Where does the error come from?
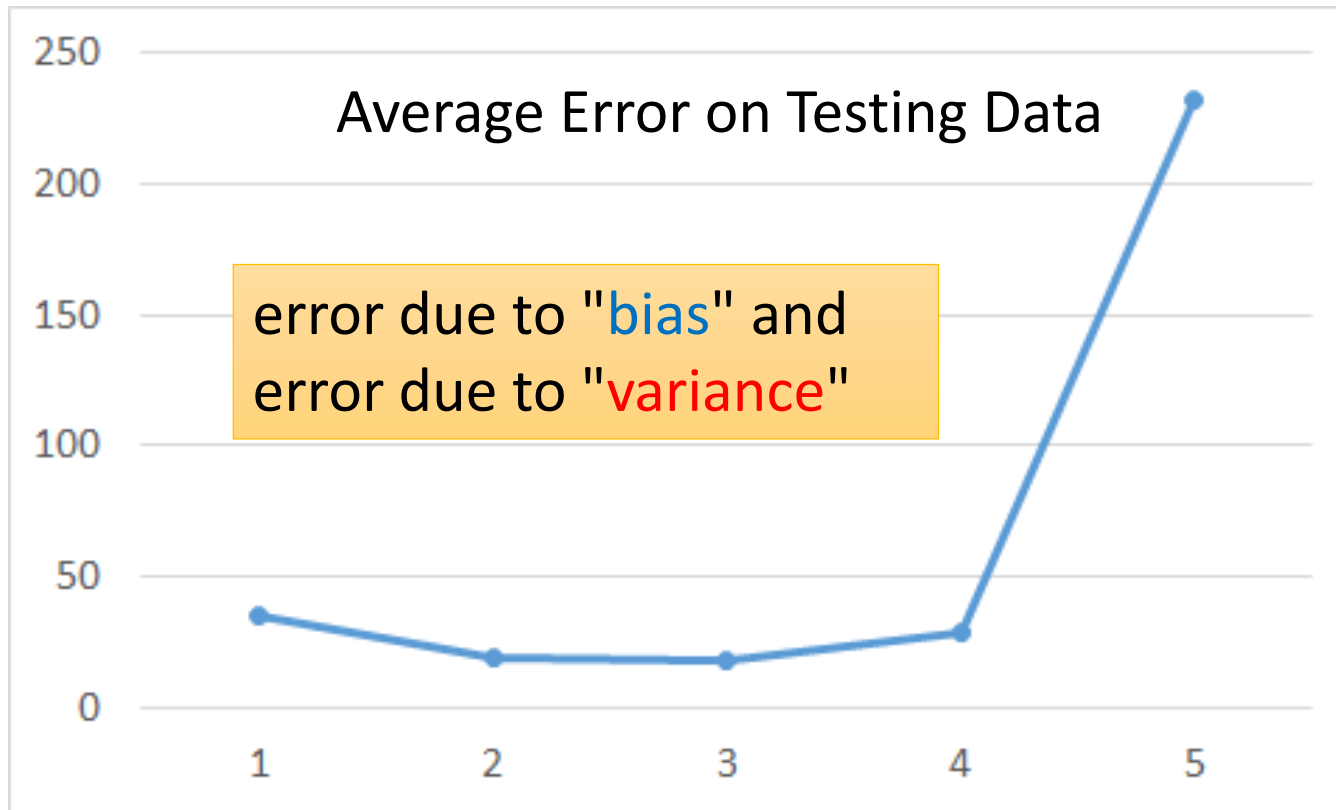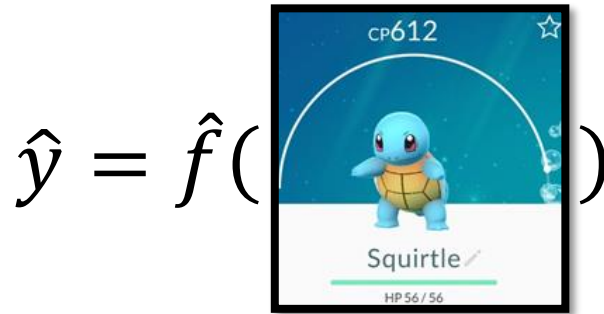
Review

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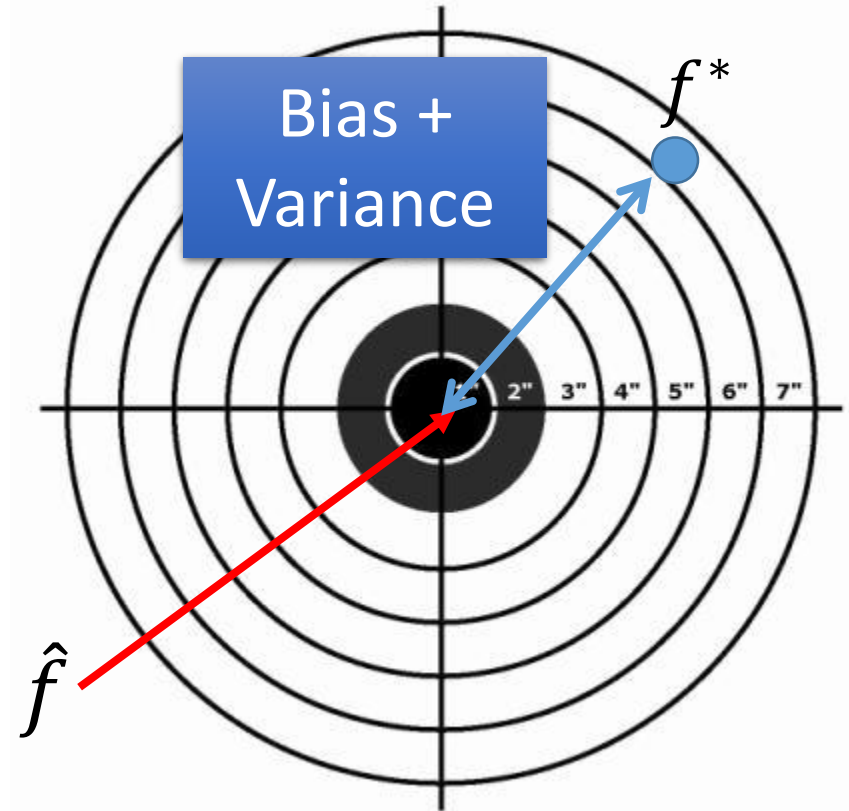
Estimator

$$\hat{y} = \hat{f}(\quad)$$

Only Niantic knows $\hat{f}$

From training data,
we find f^*

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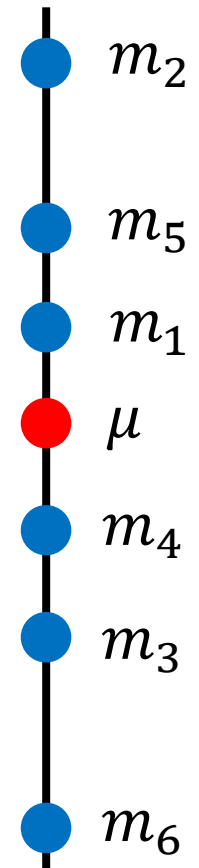
Bias + Variance

f^*

$\hat{f}$

Bias and Variance of 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 mean μ
  - Sample N points: $\{x^1, x^2, \ldots, x^N\}$

