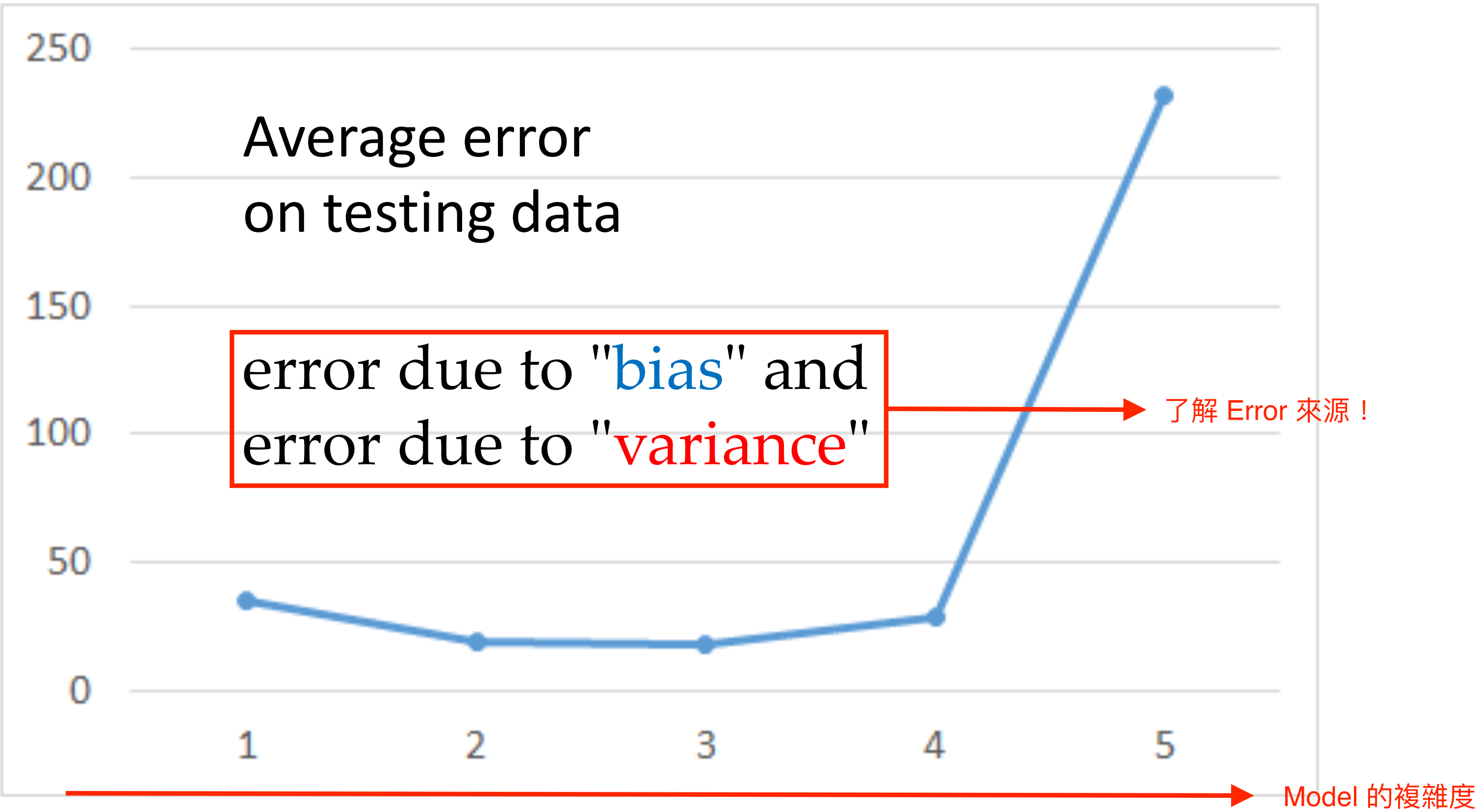
Where does the error come from?

Review



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

Estimator

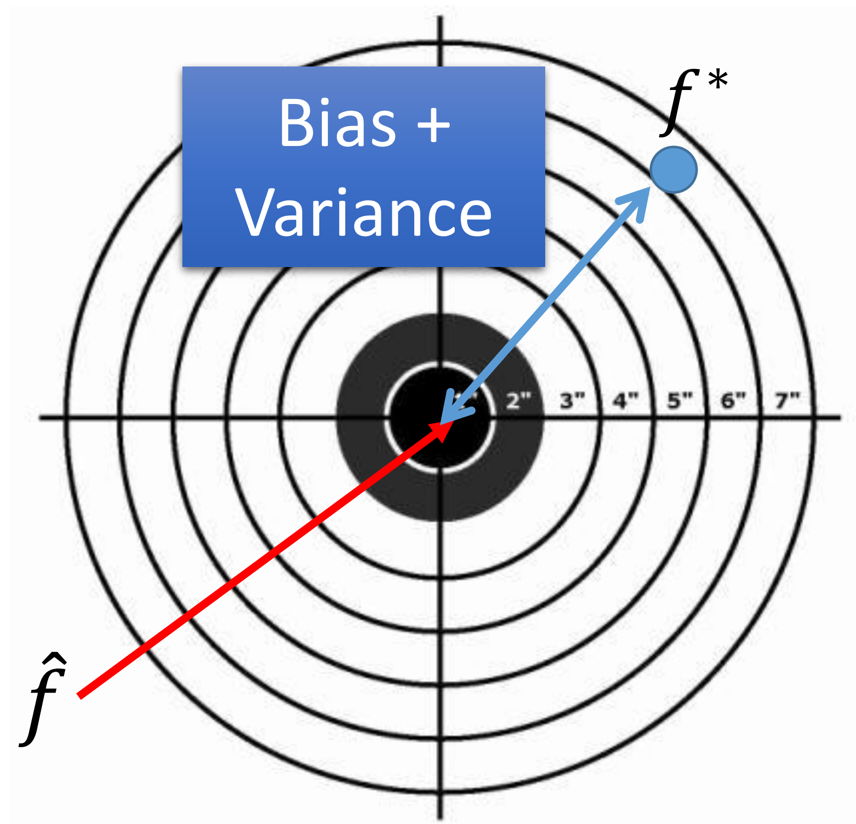
Machine 從 Function Set 中找到的 Best Function 與 真實 Function 的距離 => Bias + Variance

