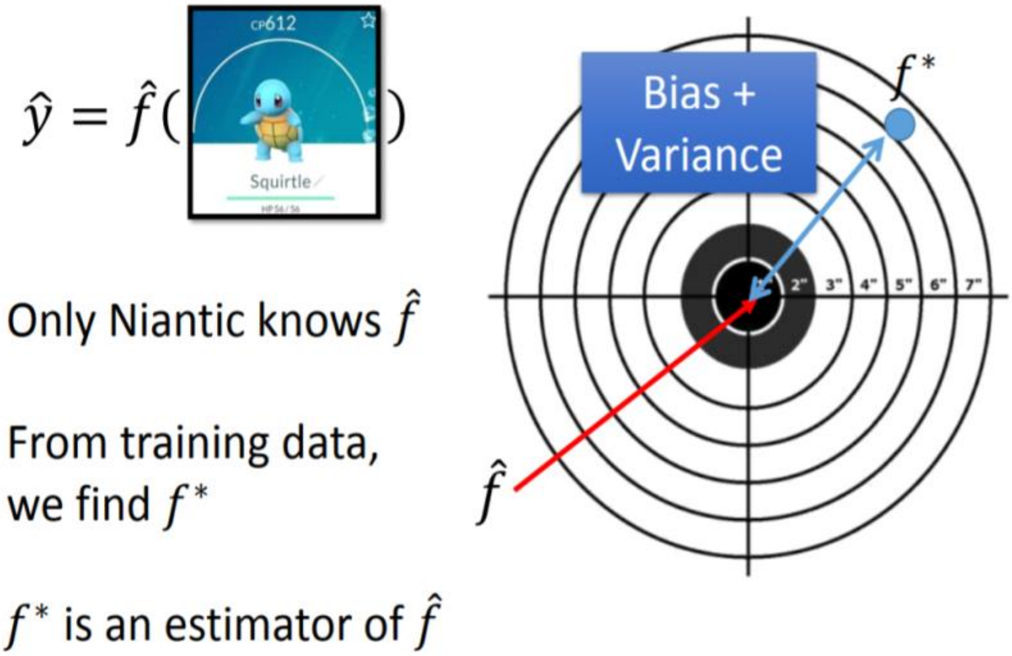
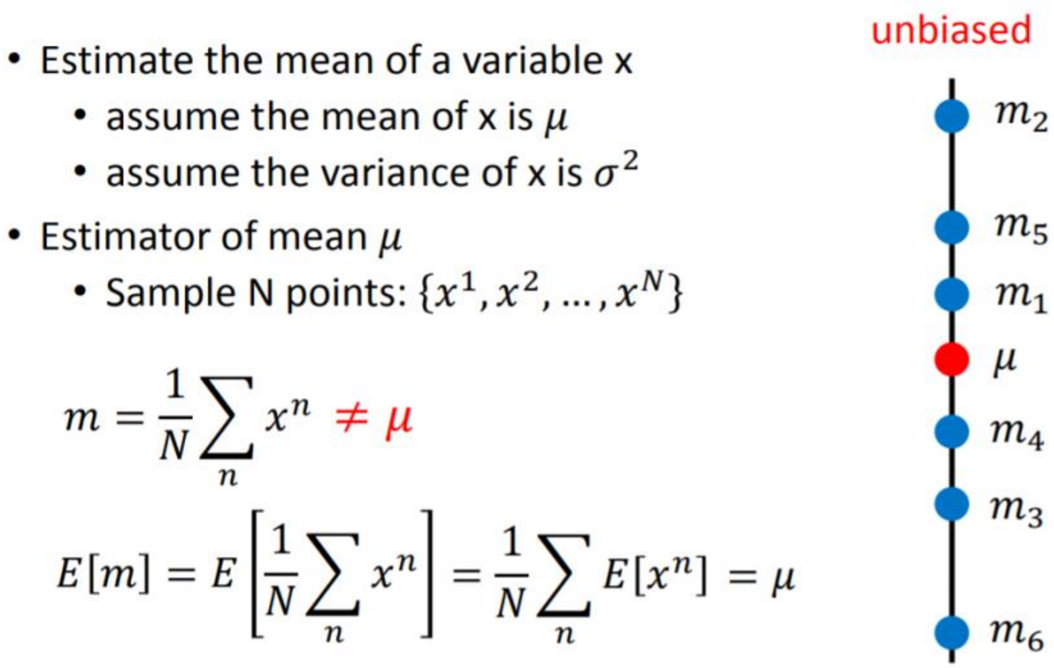
Estimator