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

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

unbiased

m_2

m_5

m_1

μ

m_4

m_3

m_6

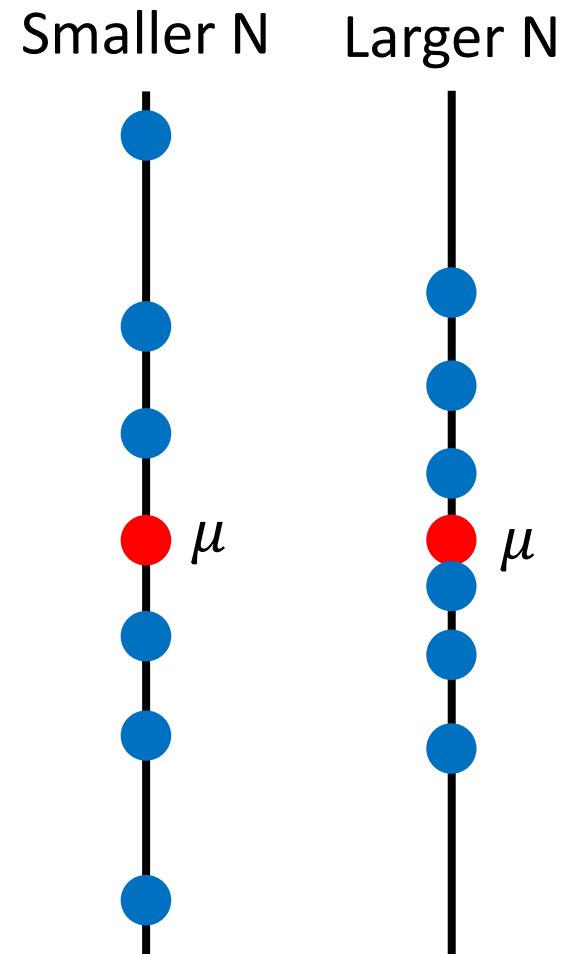# Bias and Variance of Estimator

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

- Estimator of mean μ
  - Sample N points: $\{x^1, x^2, \dots, x^N\}$

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

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

Variance depends on the number of samples

unbiased

Smaller N   Larger N

μ μ
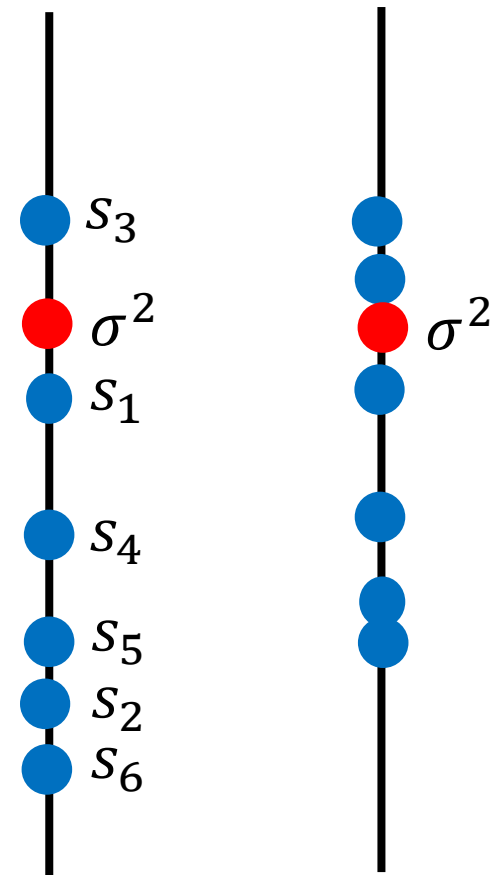
Bias and Variance of Estimator

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

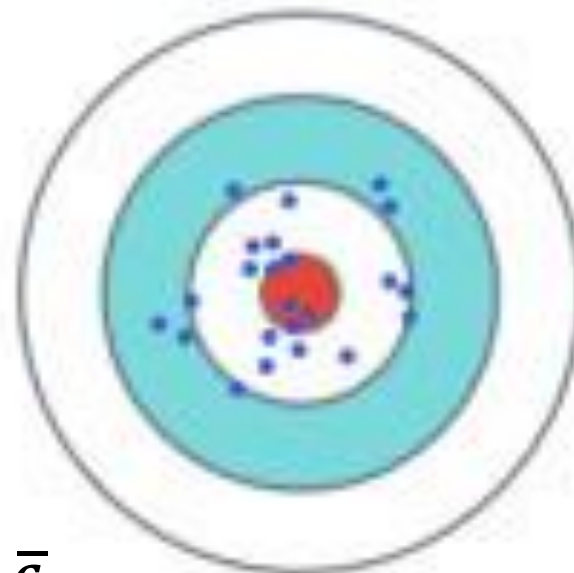$$E[s] = \frac{N-1}{N}\sigma^2 \quad \neq \sigma^2$$
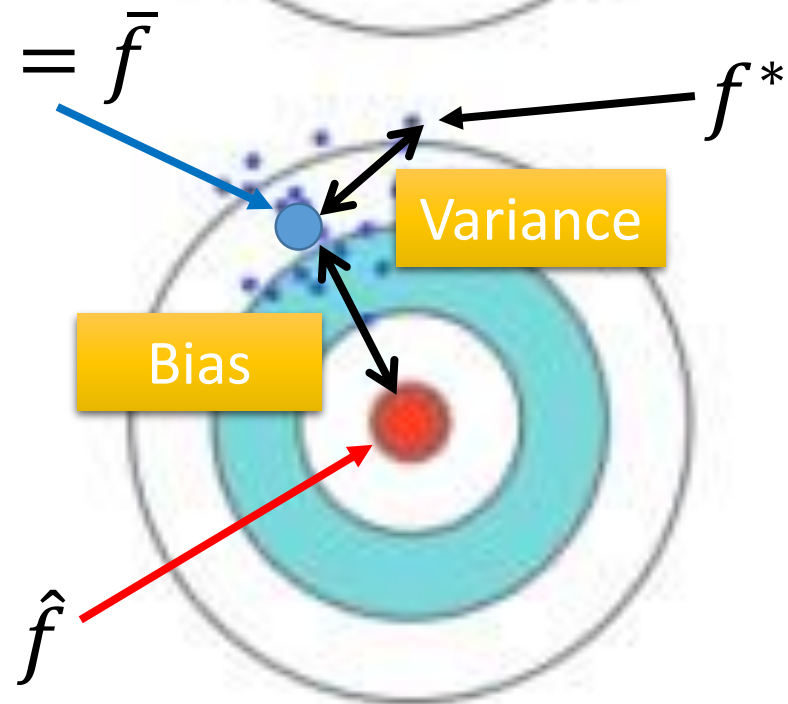
Increase N

