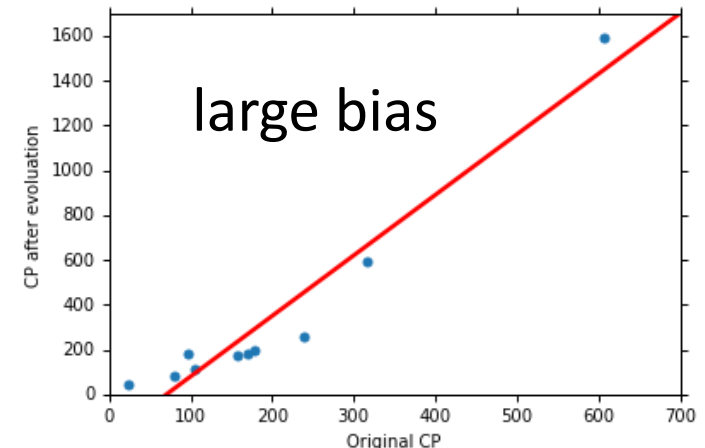
# Bias v.s. Variance



# What to do with large bias?

- Diagnosis:
  - If your model cannot even fit the training examples, then you have large bias Underfitting
  - If you can fit the training data, but large error on testing data, then you probably have large variance Overfitting
- For bias, redesign your model:
  - Add more features as input
  - A more complex model

