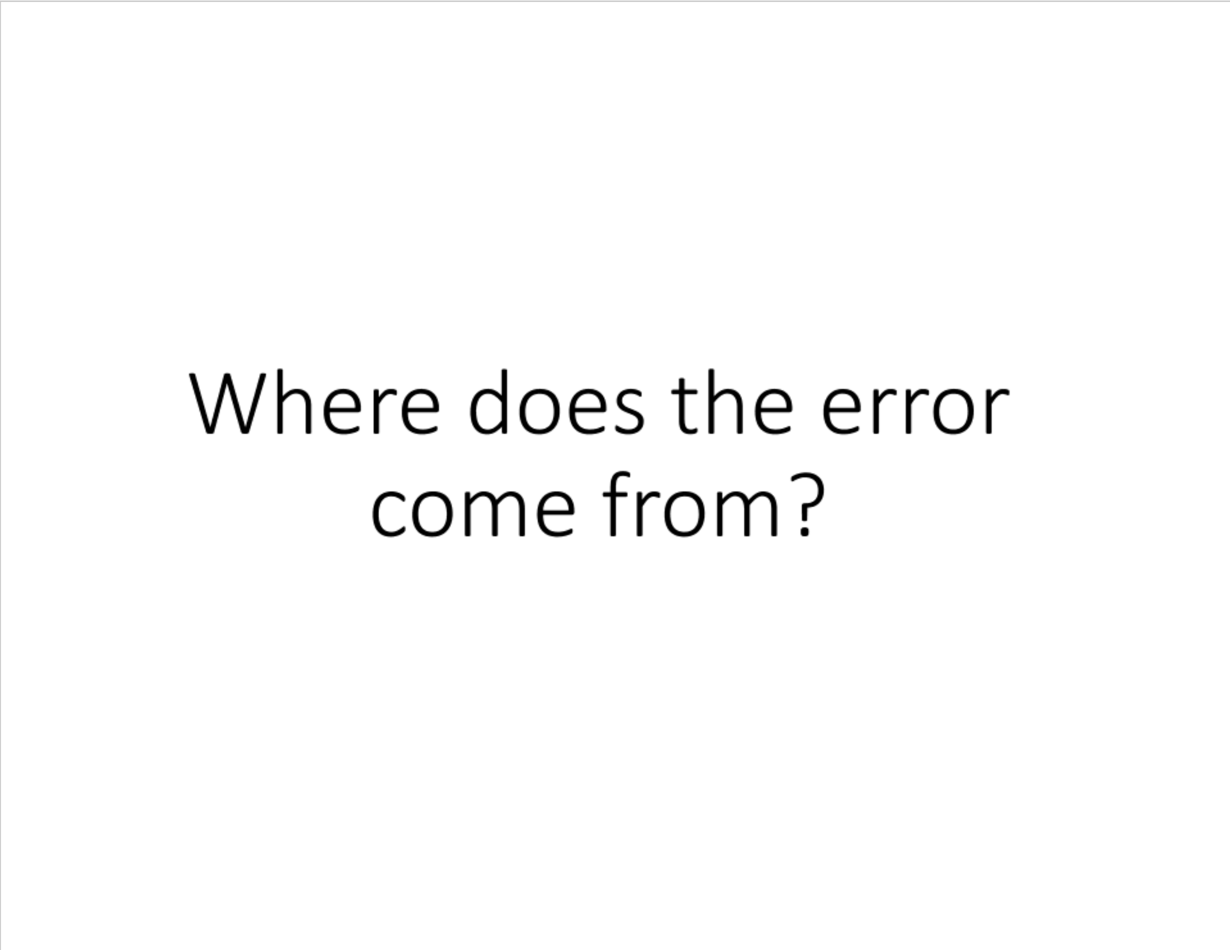
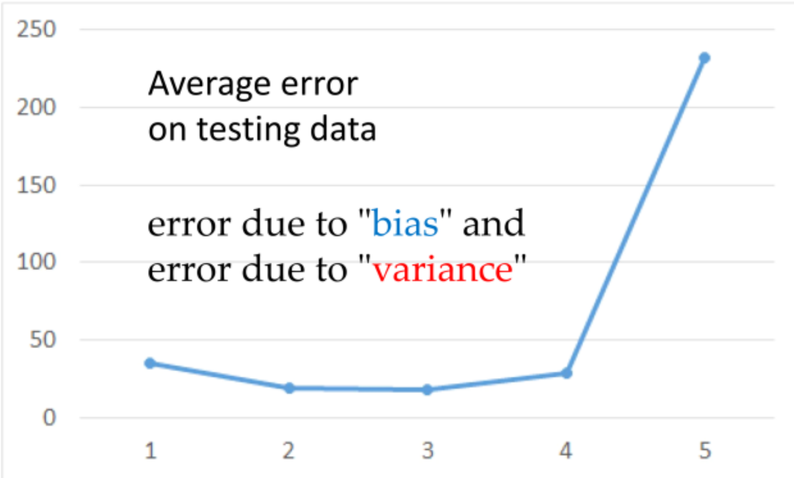
打印输出