2018年12月21日 15:41

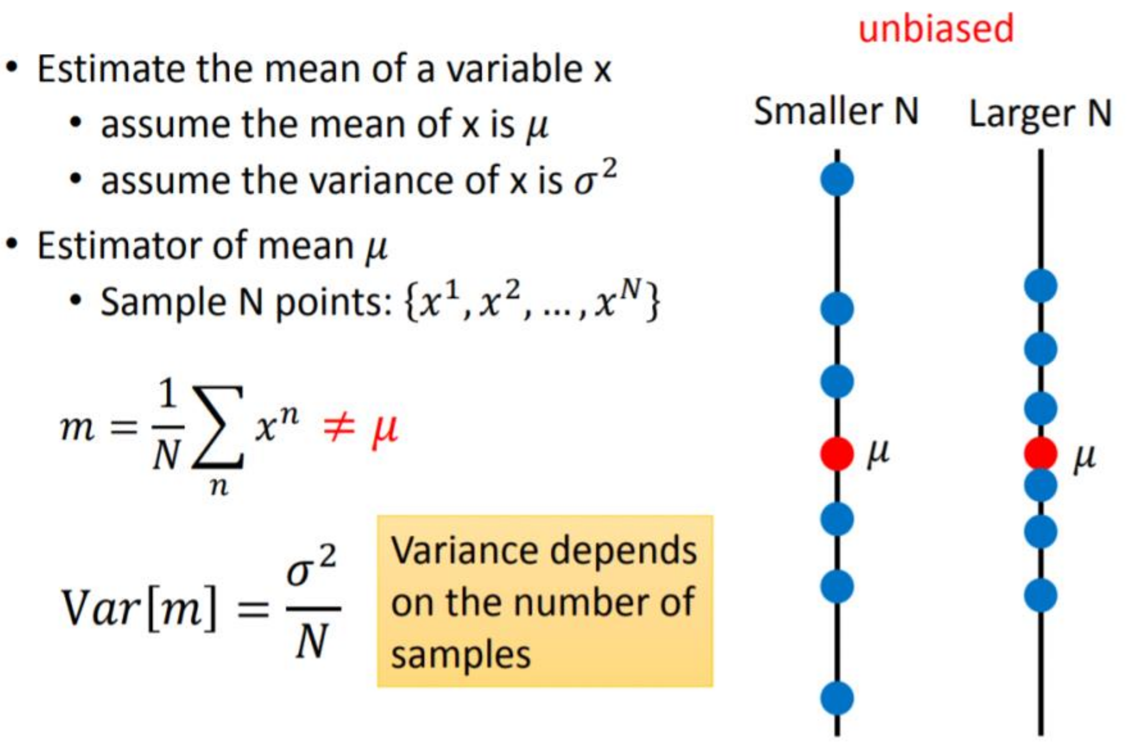
有两种Error一种来自于Bias另一种来自于Variance。



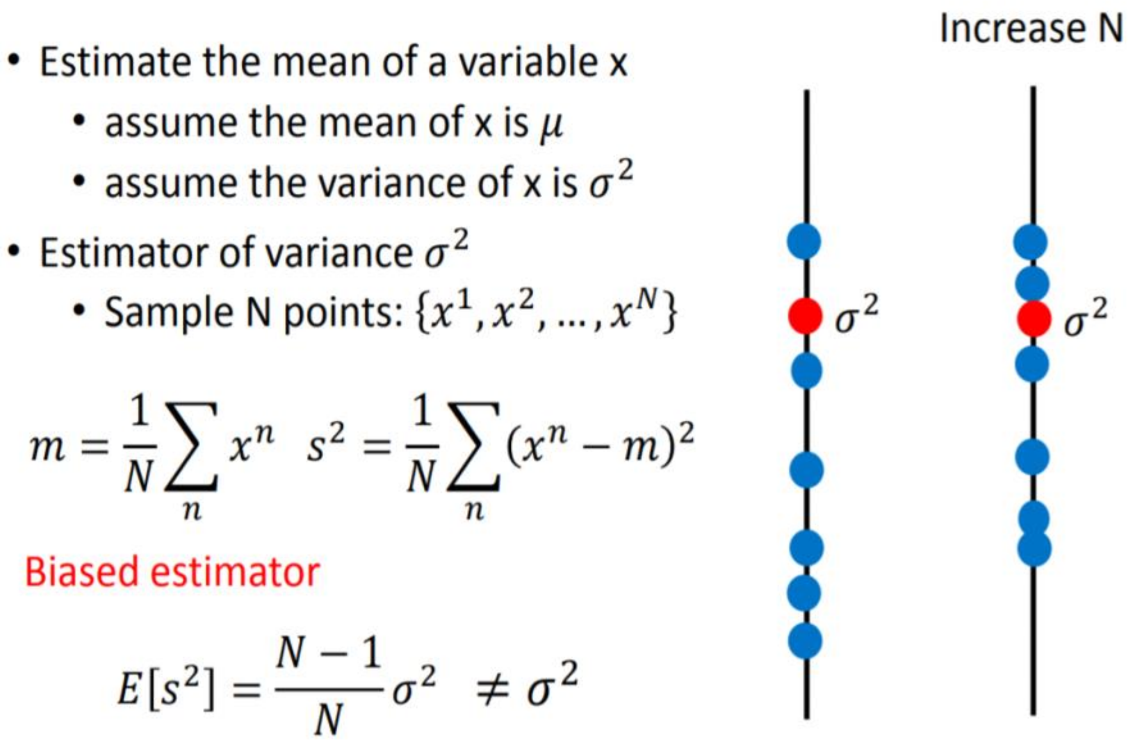
f_hat 是最好的方程, f_star 