Low Variance | High Variance

Low Bias | High Bias

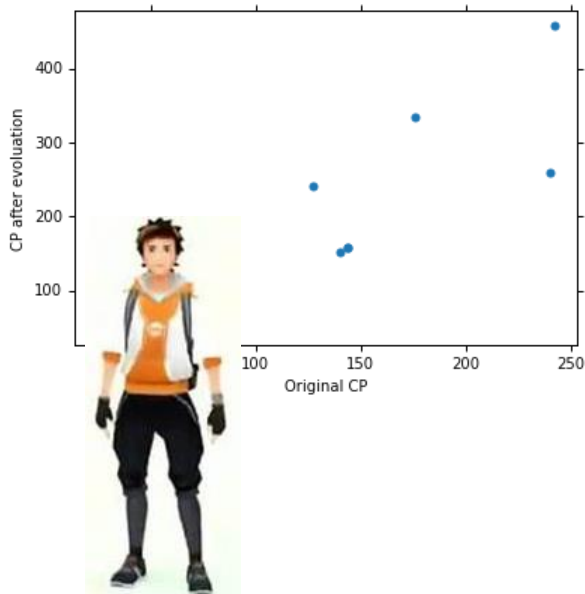$$E[f^*] = \bar{f}$$

f^*

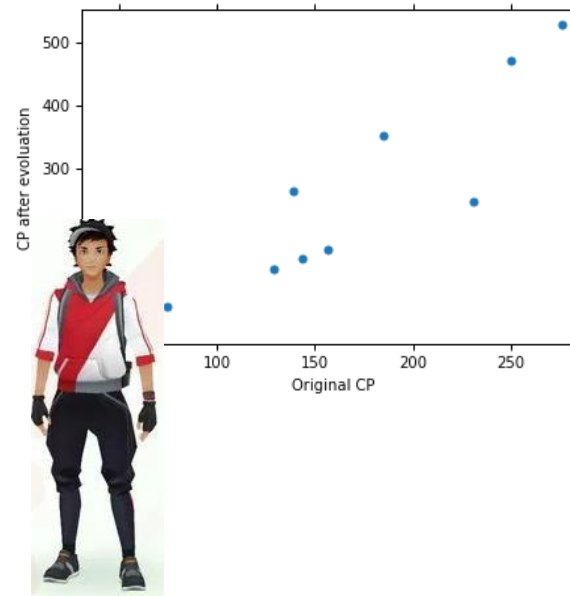Variance

Bias

$\hat{f}$

Parallel Universes

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

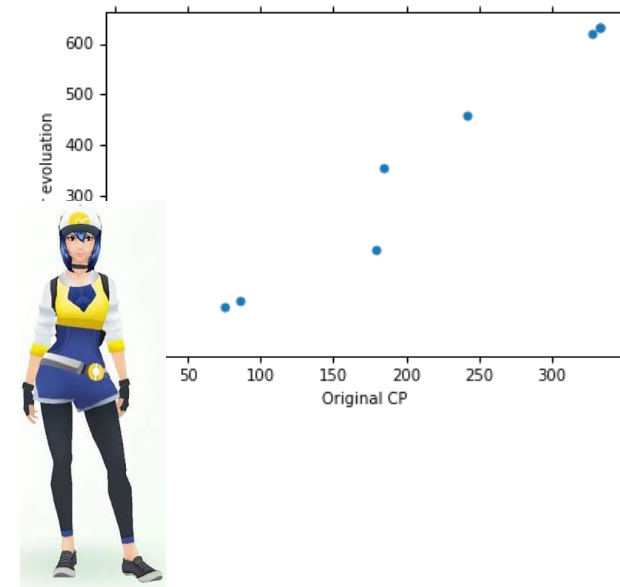$f*$：利用不同training data所train出
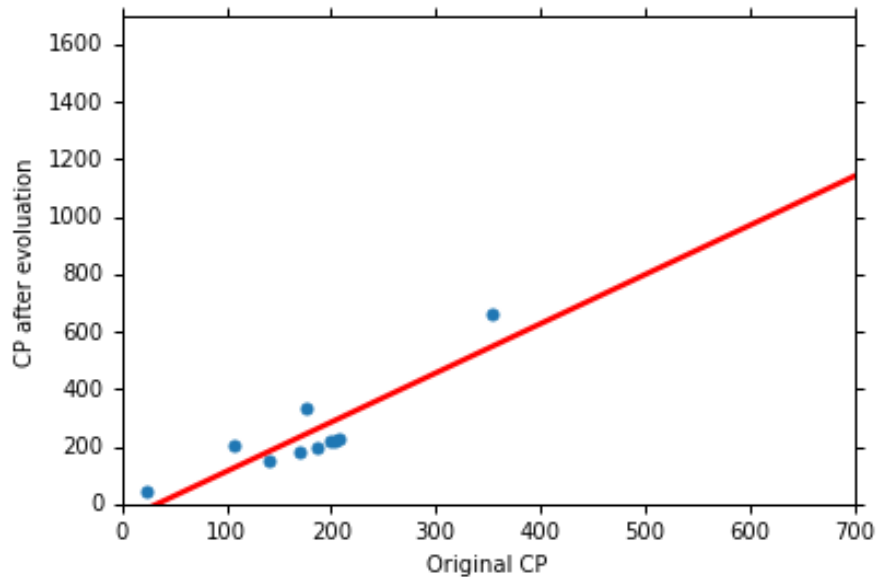來的function