2020年7月22日 21:06



Review



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

error来源 bias
variance

Estimator

进化后的cp值

input

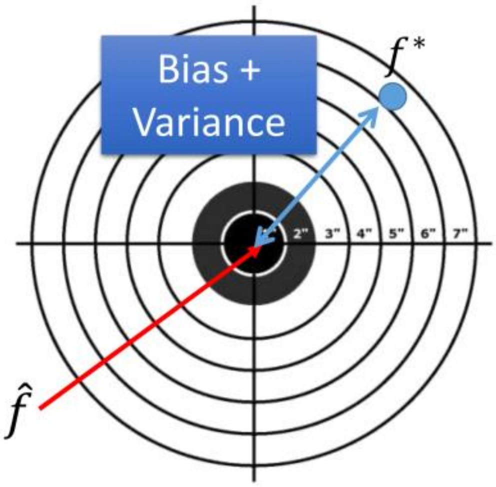
$\hat{y} = \hat{f}($  $)$

最佳的 f

Only Niantic knows $\hat{f}$

做这个的人

From training data, we find f^*



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

估测值

像是打靶

打靶

Bias and Variance of Estimator

估计

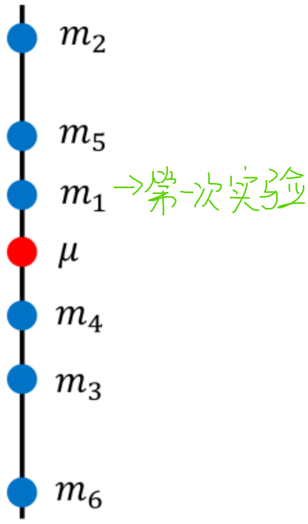
- Estimate the mean of a variable x
  - assume the mean of x is μ
  - assume the variance of x is σ^2
- Estimator of mean μ
  - Sample N points: $\{x^1, x^2, \ldots, x^N\}$

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

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

m的期望

unbiased



很多m的期望值为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 mean μ
  - Sample N points: $\{x^1, x^2, \ldots, x^N\}$

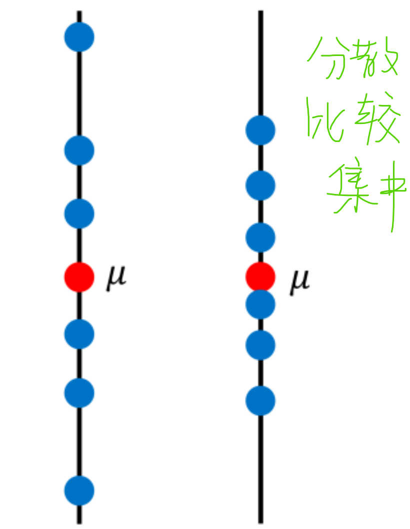
$$m = \frac{1}{N}\sum_n x^n \neq \mu$$

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

Variance depends on the number of samples

unbiased

多个N



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: $\{x^1, x^2, \dots, x^N\}$