是我们从数据中求出的方程,用打靶的方式理解，Bias和Variance就是两个靶点(方程)之间的距离。



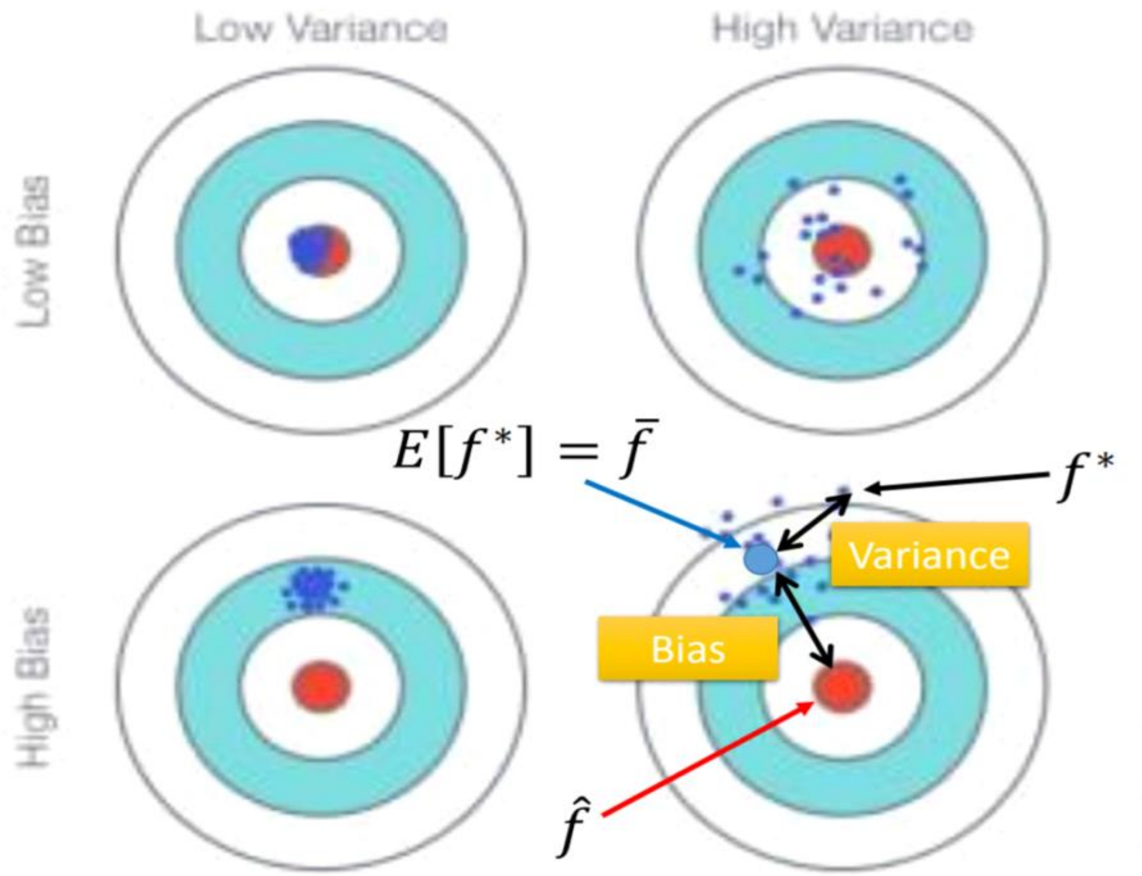
m的平均值不等于μ，但如果找了很多m，E[m]=μ说明方程是unbiased，说的形象一点就是这些点都是瞄着μ点去打的但是由于某些机械故障产生了偏移，那么究竟偏移了多少呢？就要用variance来算了。