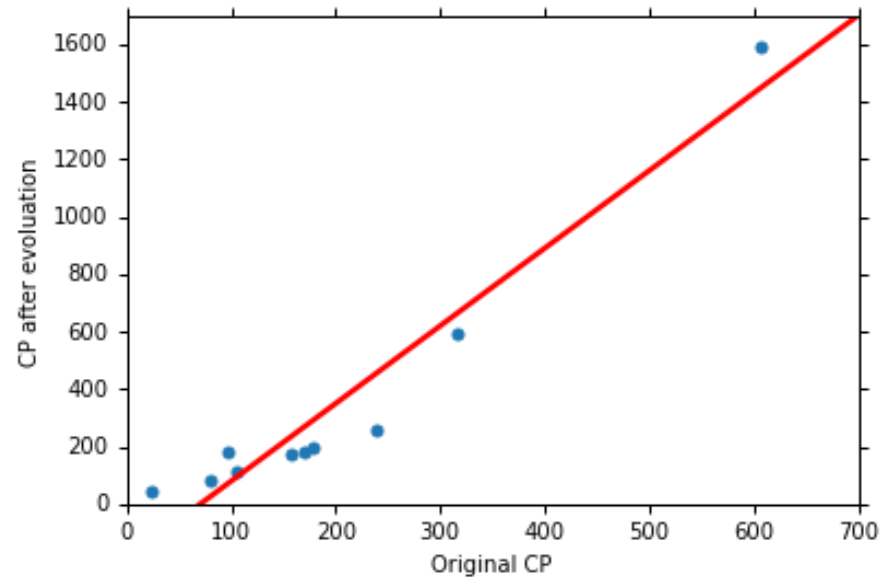Universe 1   Universe 2   Universe 3

Parallel Universes

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

Universe 123
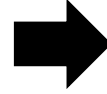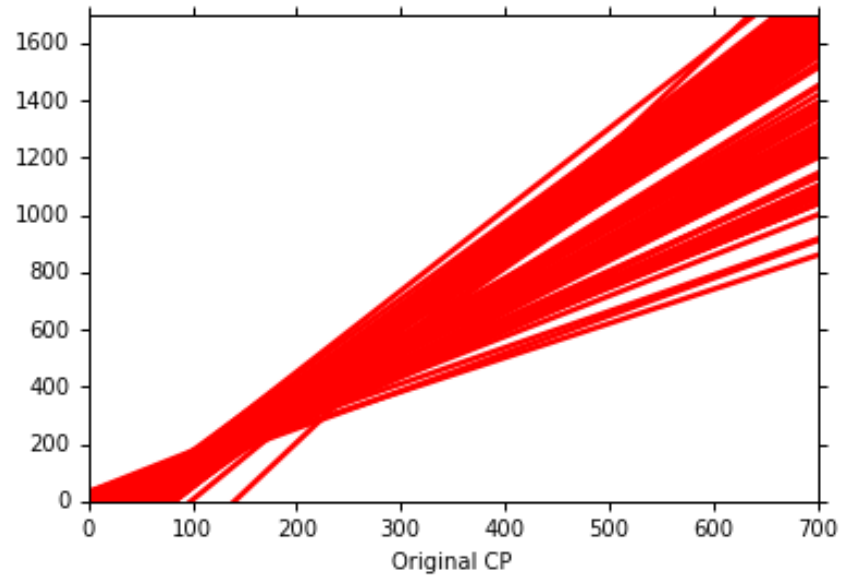
Universe 345

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

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

f^* in 100 Universes
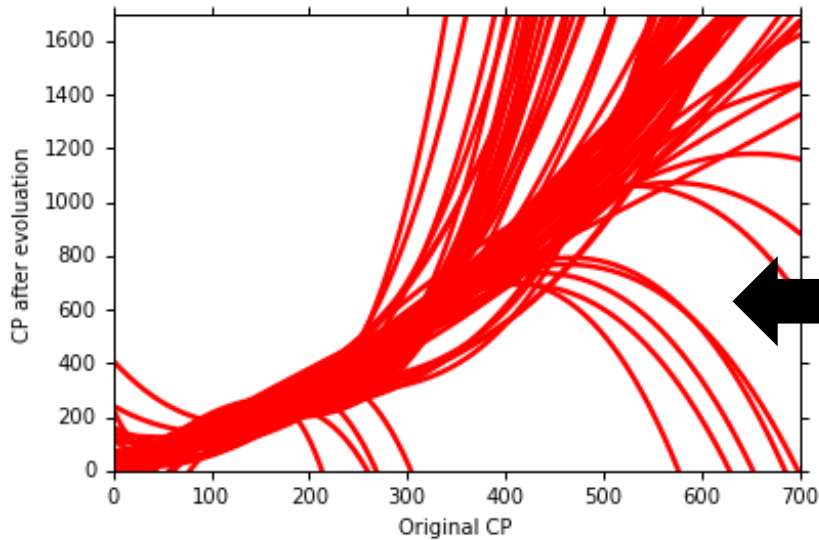
$$y = b + w \cdot x_{cp}$$
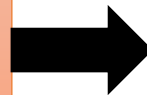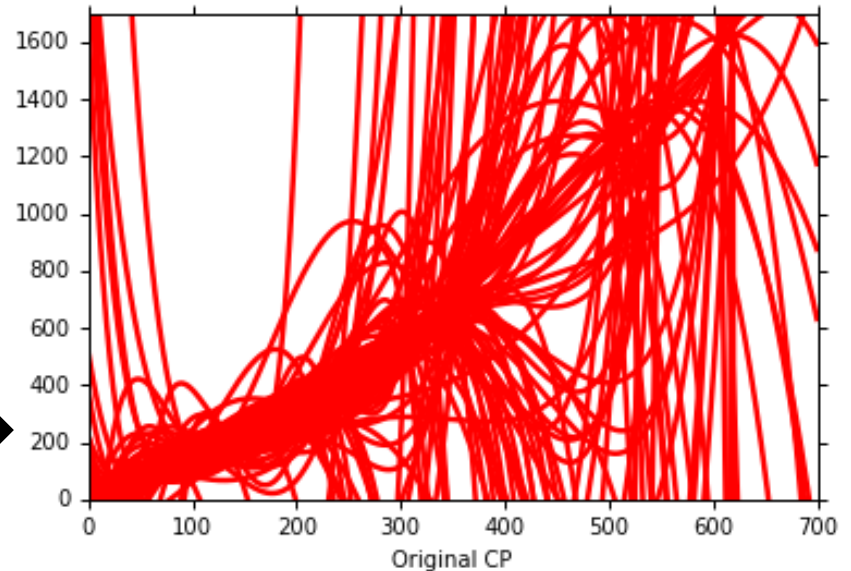