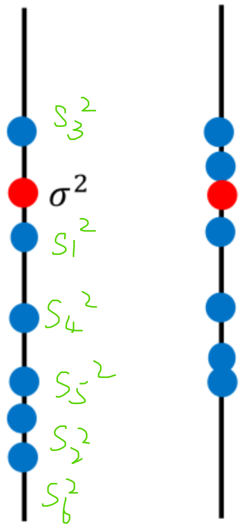
$$m = \frac{1}{N}\sum_n x^n \quad s^2 = \frac{1}{N}\sum_n (x^n - m)^2$$

Biased estimator

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

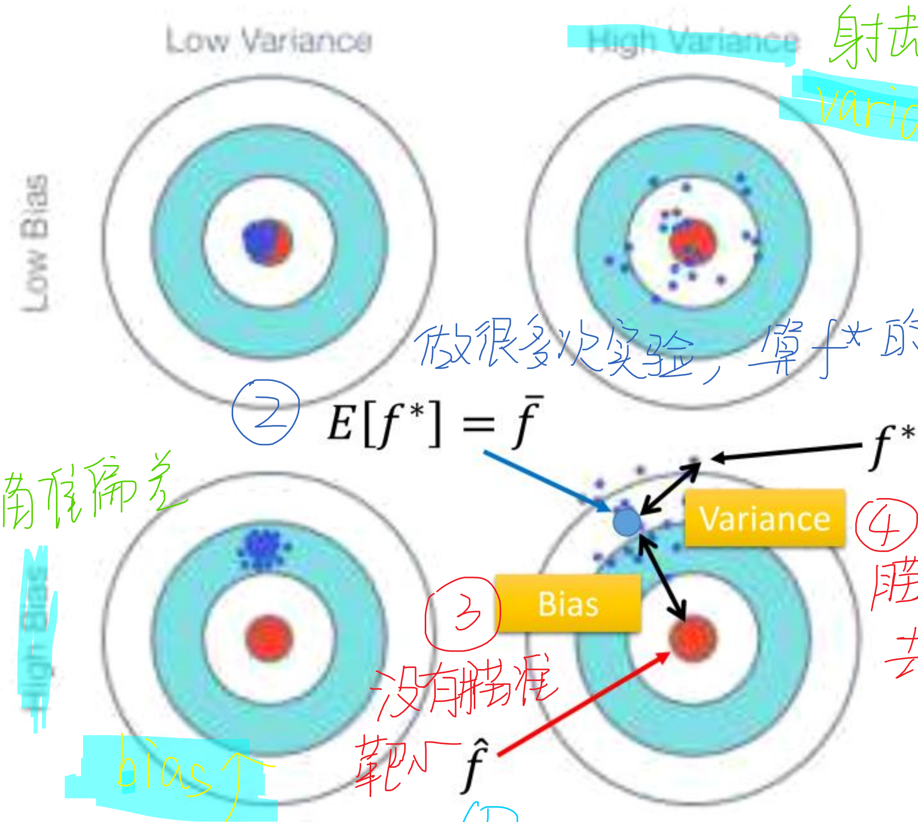
S^2估测σ^2

Increase N



N↑ S^2与σ^2之间的差距变小

S^2比σ^2小
平均而言小的次数比较多



射击偏差 variance↑

做很多次实验，算f^*的期望值

② ① ④ ③ ①

瞄准偏差

瞄准后射出去有偏移

没有瞄准靶心

bias↑