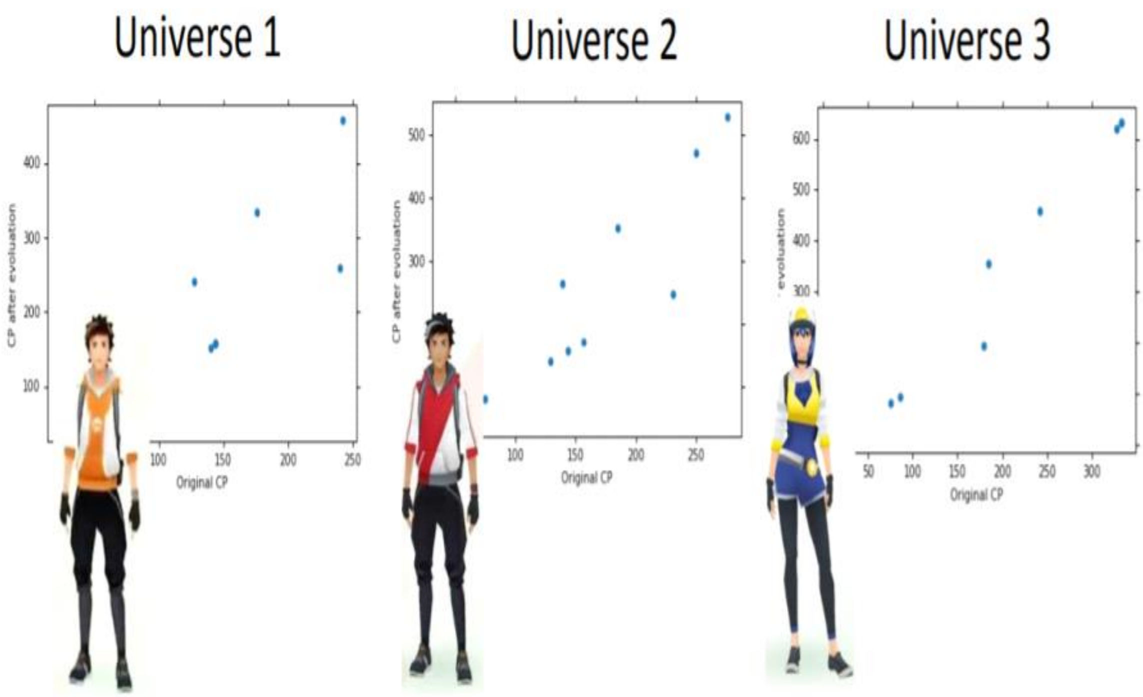
variance的求法如下图：



方程式Biased的，随着N的增大E[s^2] ≈σ^ 2



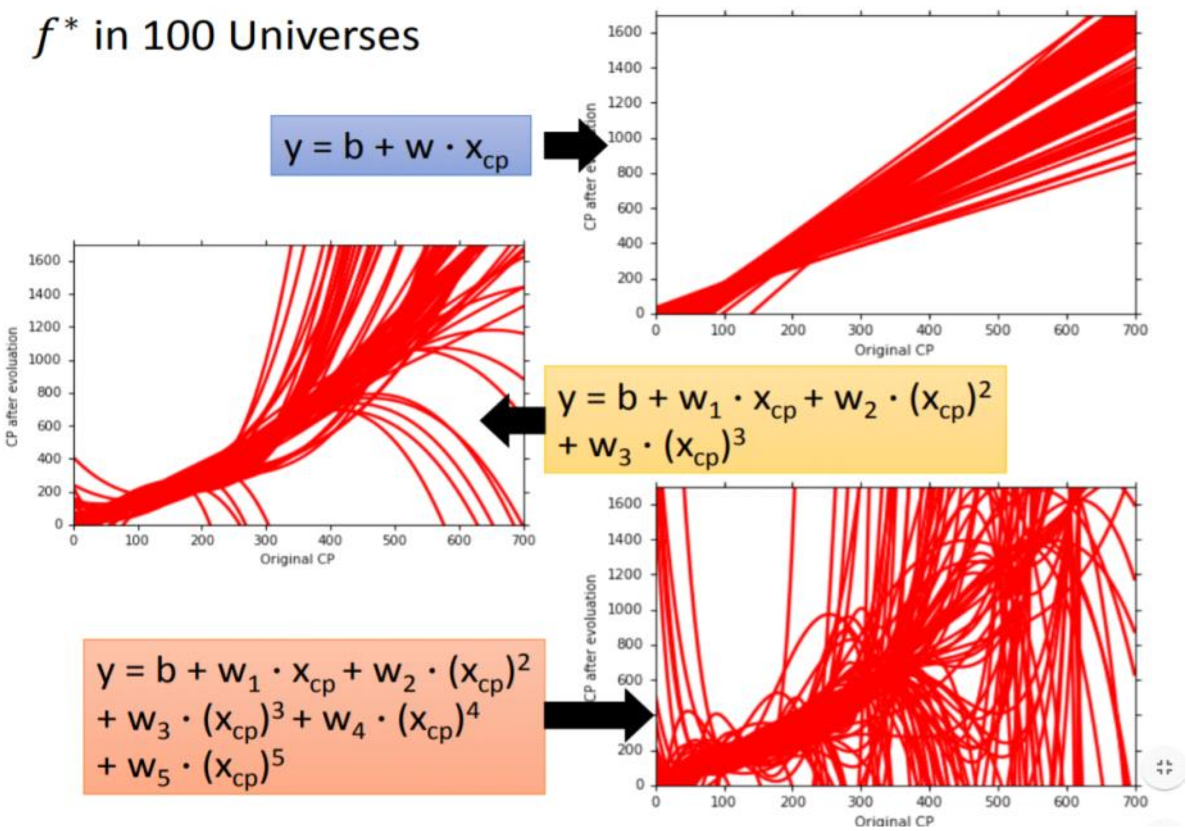
距离就是Bias，分散程度就是Variance，一个模型是如何生成这么多点的呢，可以用平行宇宙的思想来考虑。