因model本身就on不好



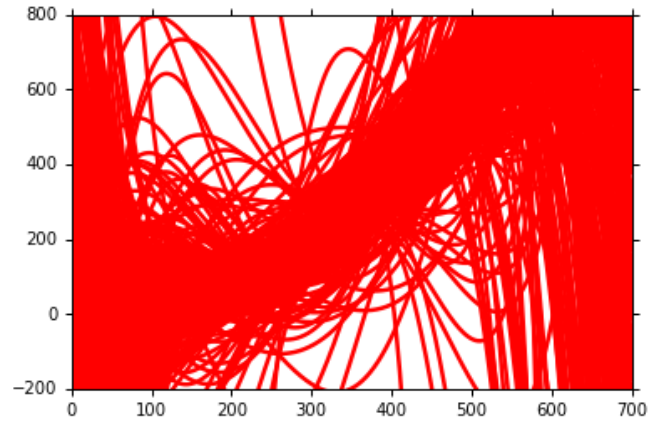


# What to do with large variance?

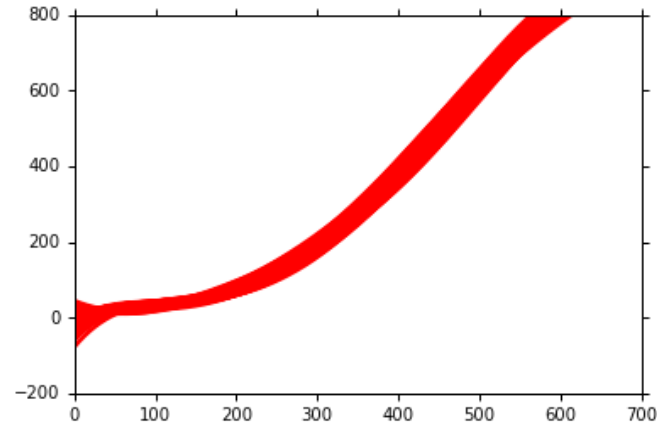
蒐集更多資料

- More data

Very effective,  
but not always  
practical



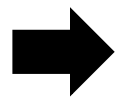
10 examples



100 examples

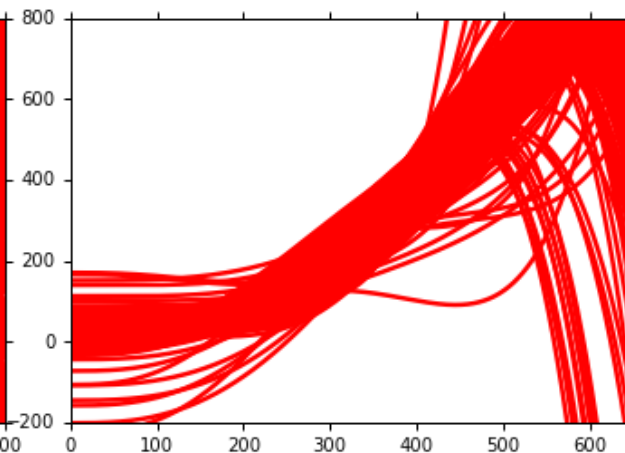
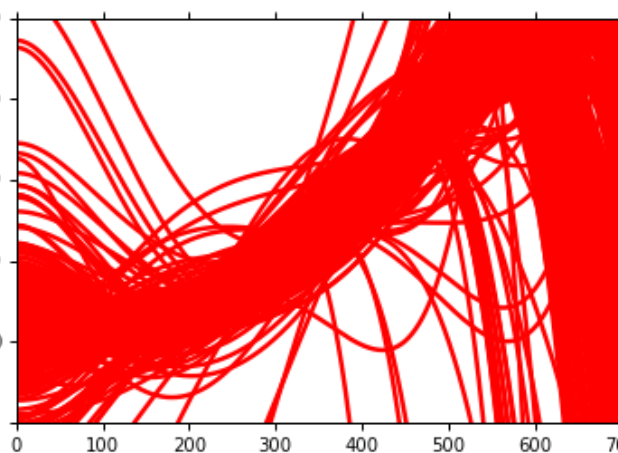
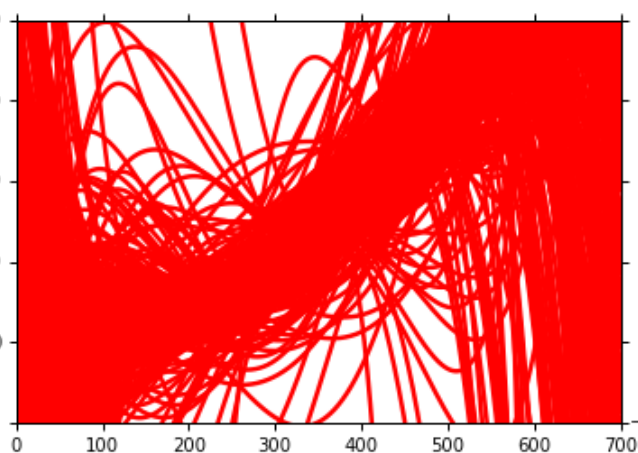
若實際執行上無法蒐集更多資料，  
可以->

- Regularization



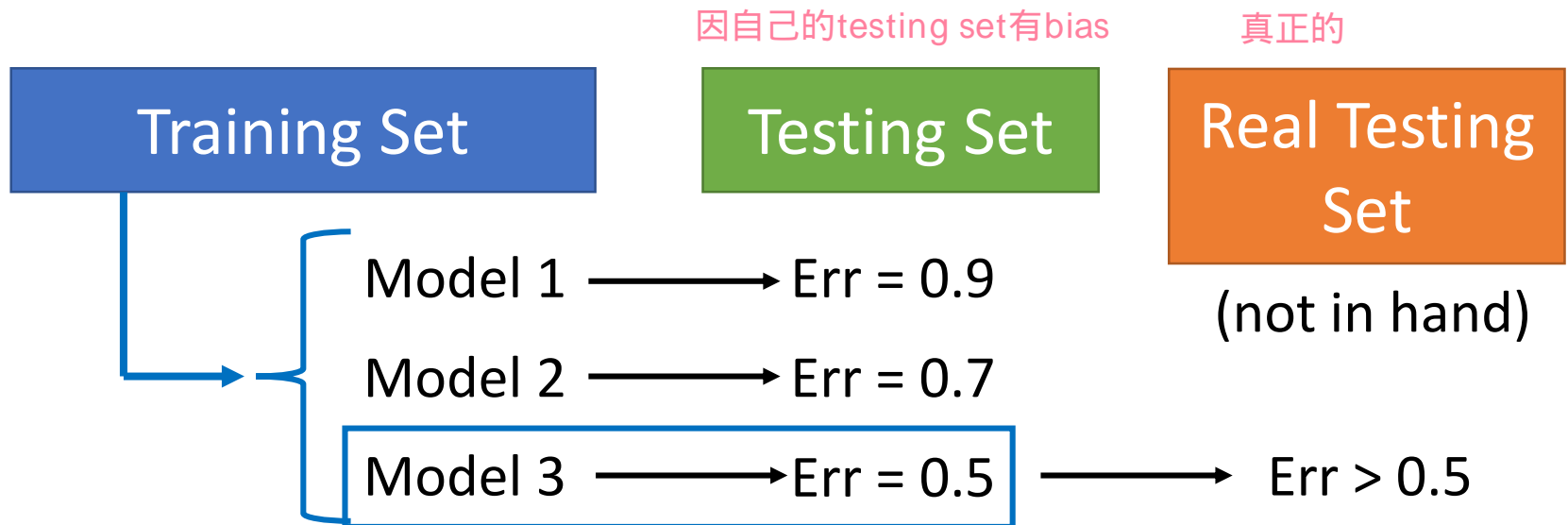
May increase bias

(雖然變異變小，但誤差可能變大，  
故要取得平衡)