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

Only Niantic knows $\hat{f}$

From training data, we find f^*

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

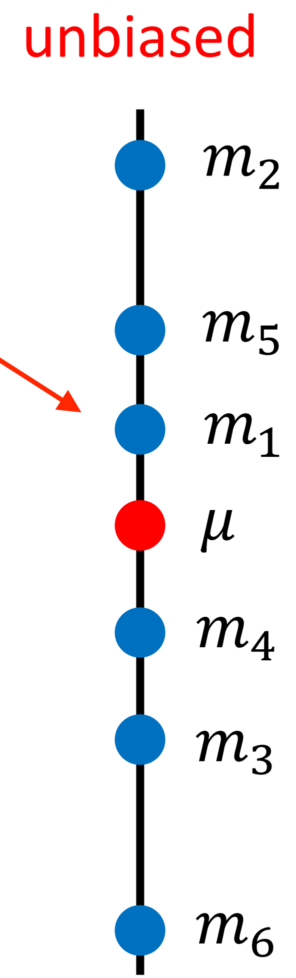


Bias and Variance of Estimator

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

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

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



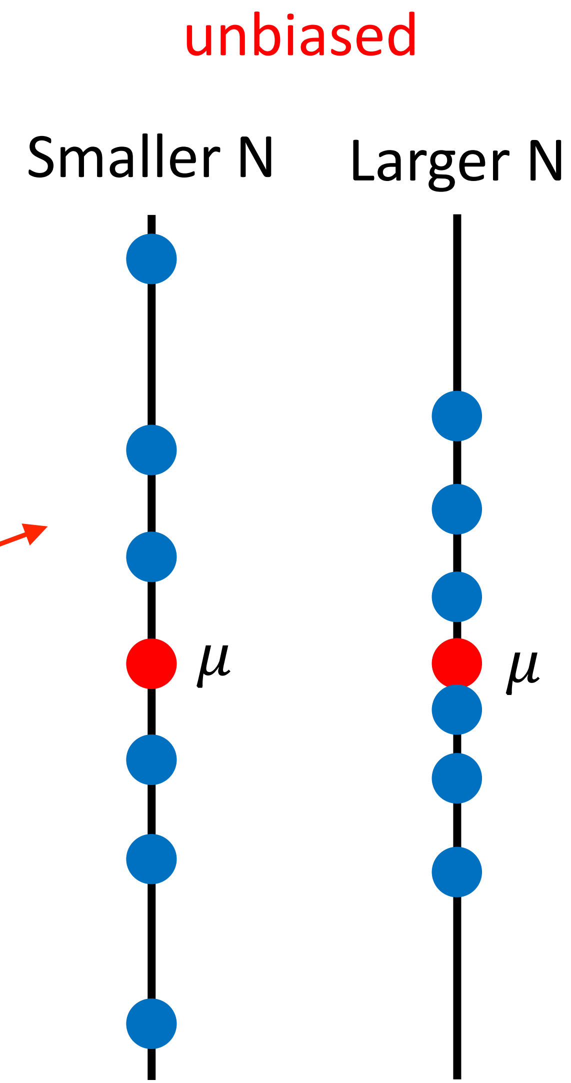
Bias and Variance of Estimator

- 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: $\{x^1, x^2, \ldots, x^N\}$

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

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

Variance depends on the number of samples



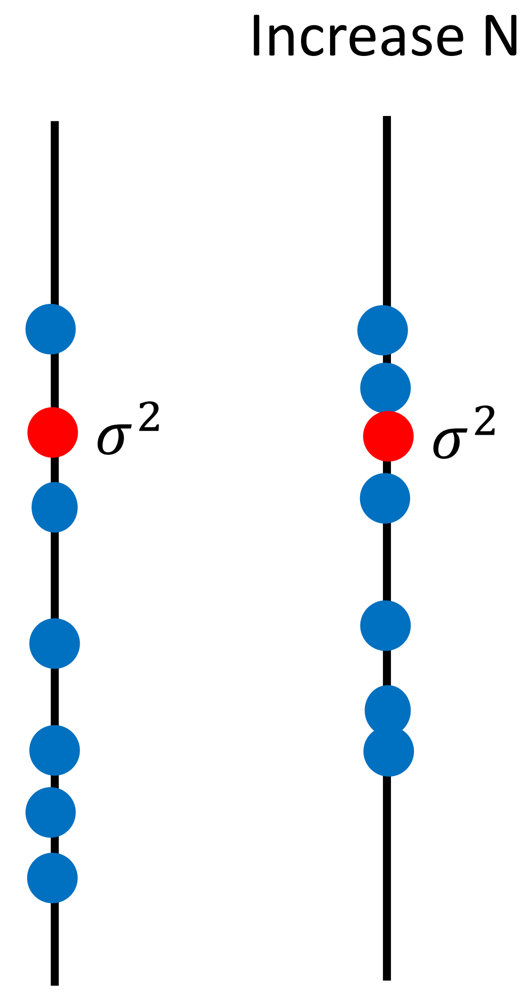
Bias and Variance of Estimator

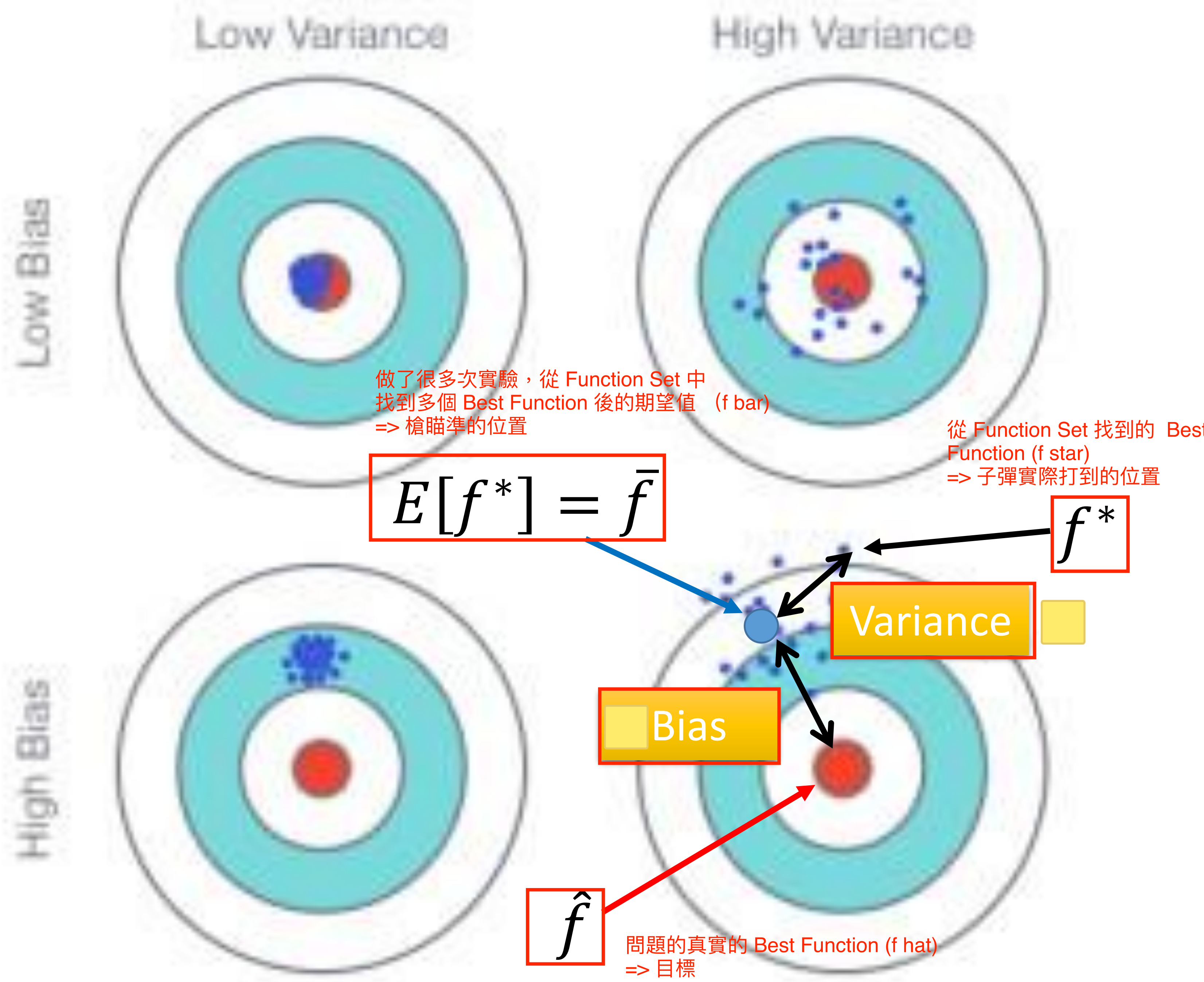
- Estimate the mean of a variable x
  - assume the mean of x is μ
  - assume the variance of x is σ^2
- Estimator of variance σ^2
  - Sample N points: $\{x^1, x^2, \ldots, x^N\}$