bias < Variance

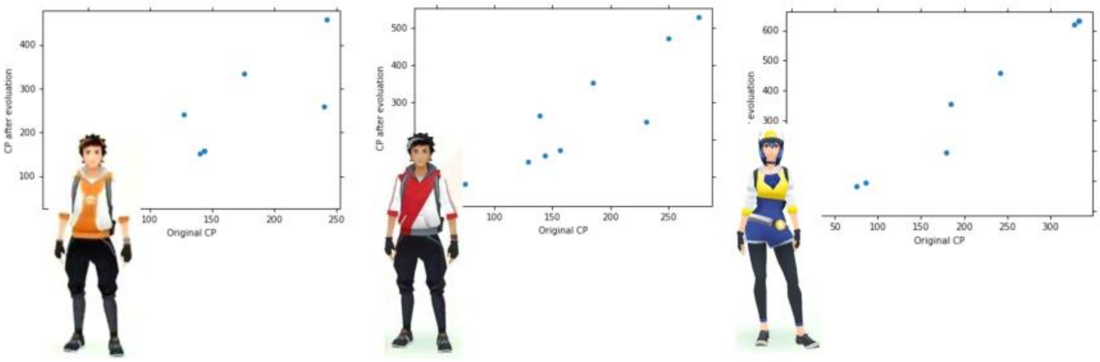
Variance

Parallel Universes

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

不同宇宙中的数据

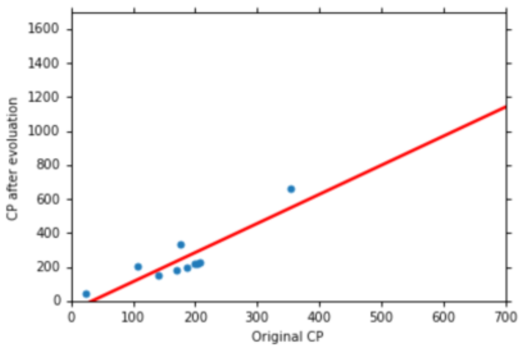
Universe 1 Universe 2 Universe 3



Parallel Universes

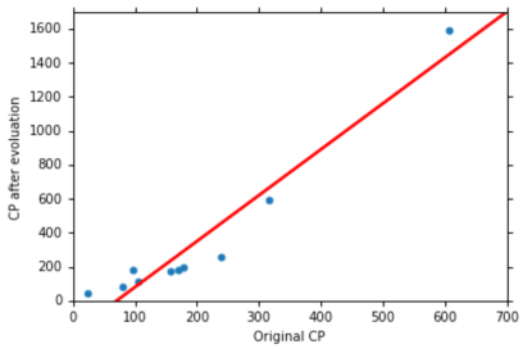
- In different universes, we use the same model, but obtain different f^*