# Model Selection

- There is usually a trade-off between bias and variance.
- Select a model that balances two kinds of error to minimize total error
- What you should NOT do:



# Homework 舉kaggle的例子

public

private

Training Set

Testing Set

Testing Set

Model 1  $\longrightarrow$  Err = 0.9

Model 2  $\longrightarrow$  Err = 0.7

Model 3  $\longrightarrow$  Err = 0.5  $\longrightarrow$  Err > 0.5

I beat baseline!

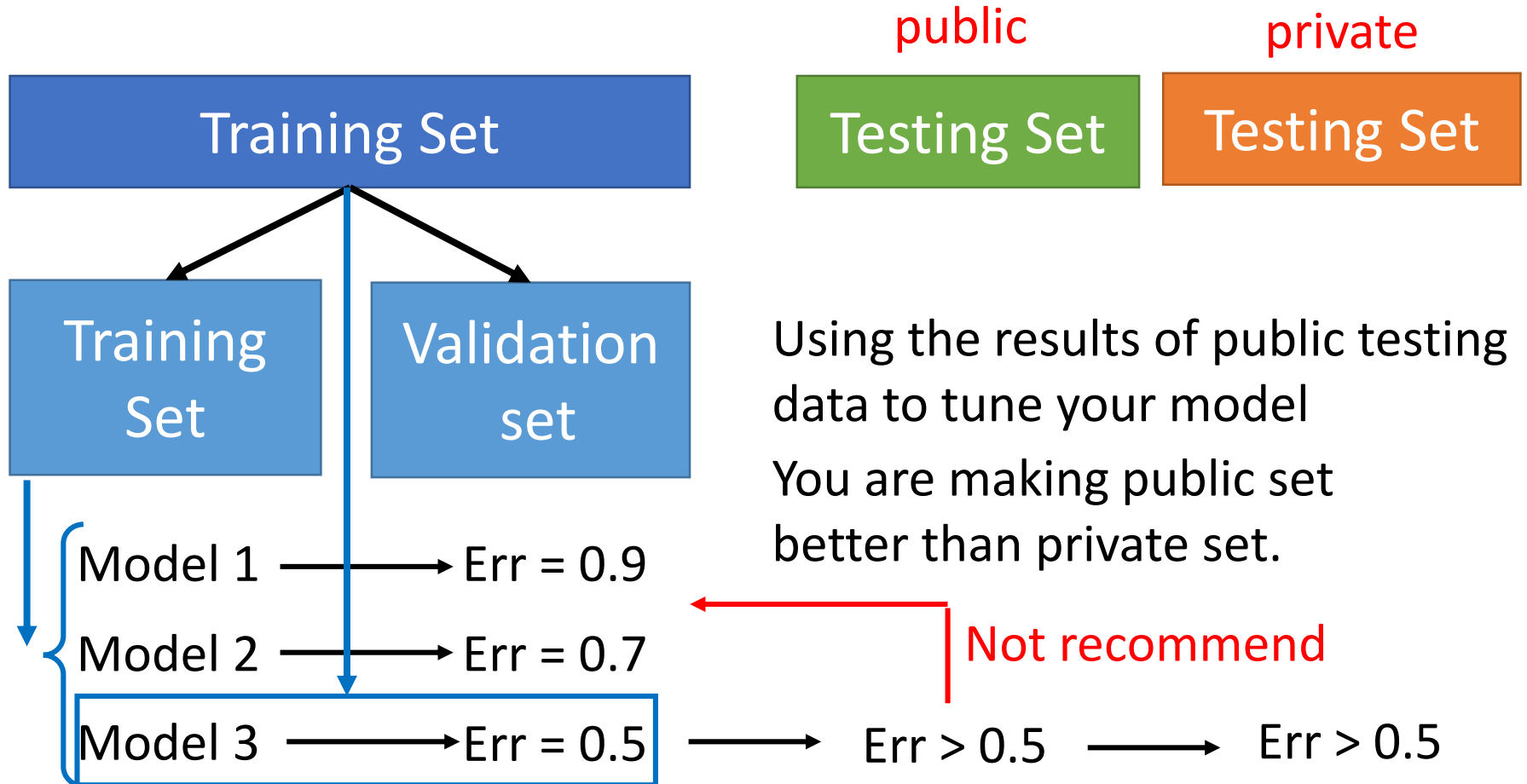
No, you don't

What will happen?