$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3$$

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

Variance

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

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

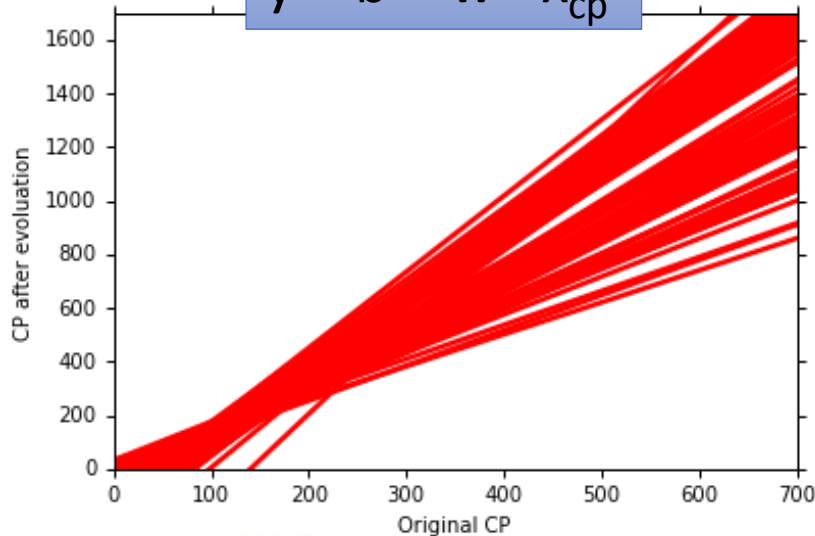Small Variance

Large Variance

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

Assume this is $\hat{f}$

Black curve: the true function $\hat{f}$