Universe 123



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

Universe 345



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

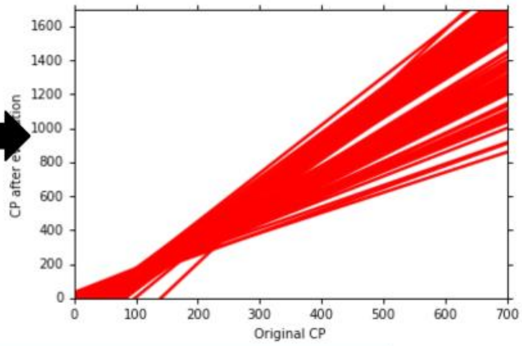
f^*的分布

做100次实验，每次抓10只

做100次实验，每次抓10只

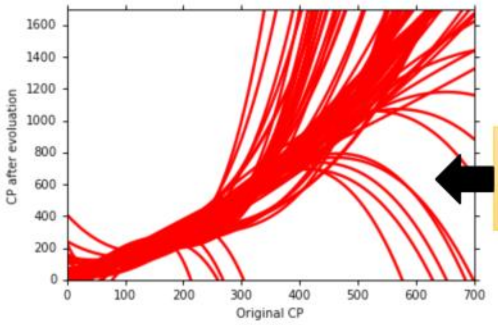
f^* in 100 Universes

① $y = b + w \cdot x_{cp}$



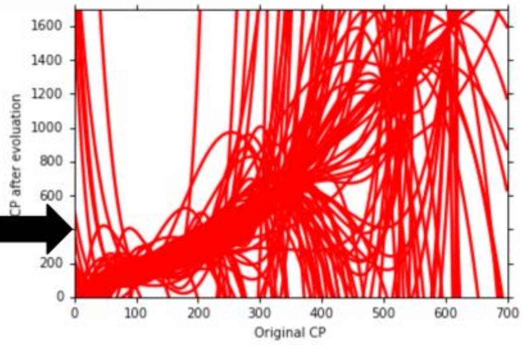
② 换model

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



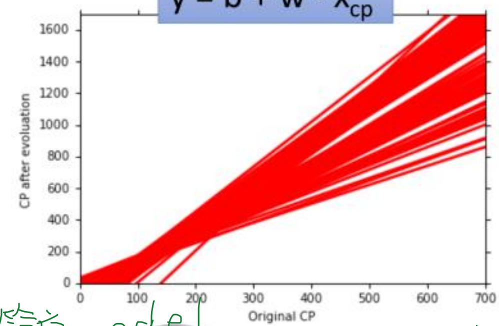
再换model

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

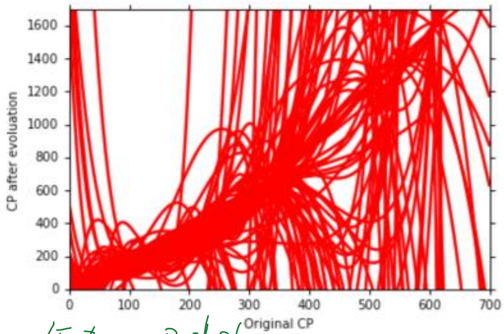


Variance

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



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



简单model

Small Variance

散布比较集

复杂model

Large Variance

散布比较开

Simpler model is less influenced by the sampled data

不太会受data的影响

Consider the extreme case f(x) = c

简单model

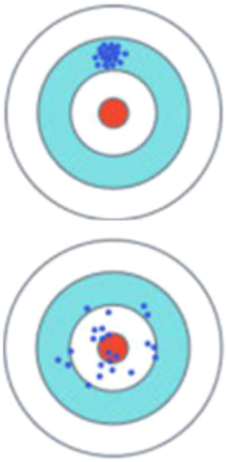
受data影响较大

Bias

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

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

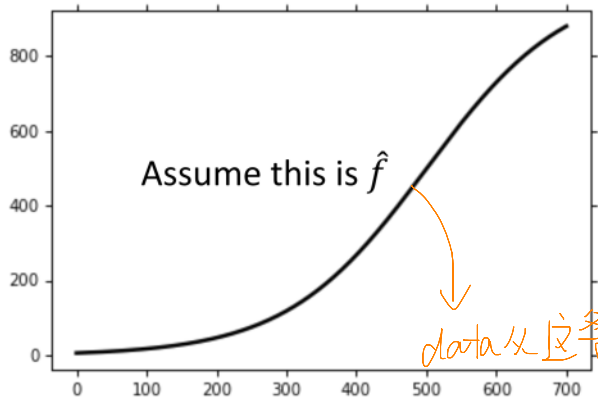
靶心



Large Bias

Small Bias

平均值和靶心很接近



data从这条线 sample出来

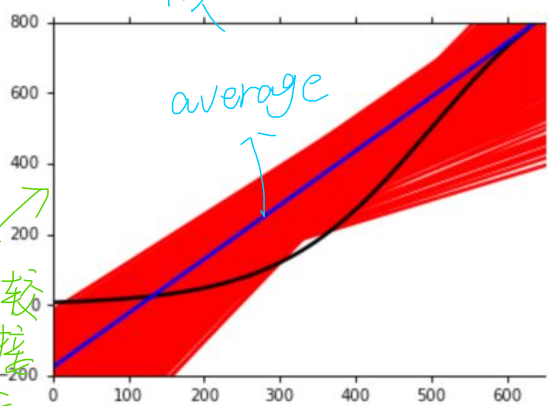
Black curve: the true function $\hat{f}$