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

Biased estimator

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





Low Variance
High Variance
Low Bias
High Bias
做了很多次實驗，從 Function Set 中
找到多個 Best Function 後的期望值 (f bar)
=> 槍瞄準的位置
$E[f^*] = \bar{f}$
從 Function Set 找到的 Best
Function (f star)
=> 子彈實際打到的位置
f^*
Variance
Bias
$\hat{f}$
問題的真實的 Best Function (f hat)
=> 目標

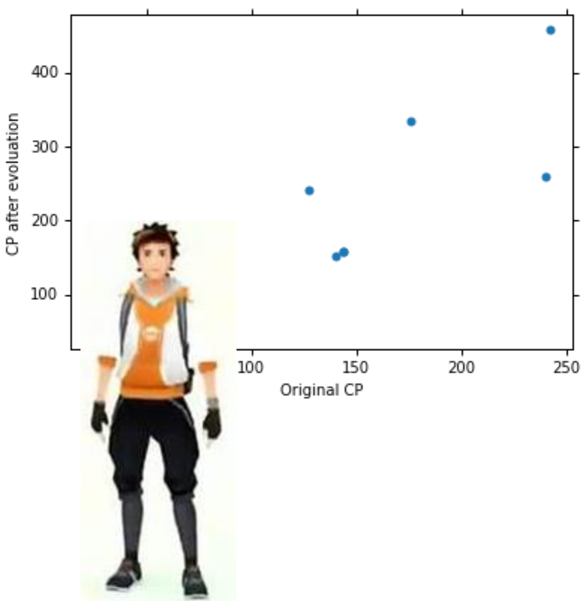
如何去找到「槍瞄準的位置」=> f bar => f star 的期望值

Parallel Universes

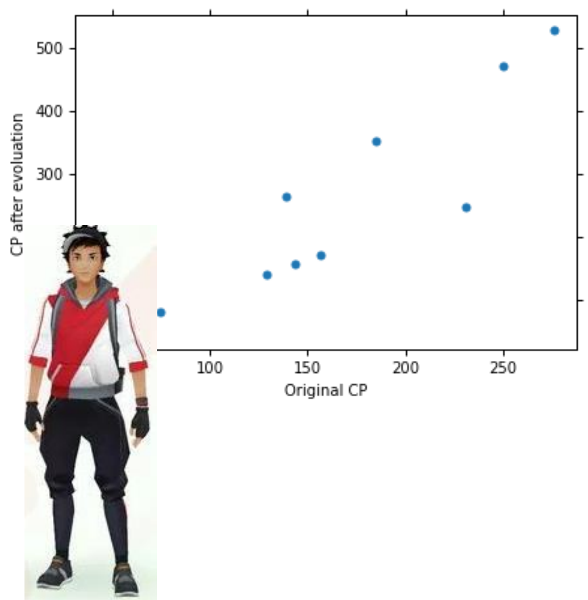
- In all the universes, we are collecting (catching) 10 Pokémons as training data to find f^*

假設有很多個平行宇宙，在每一個平行宇宙中都做相同的實驗：抓十隻寶可夢，預測寶可夢進化後的 CP

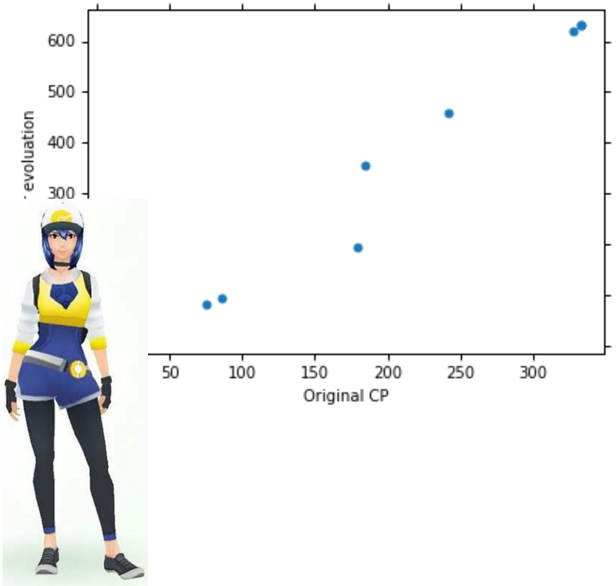
Universe 1



Universe 2



Universe 3

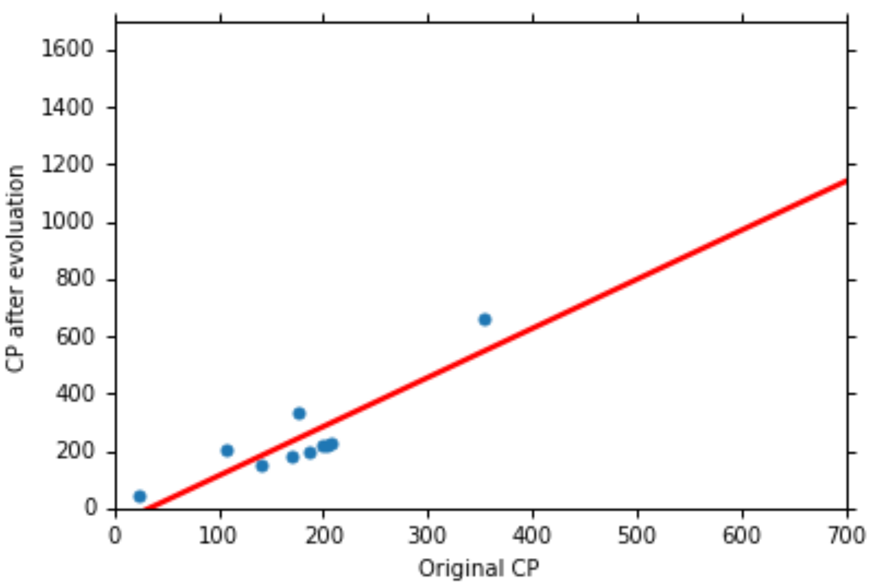


Parallel Universes

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

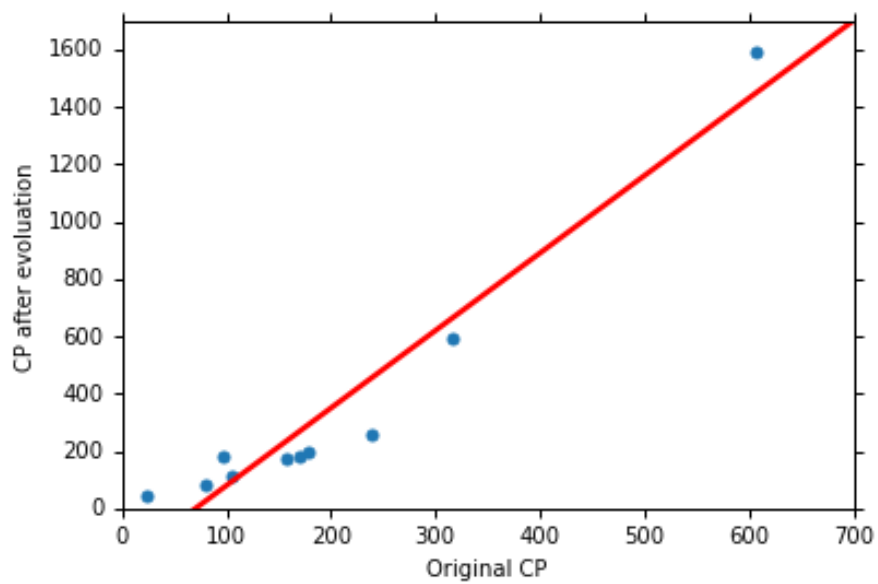
在不同的平行宇宙中，都使用相同的 Model (Function Set) 與相同的 Loss function，因為抓到的十隻寶可夢（Training Data）都不同，所以從 Function Set 中找到的 Best Function 也會不同！