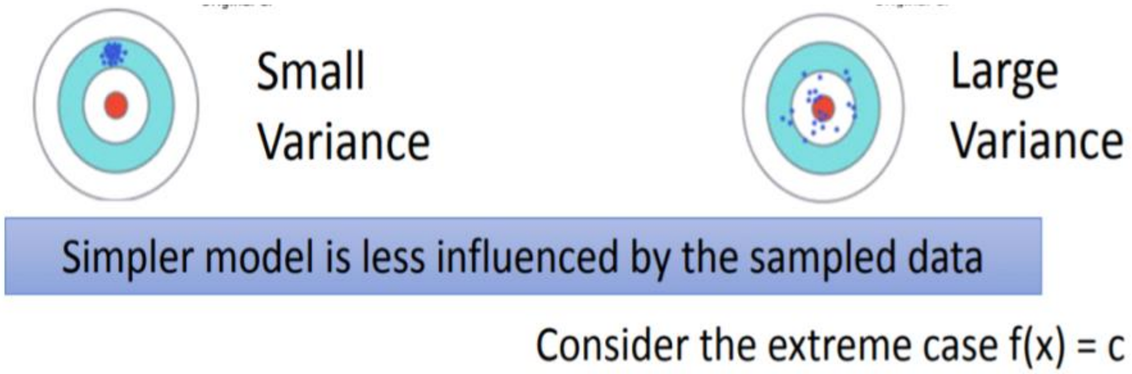
Variance

2018年12月24日　　16:11

在三个宇宙中分别抓到了10只不同的宝可梦,以100只为例：



model越复杂，函数的分布就越开，它的Variance就越大。



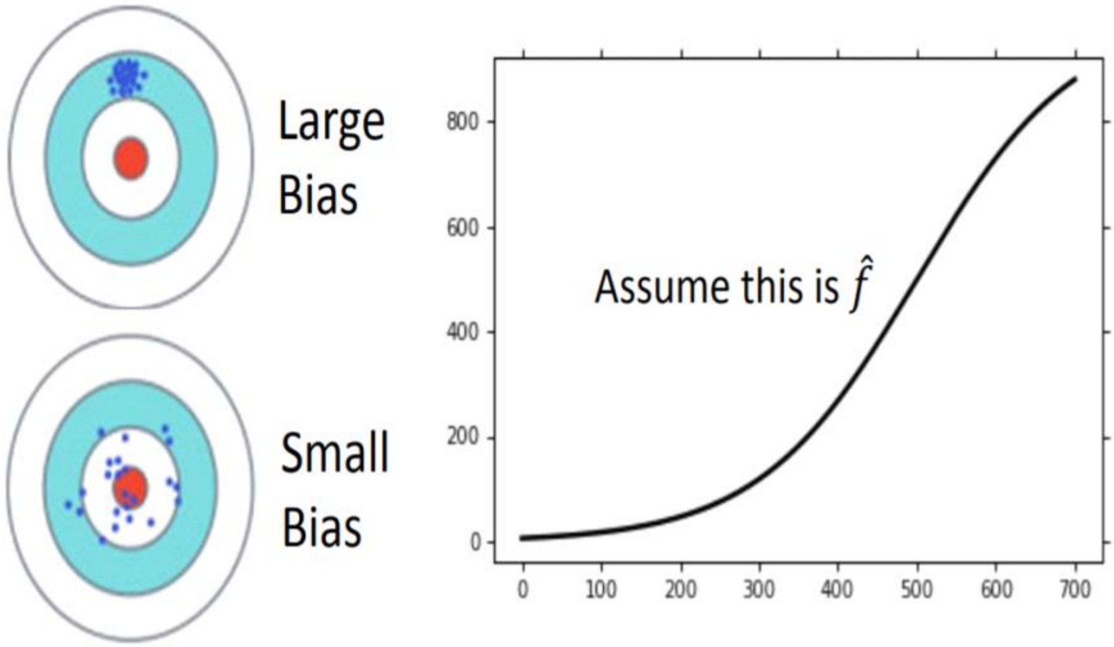