Red curves: 5000 f^*
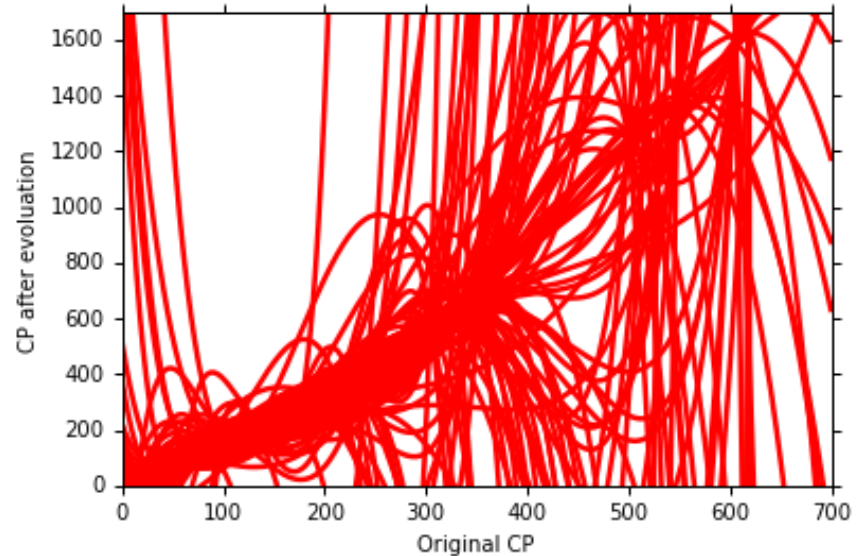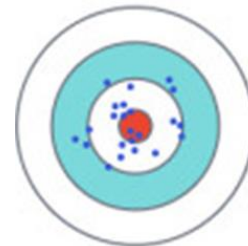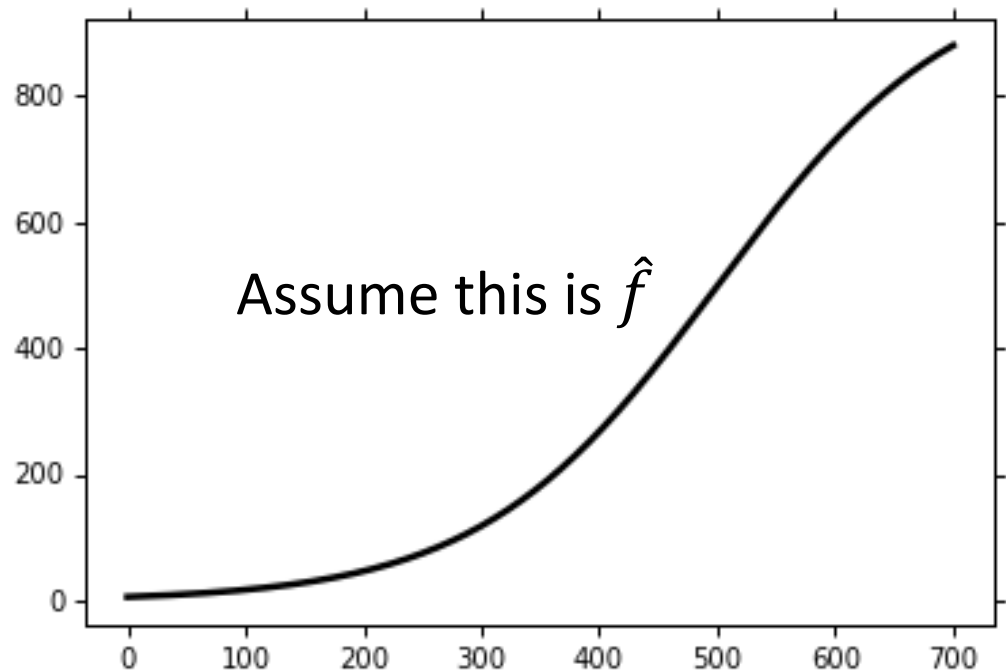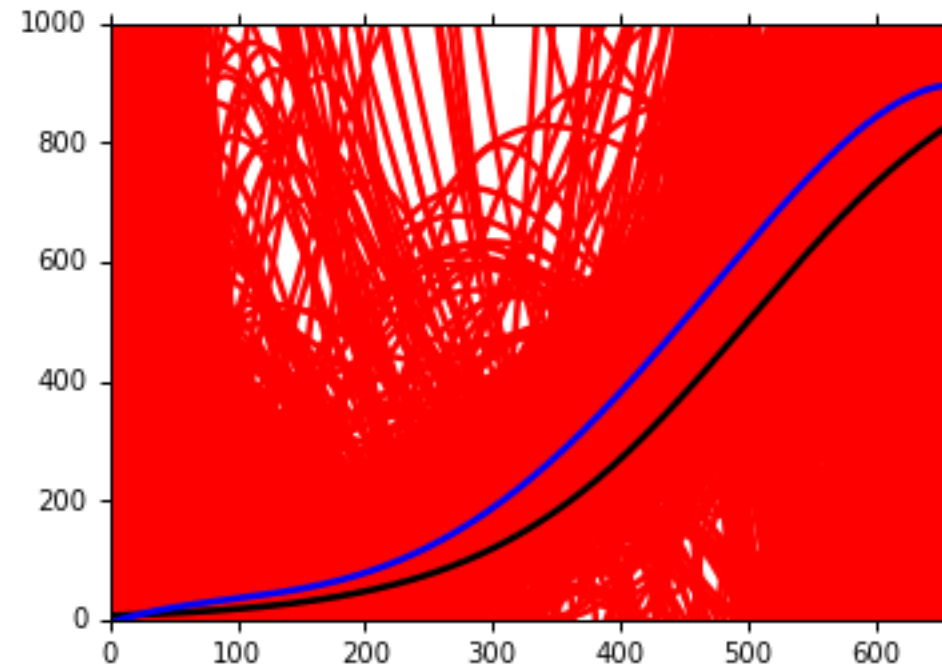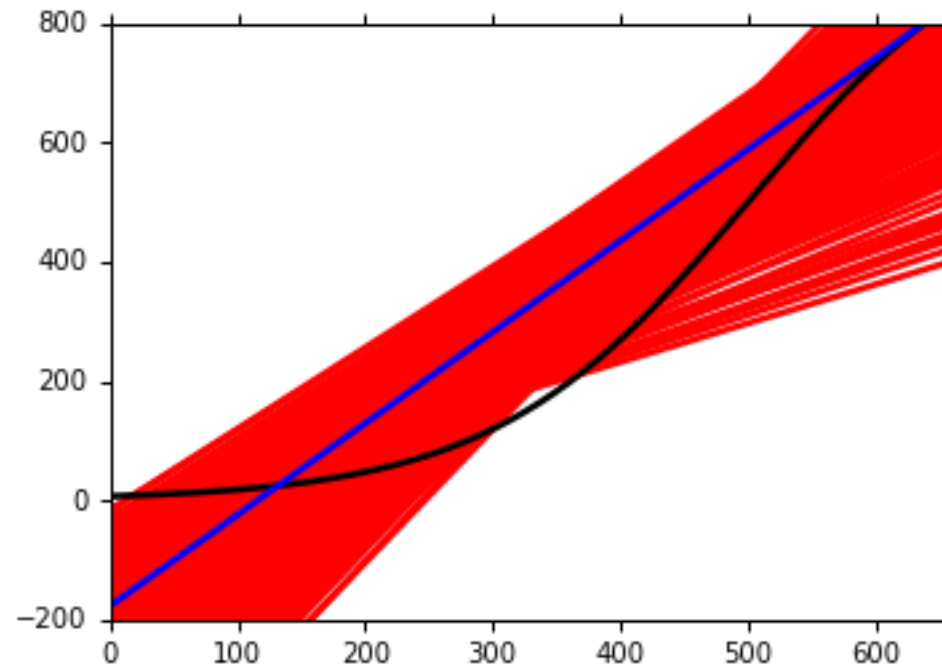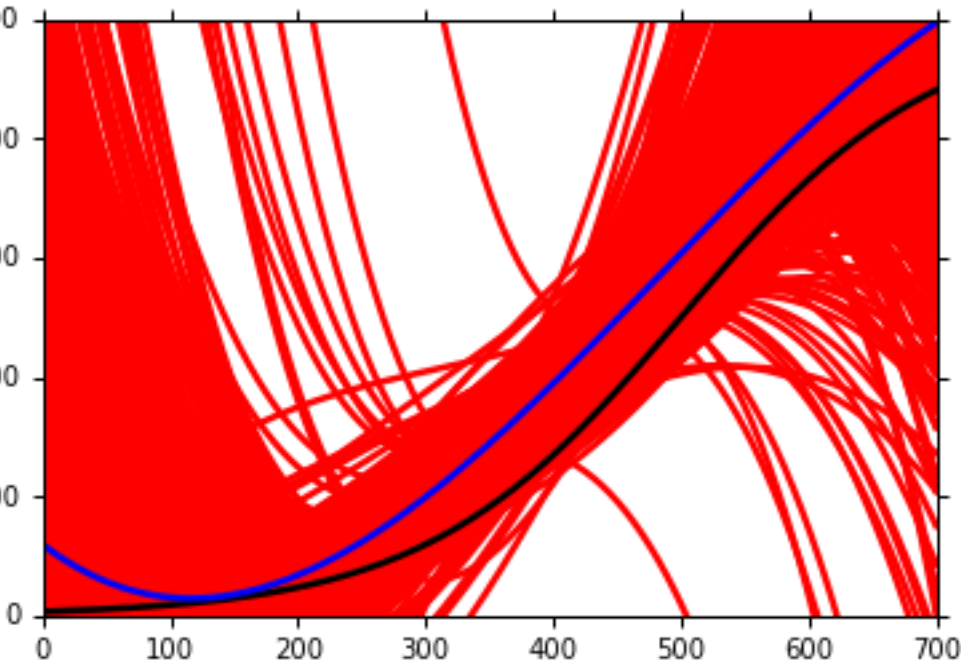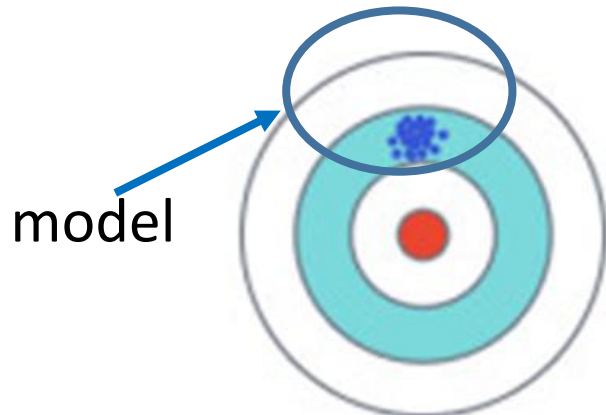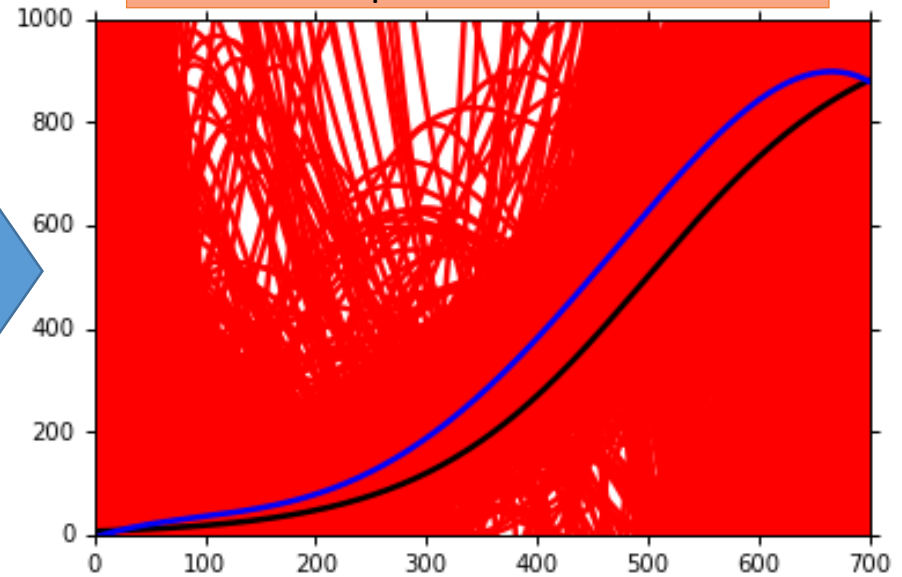Blue curve: the average of 5000 f^*
$= \bar{f}$