Universe 123



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

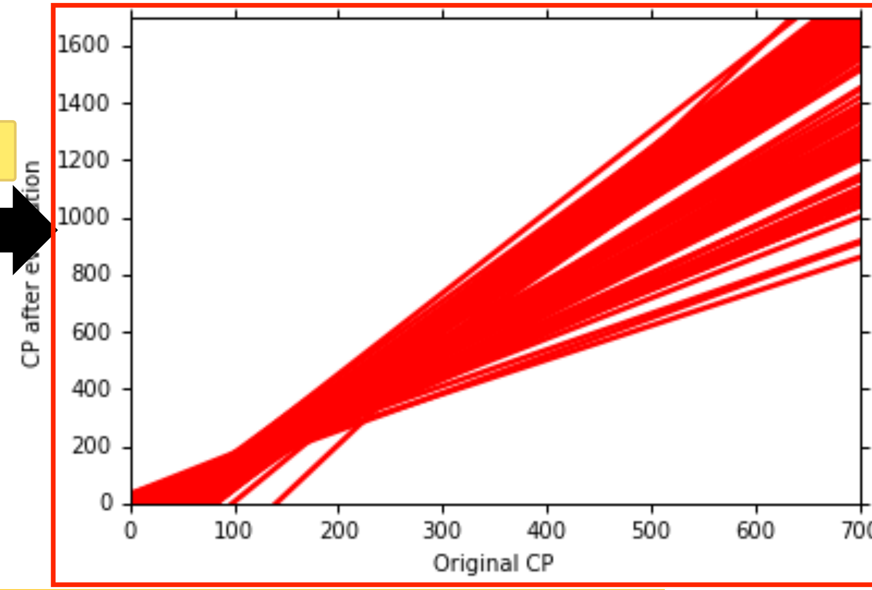
Universe 345



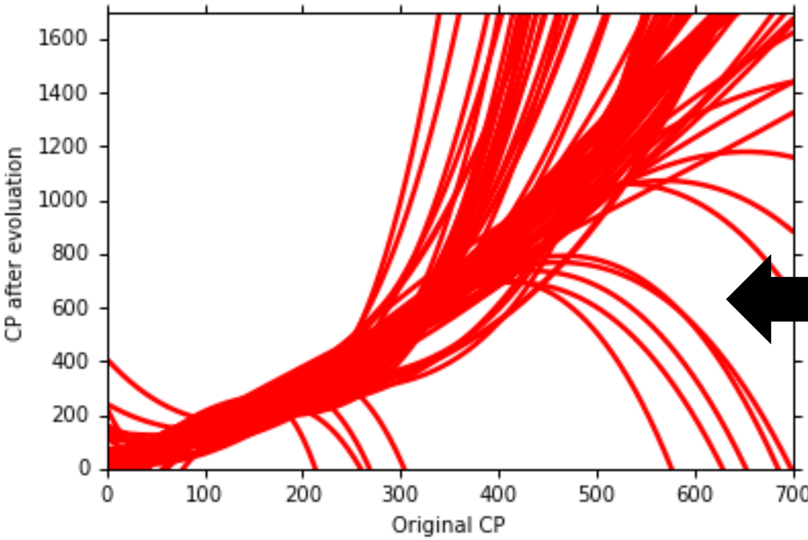
$$y = b + w \cdot x_{cp}$$

f^* in 100 Universes

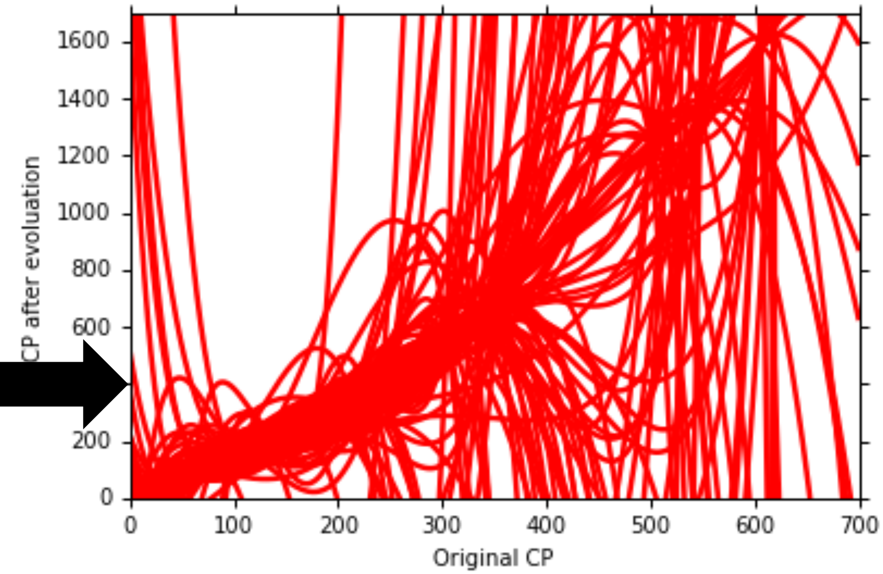
$$y = b + w \cdot x_{cp}$$



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

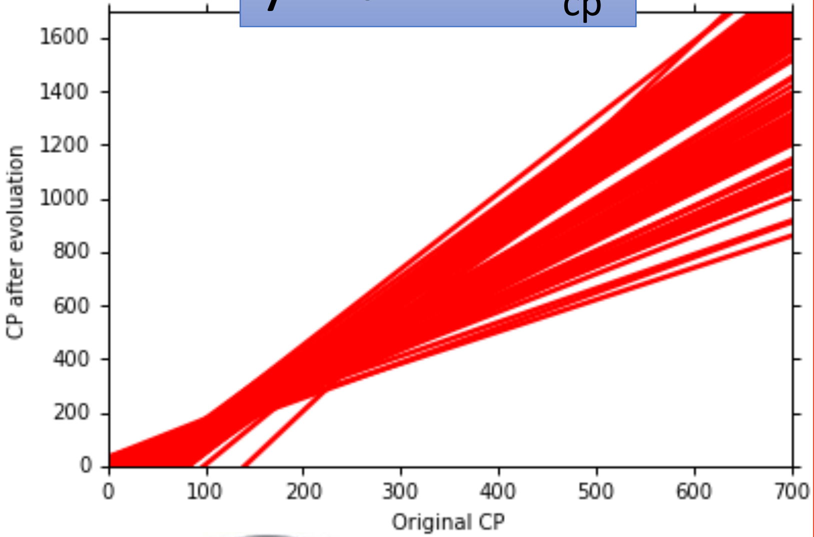


$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4 + w_5 \cdot (x_{cp})^5$$



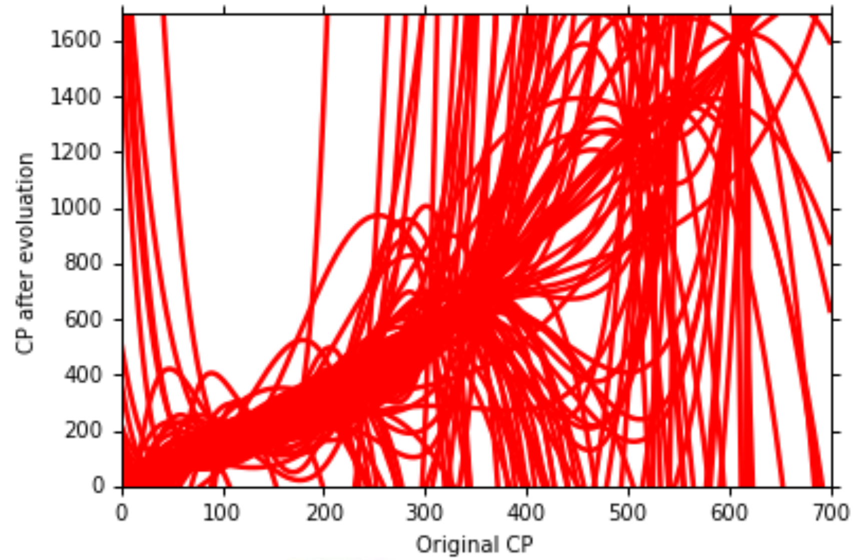
Variance

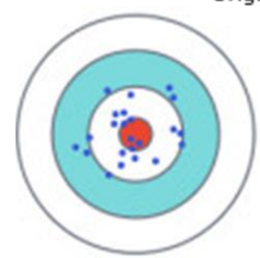
$$y = b + w \cdot x_{cp}$$



Small Variance