那么为什么模型越简单，分布越集中呢？因为简单的模型受到data的影响很小。

Bias

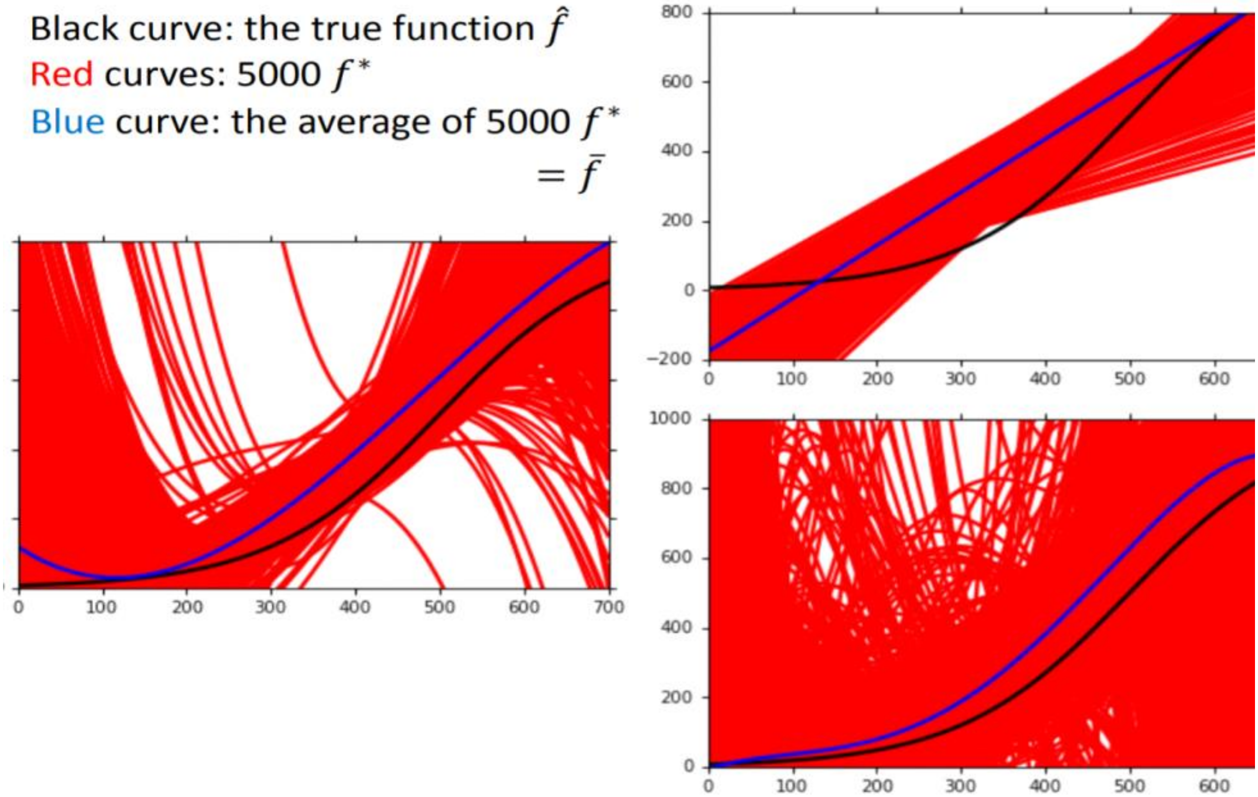
2018年12月24日 16:12

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