Bias

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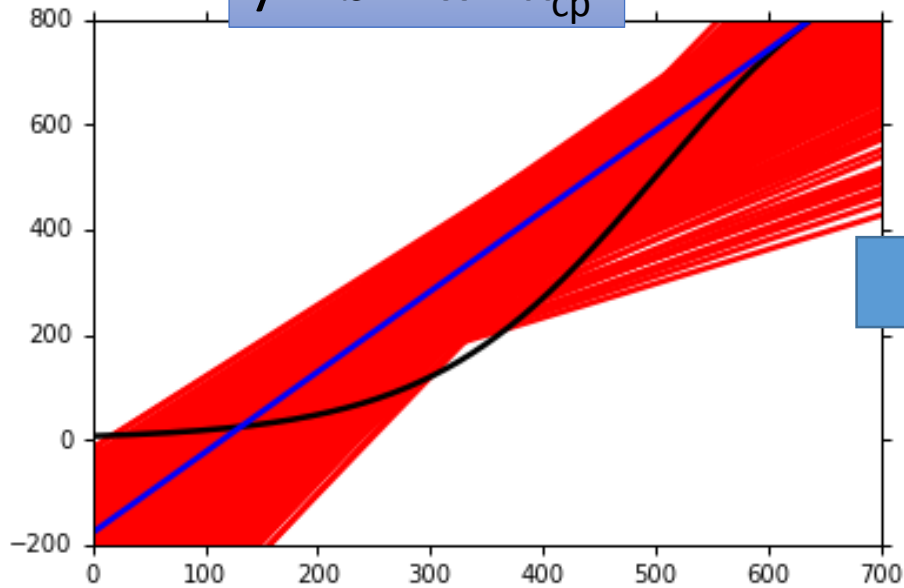
$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4 + w_5 \cdot (x_{cp})^5$$

model
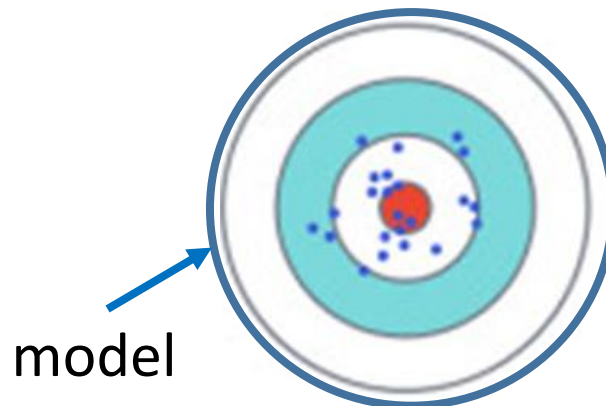
Large Bias

model

Small Bias

Bias v.s. Variance

Error from bias
Error from variance
Error observed

Underfitting

Overfitting

複雜度

Large Bias

Small Variance

Small Bias

Large Variance
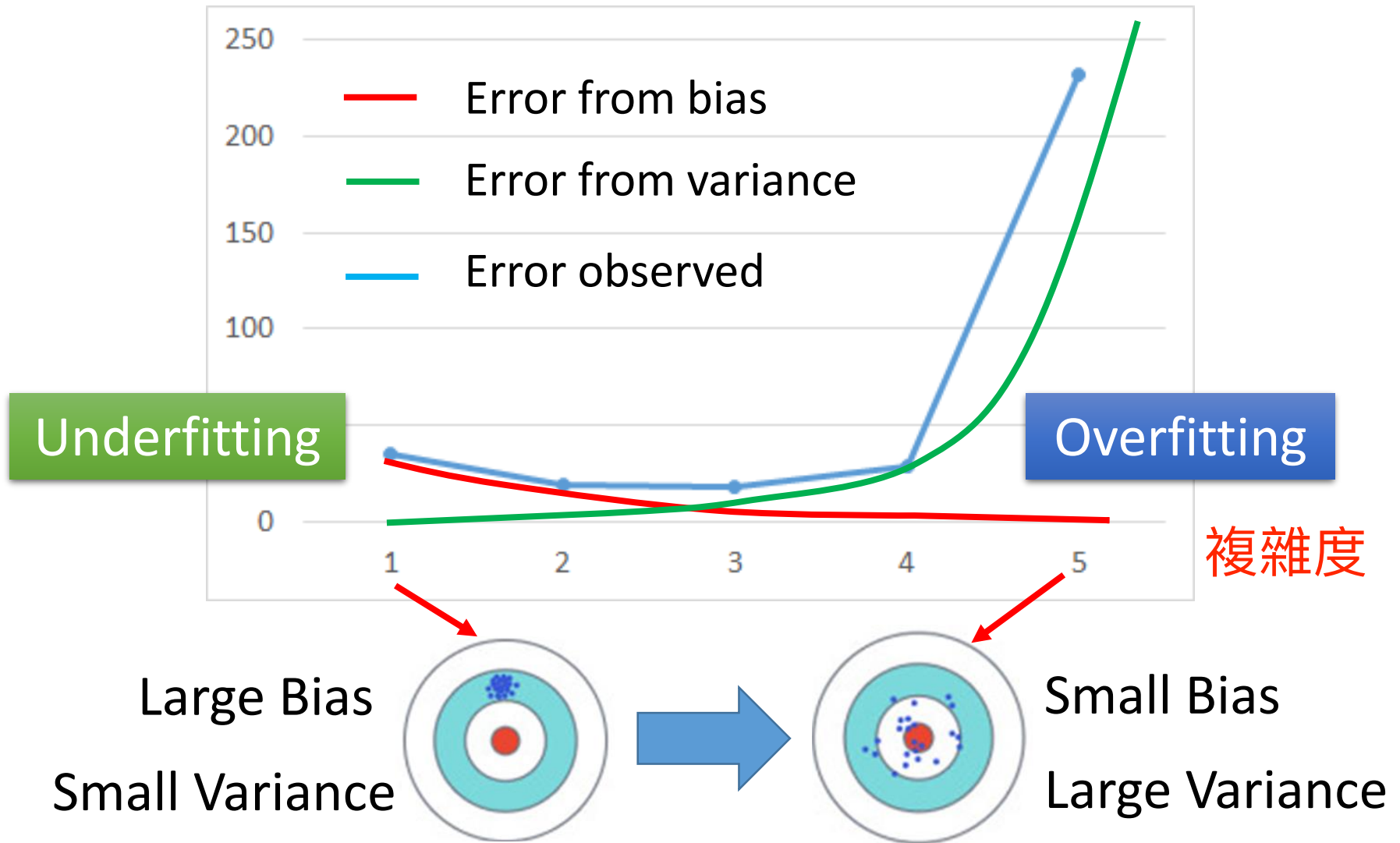
What to do with large bias?