Red curves: 5000 f^* 5000次

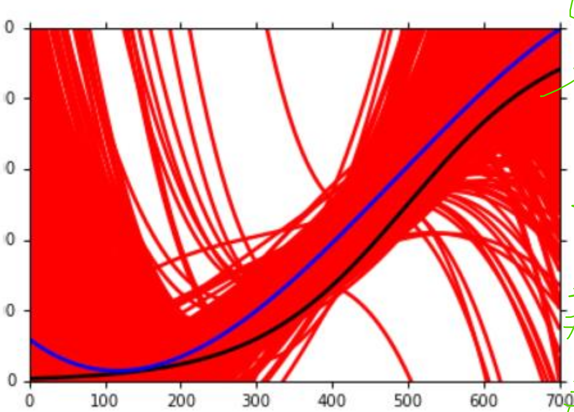
Blue curve: the average of 5000 f^* $= \bar{f}$

1次

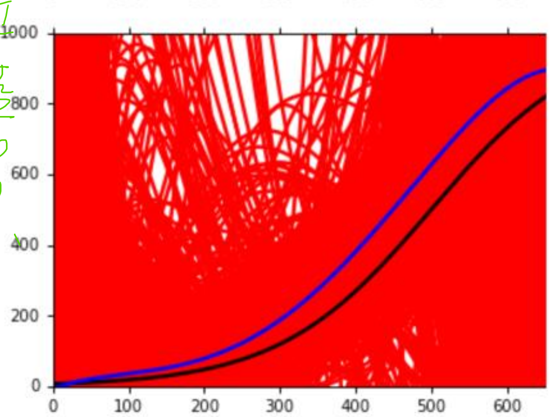
average



较接近 蓝和黑



2次

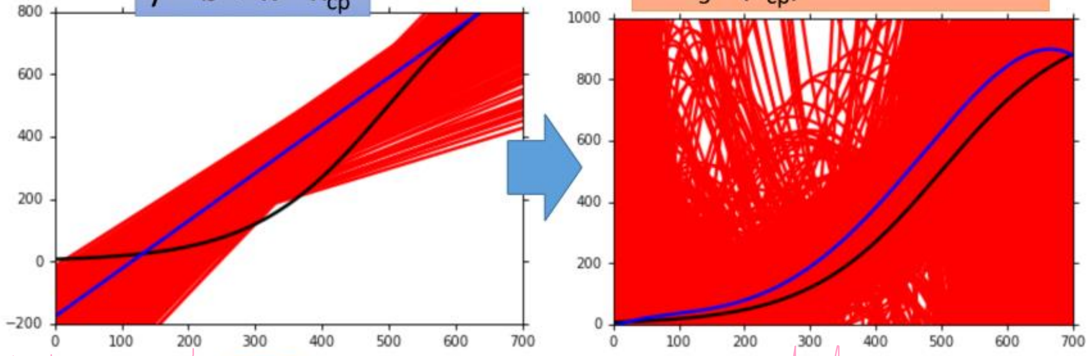


3次

Bias

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

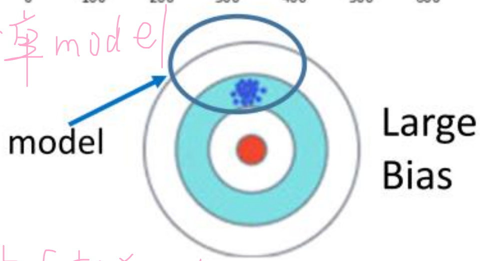
$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4 + w_5 \cdot (x_{cp})^5$