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



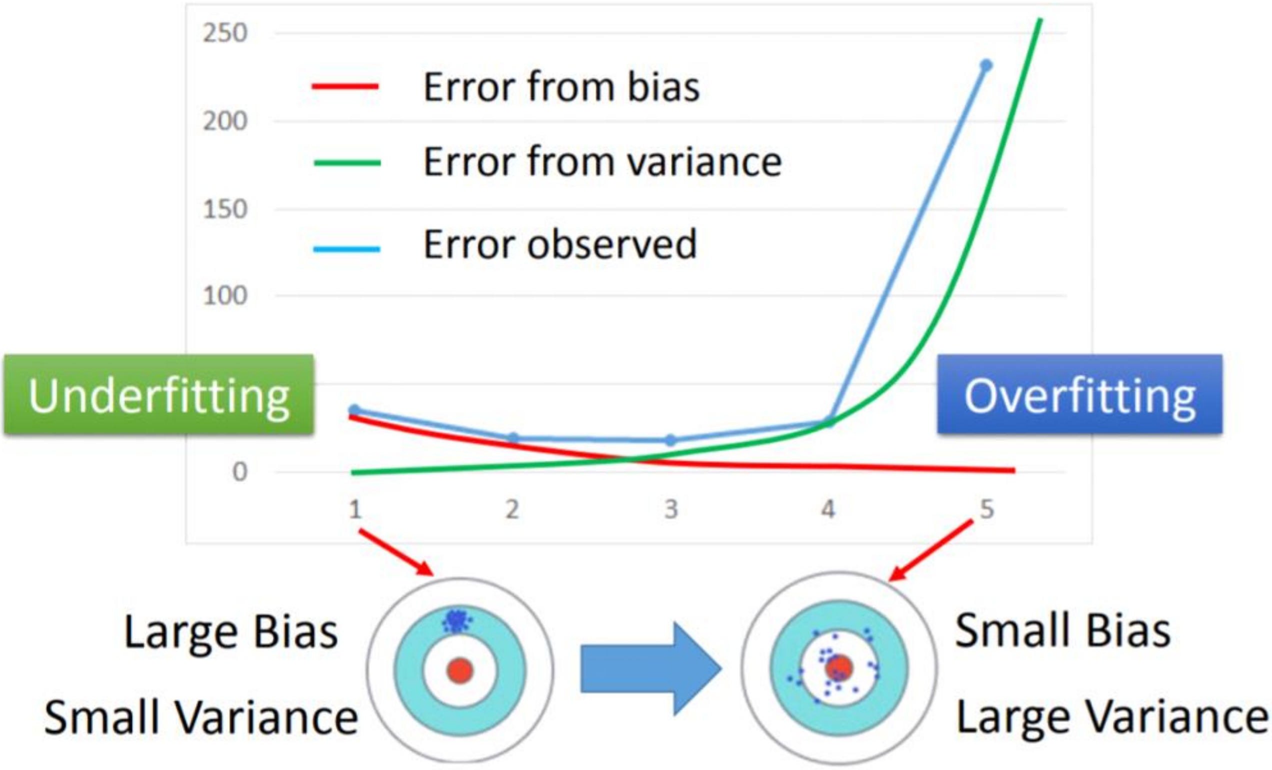
Bias在图中应该如何测量呢？假设黑色的线就是要找的function，红线是用黑线生成5000次样本训练出来的，蓝线代表红线的平均值。看样子是模型越复杂Bias越小Variance越大。