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

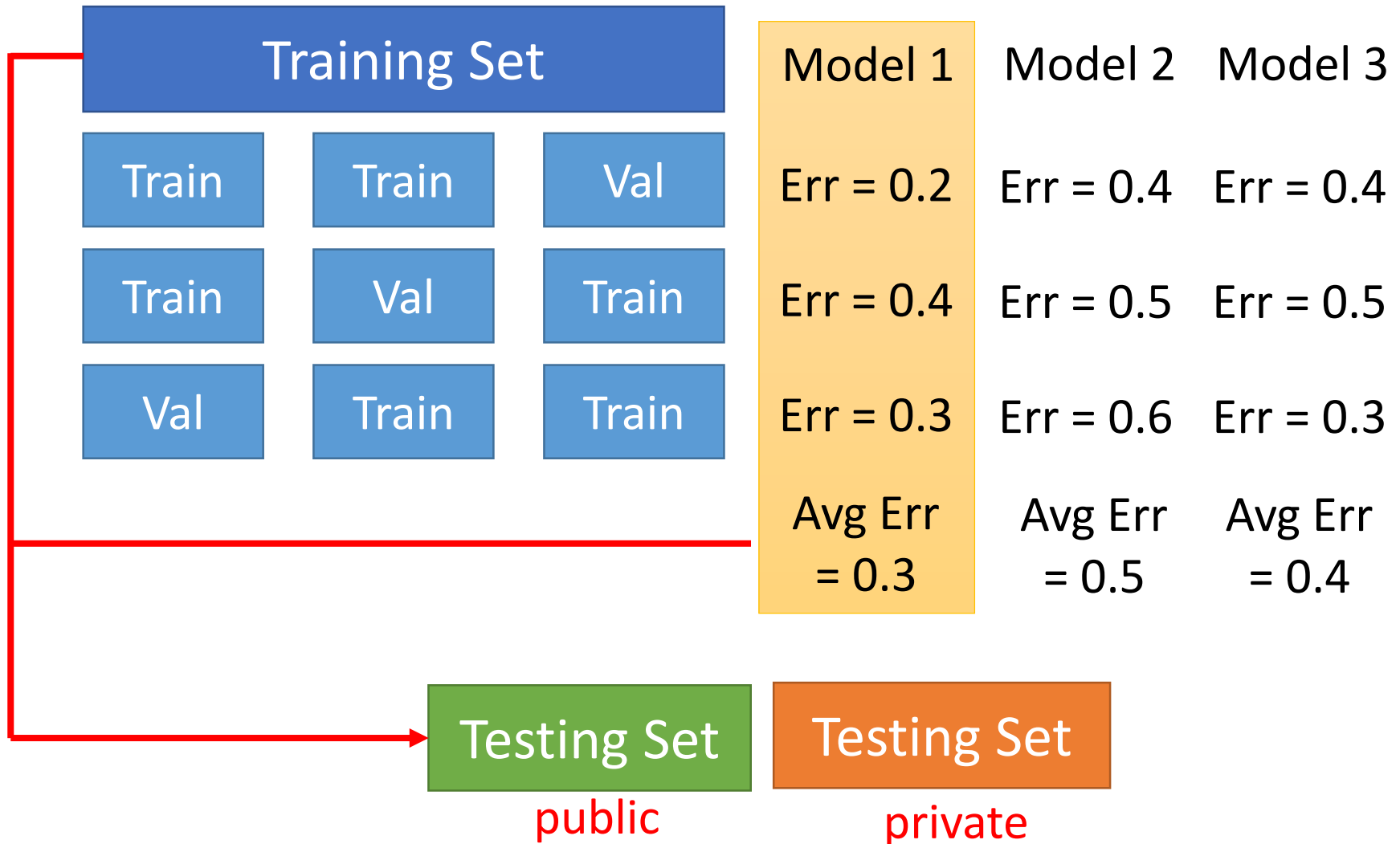


# Cross Validation



# N-fold Cross Validation

分很多種不同的樣子



# Reference

- Bishop: Chapter 3.2