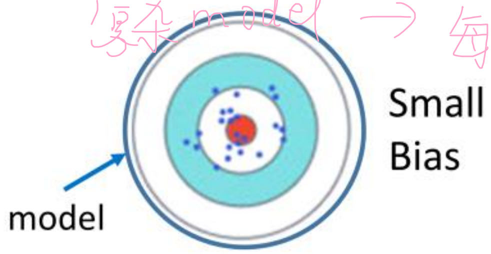
简单model

model

每次打都差不多，集中某处 远离靶心



复杂model → 每次打差别较大，但在靶心周围



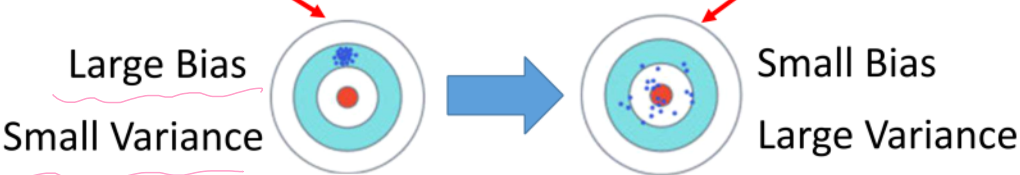
Bias v.s. Variance

Variance ↑ 瞄准 射得远

bias大

variance大

bias↓ 瞄得准

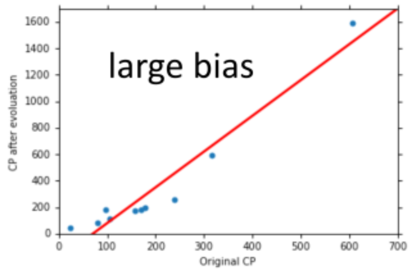


简单

复杂

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



不能fit

和正确model有距离

training data是可以的 但test data error大

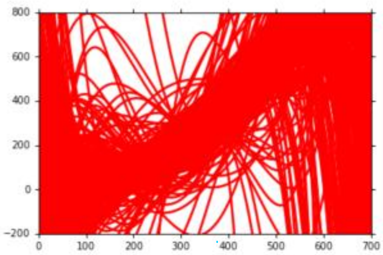
收集更多data

model不好

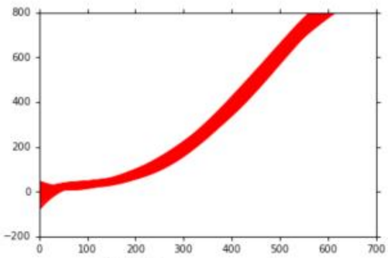
What to do with large variance?

- More data

Very effective, but not always practical



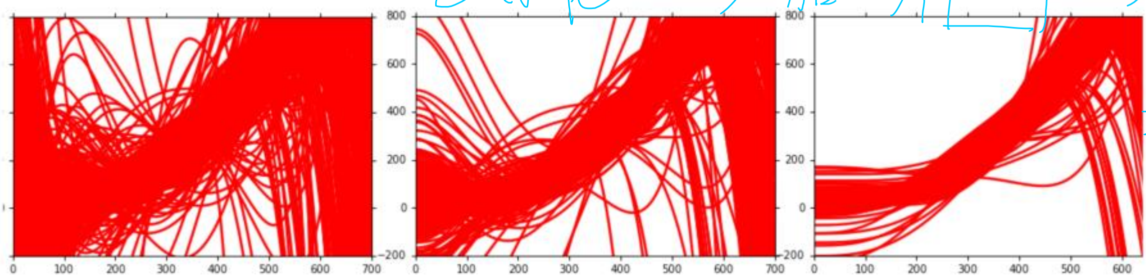
10 examples



100 examples

- Regularization

正则化 → 加入 → 使曲线更平滑



Model Selection

权衡

- 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:

自己手上的 有自己的 bias

≠

真正的是不知道的

Training Set | Testing Set | Testing Set (real)

Model 1 → Err = 0.9

Model 2 → Err = 0.7

Model 3 → Err = 0.5 → Err > 0.5

Homework

我有 — Training Set

我有 — 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 next Friday?

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

Cross Validation

将traingset分成两组

Training Set	public Testing Set	private Testing Set

② Training Set | Validation set

① train model

Model 1 → Err = 0.9

Model 2 → Err = 0.7

Model 3 → Err = 0.5 → Err > 0.5 → Err > 0.5

选model

Using the results of public testing data to tune your model ③

You are making public set better than private set.

Not recommend

看起来比training的err大

但反映的是private的error

①TraingSet train model　Validationset 选 model

②选出model 3　将所有Traing set 用于model 3

★担心分坏

N-fold Cross Validation

① 分3份的情况用于这些model

Training Set

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

② Testing Set (public) | Testing Set (private)

少在意public上的error 进而调整model

说不定private比较糟糕 test

说不定 private

Reference

- Bishop: Chapter 3.2