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





Large Variance

Simpler model is less influenced by the sampled data

Consider the extreme case f(x) = c

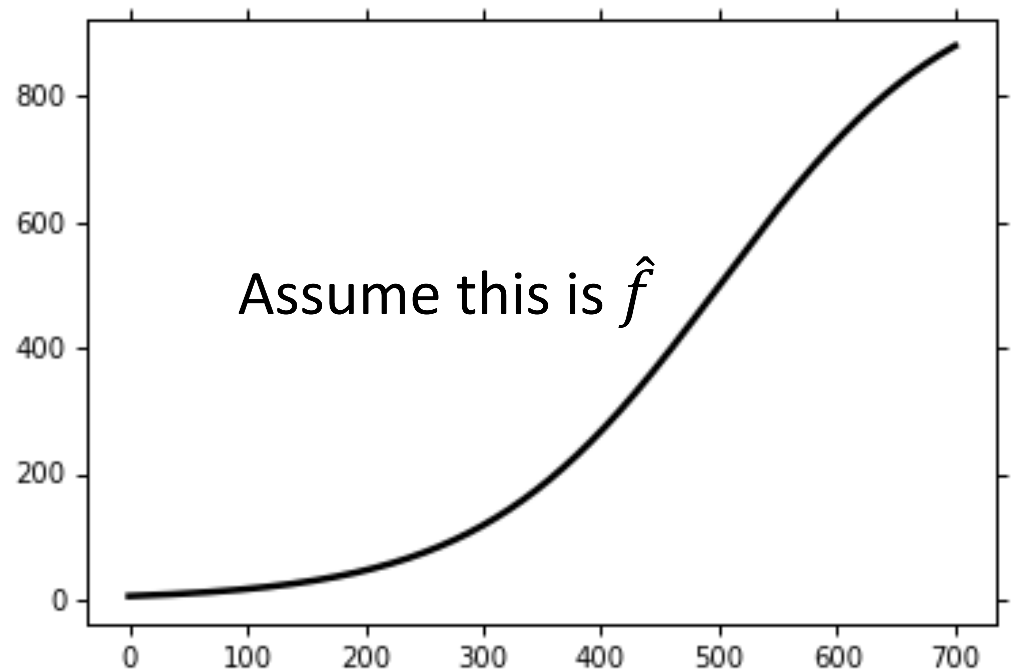
Bias

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

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

Large Bias

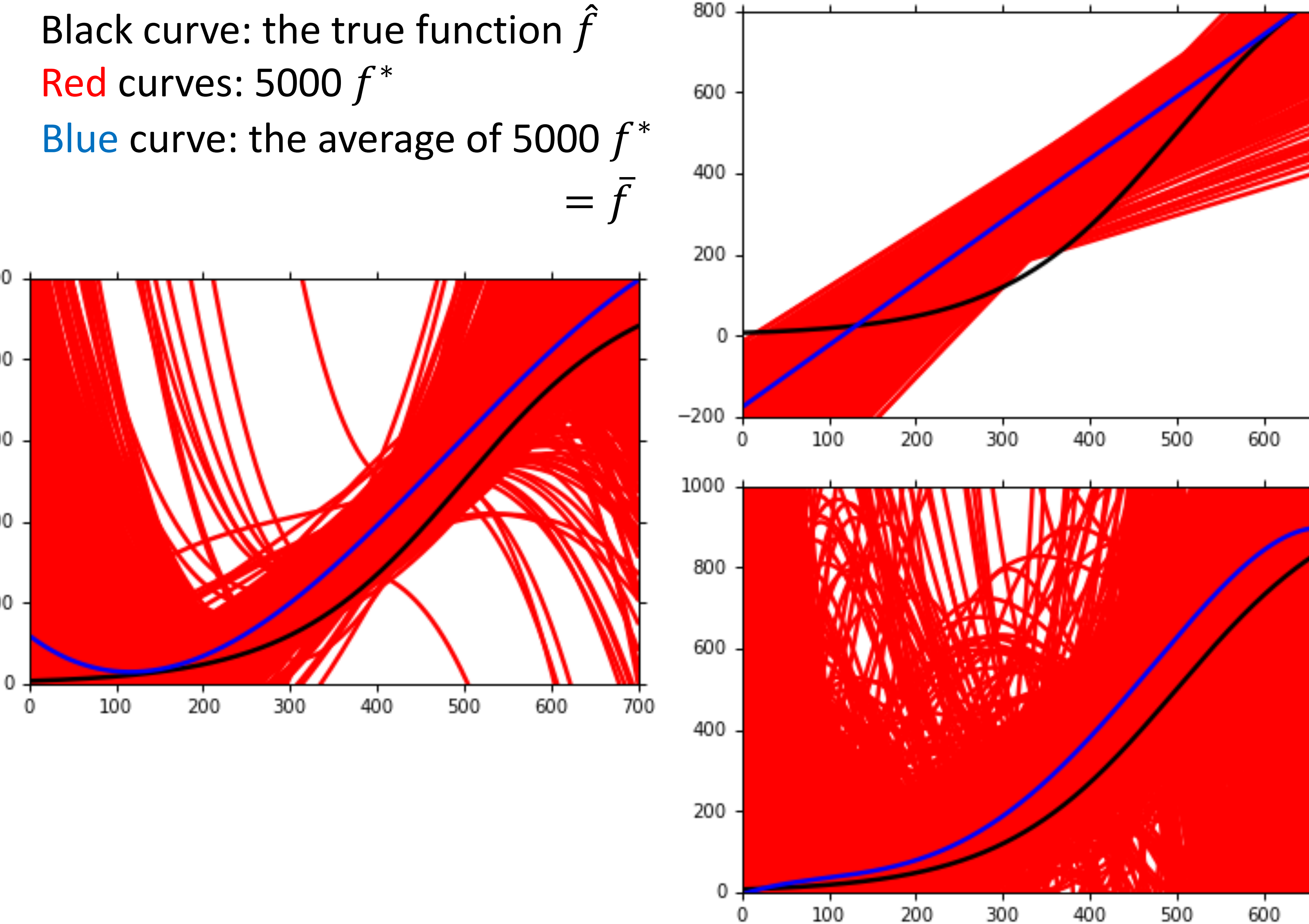
Small Bias



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

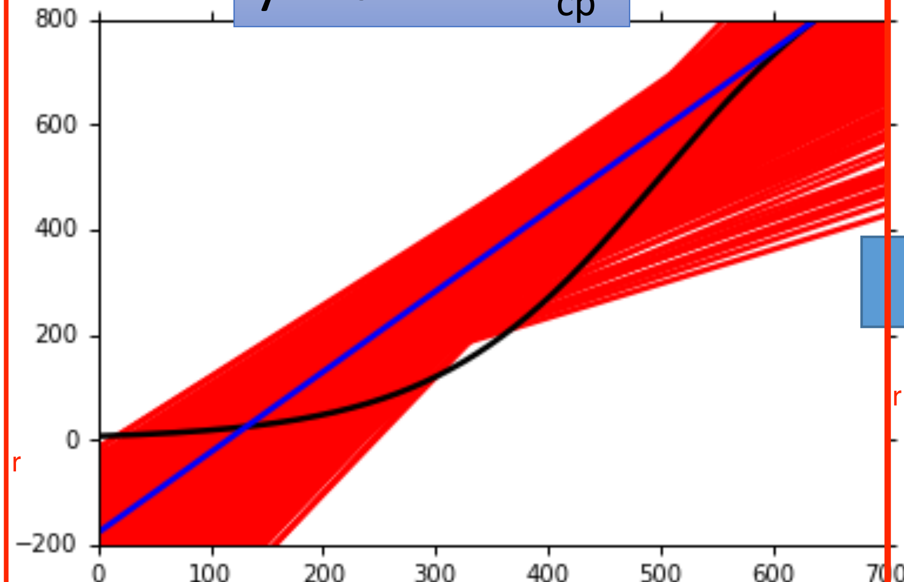
Red curves: 5000 f^*

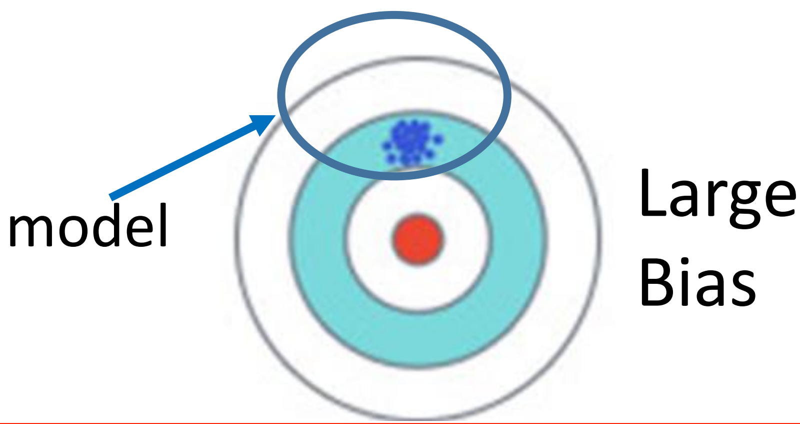
Blue curve: the average of 5000 f^* $= \bar{f}$