- 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
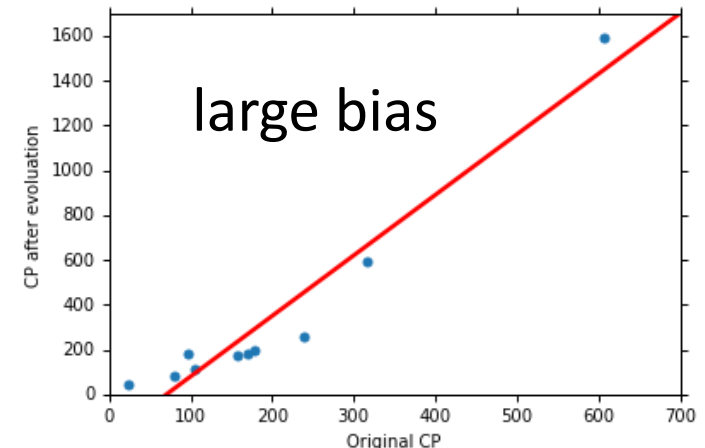
large bias

What to do with large variance?

• More data

ex. 圖片翻轉幾度、增加雜訊…

Very effective, but not always practical

10 examples

100 examples
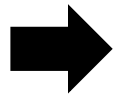
• Regularization ➡️ May increase bias

選擇較平滑的function，但可能因此忽略正確的f而傷害了bias
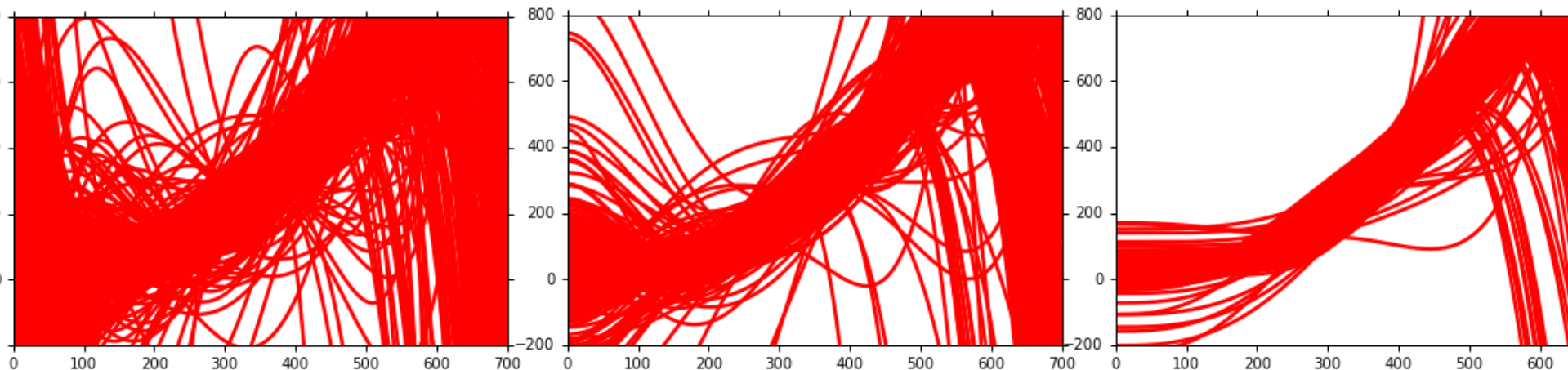
Model Selection

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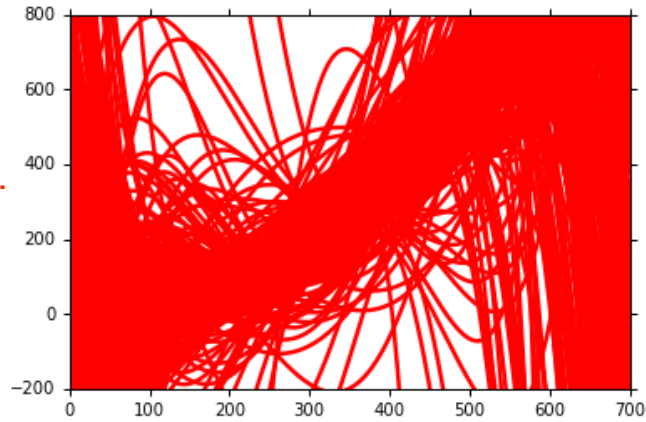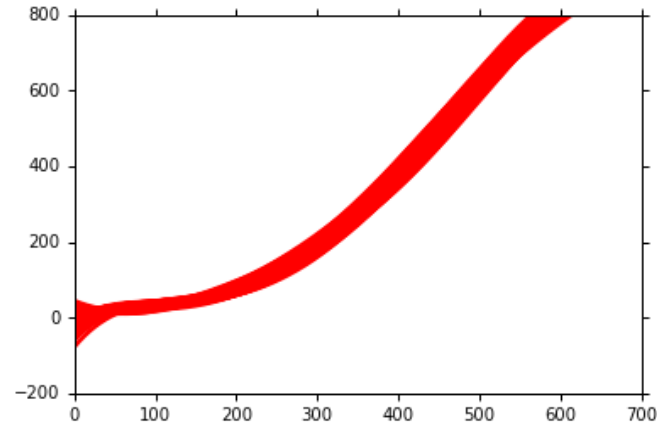
- Select a model that balances two kinds of error to minimize total error

- What you should NOT do:

Homework

Training Set	public Testing Set	private Testing Set

Model 1 ———→ Err = 0.9

Model 2 ———→ Err = 0.7

Model 3 ———→ Err = 0.5 ——→ 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/

TOP 10 IN PUBLIC LEADERBOARD

RANKED 3XX IN PRIVATE LEADERBOARD

imgflip.com

Cross Validation

先使用validation set 找出可能適合的 modeI，再將所有所有training data透過這個modeI去train

public

private

| Training Set | Testing Set | Testing Set |

文字

Using the results of public testing data to tune your model

You are making public set better than private set.

Training Set → Validation set

Model 1 ⟶ Err = 0.9

Model 2 ⟶ Err = 0.7

Model 3 ⟶ Err = 0.5 ⟶ Err > 0.5 ⟶ Err > 0.5

Not recommend

N-fold Cross Validation

	Model 1	Model 2	Model 3
Train / Train / Val	Err = 0.2	Err = 0.4	Err = 0.4
Train / Val / Train	Err = 0.4	Err = 0.5	Err = 0.5
Val / Train / Train	Err = 0.3	Err = 0.6	Err = 0.3
	Avg Err = 0.3	Avg Err = 0.5	Avg Err = 0.4

Training Set

Testing Set — public

Testing Set — private

Reference

- Bishop: Chapter 3.2