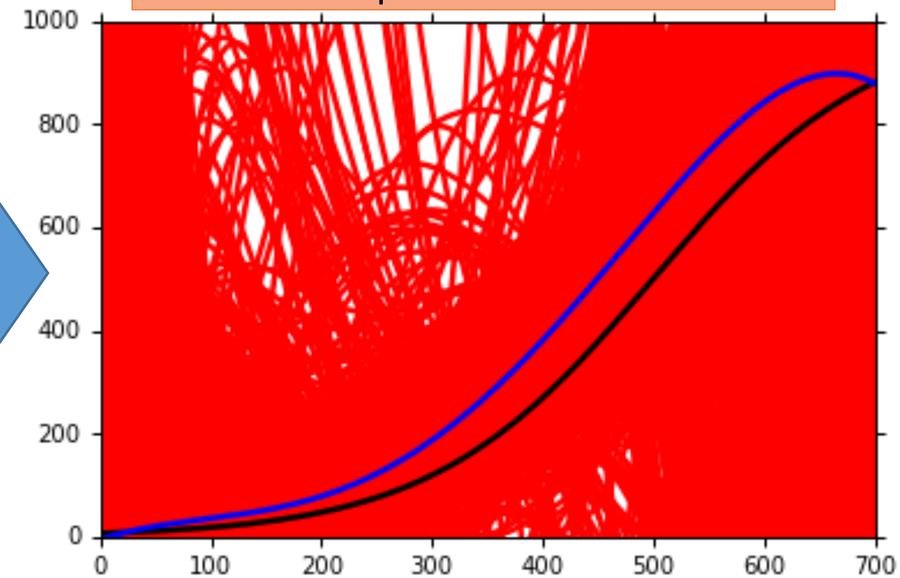
Bias

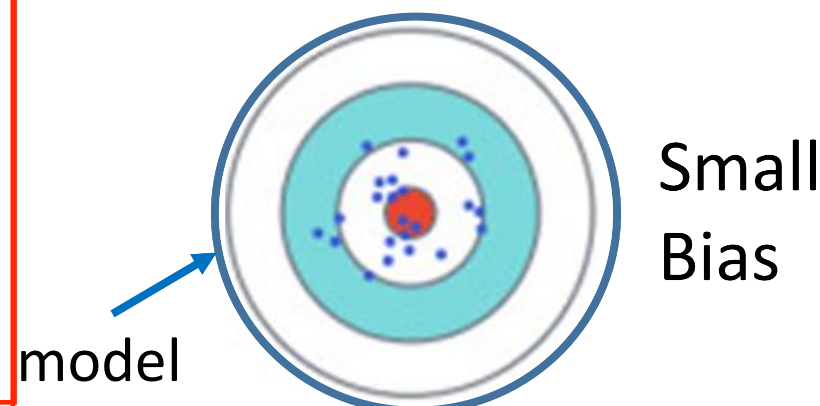
$$y = b + w \cdot x_{cp}$$



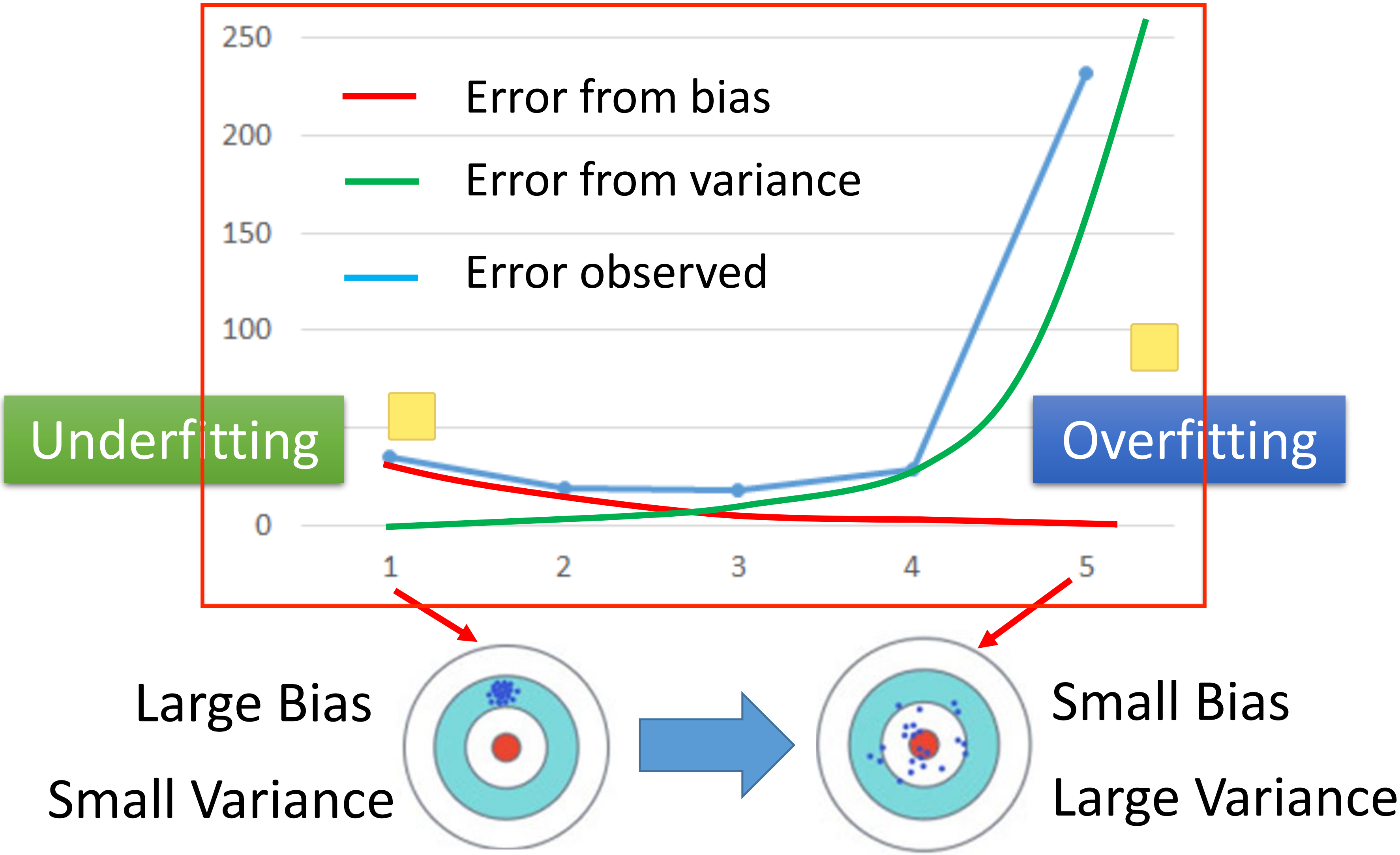


$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4 + w_5 \cdot (x_{cp})^5$$





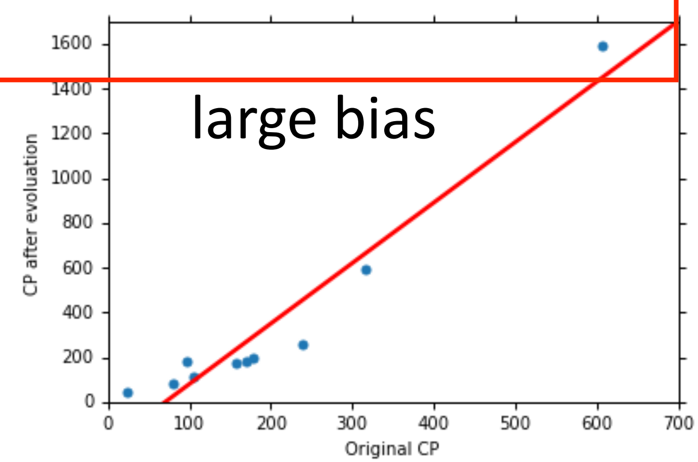
Bias v.s. Variance



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

解決 Underfitting (Bias 大) 問題

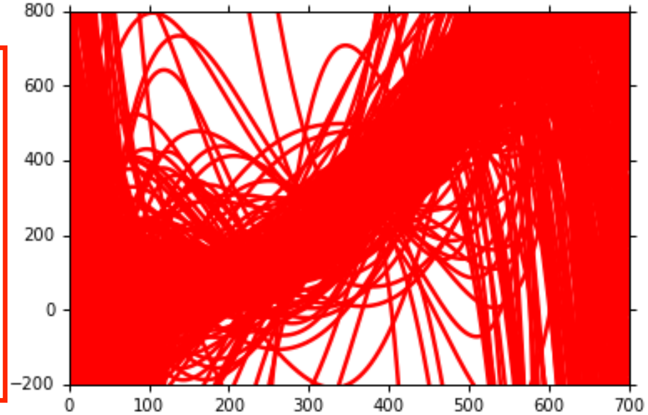


What to do with large variance?

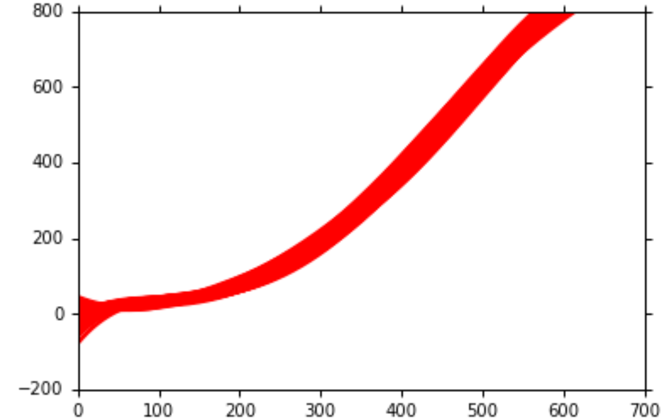
- More data

Very effective, but not always practical

解決 Overfitting (Variance 大) 的問題

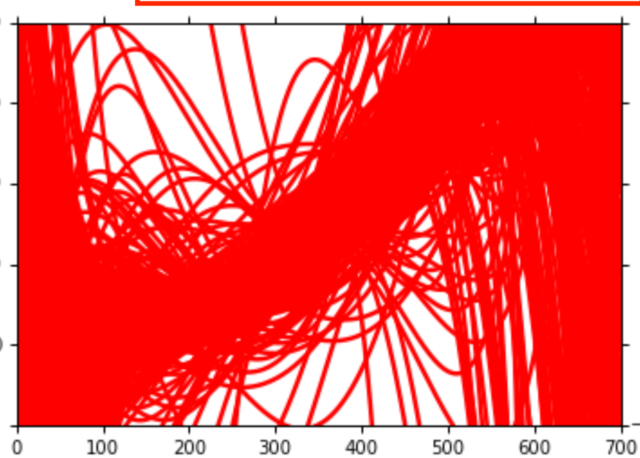


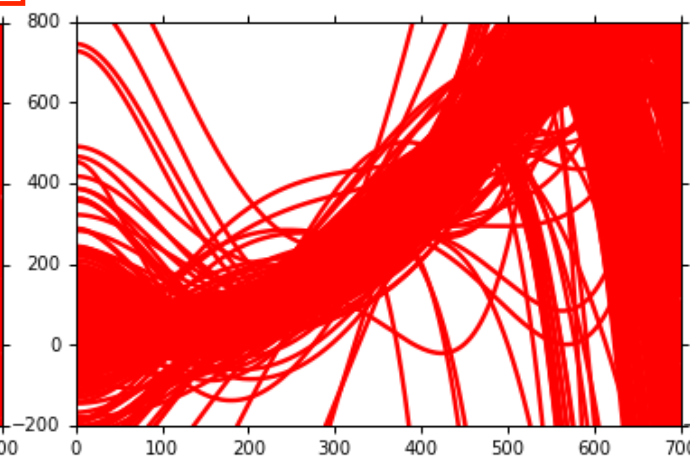
10 examples

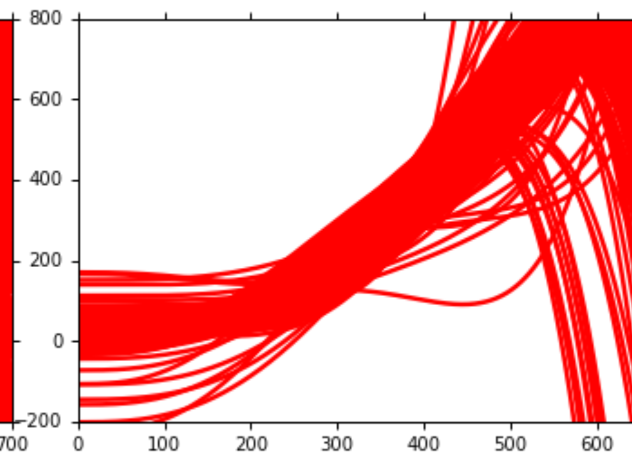


100 examples

- Regularization

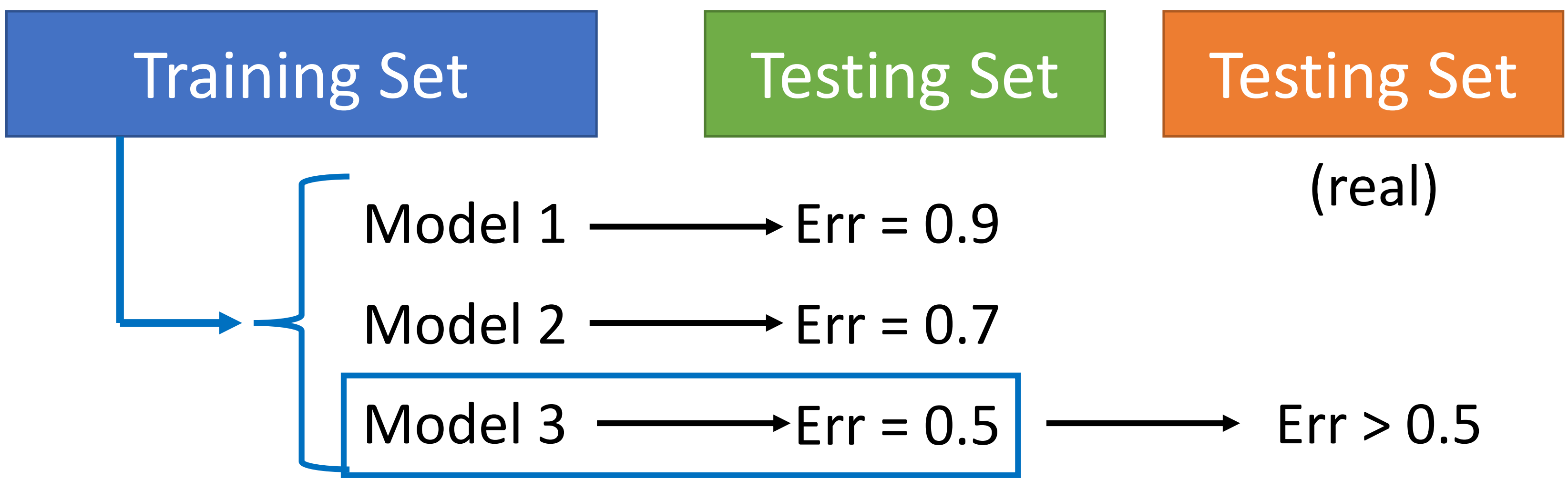






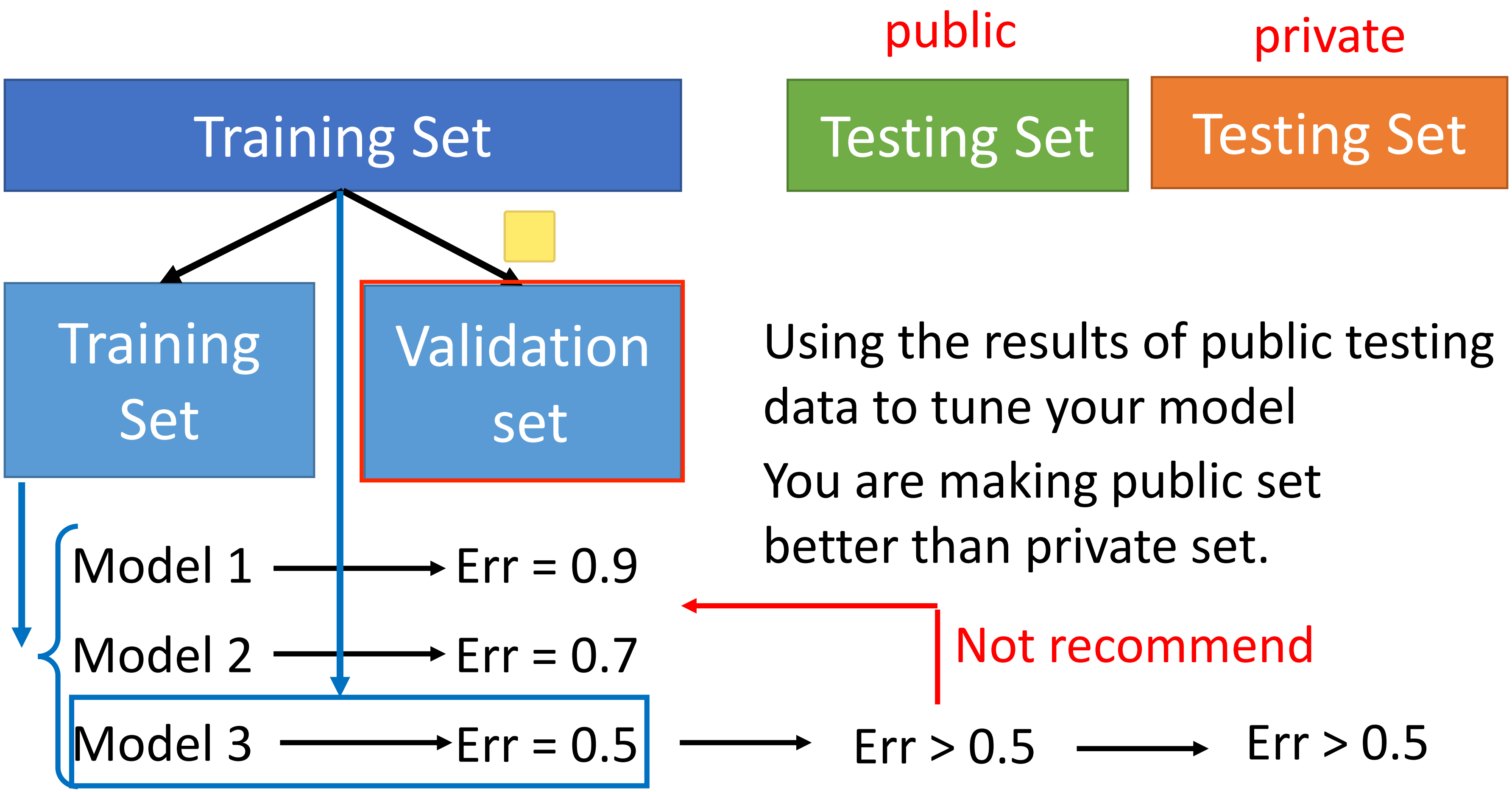
Model Selection

- 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:



利用 Cross Validation 使我們擁有的 Testing Set Error Rate 能夠反映 “Real World” Testing Set Error Rate

Cross Validation



Using the results of public testing data to tune your model

You are making public set better than private set.

將 Training Set 分成 N 份

N-fold Cross Validation