Black curve: the true function $\hat{f}$
Red curves: 5000 f^*
Blue curve: the average of 5000 f^*
$= \bar{f}$
0
100
200
300
400
500
600
700
800
600
400
200
0
−200
0
100
200
300
400
500
600
1000
800
600
400
200
0
0
100
200
300
400
500
600

Bias v.s. Variance

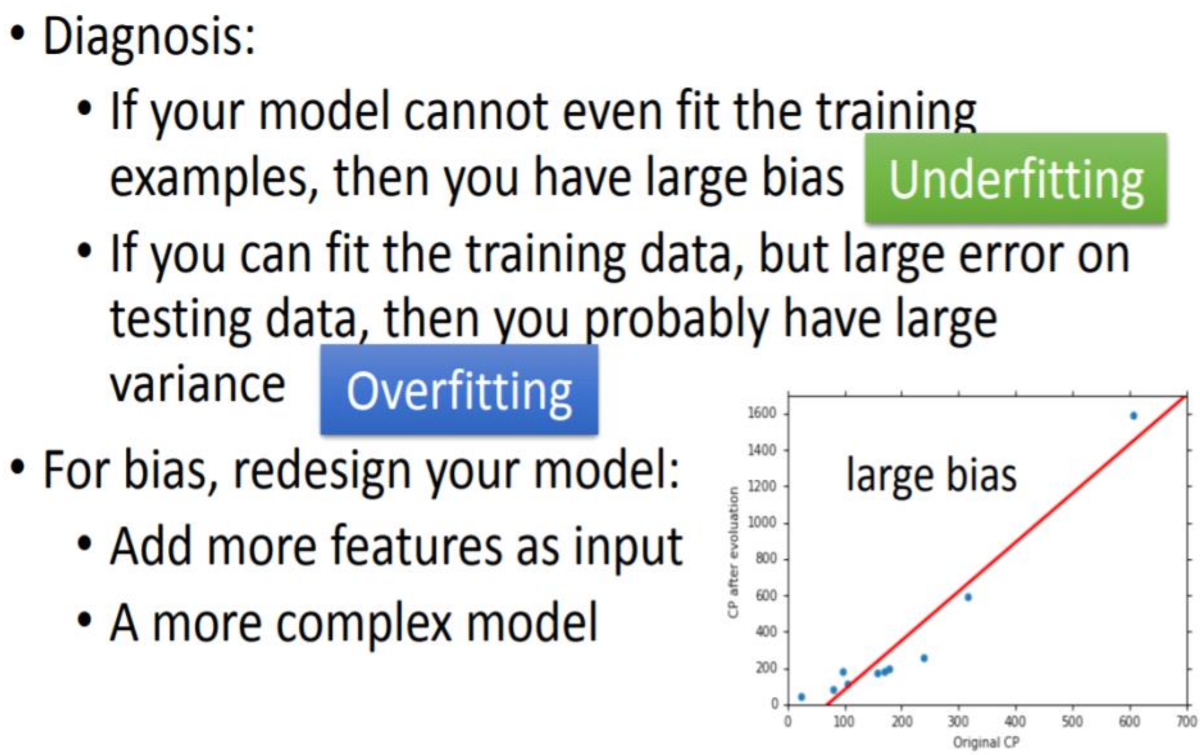
2018年12月24日　　16:26

从全局来看的整个图表：一开始模型比较简单所以Bias大，Variance小，到后来模型复杂后Bias就变小了而Variance就变大了。根据模型复杂程度由小到大可以把它想成从一堆远离中心点的变化到靠近中心分散的点。



What to do with large bias?

2018年12月24日 16:39

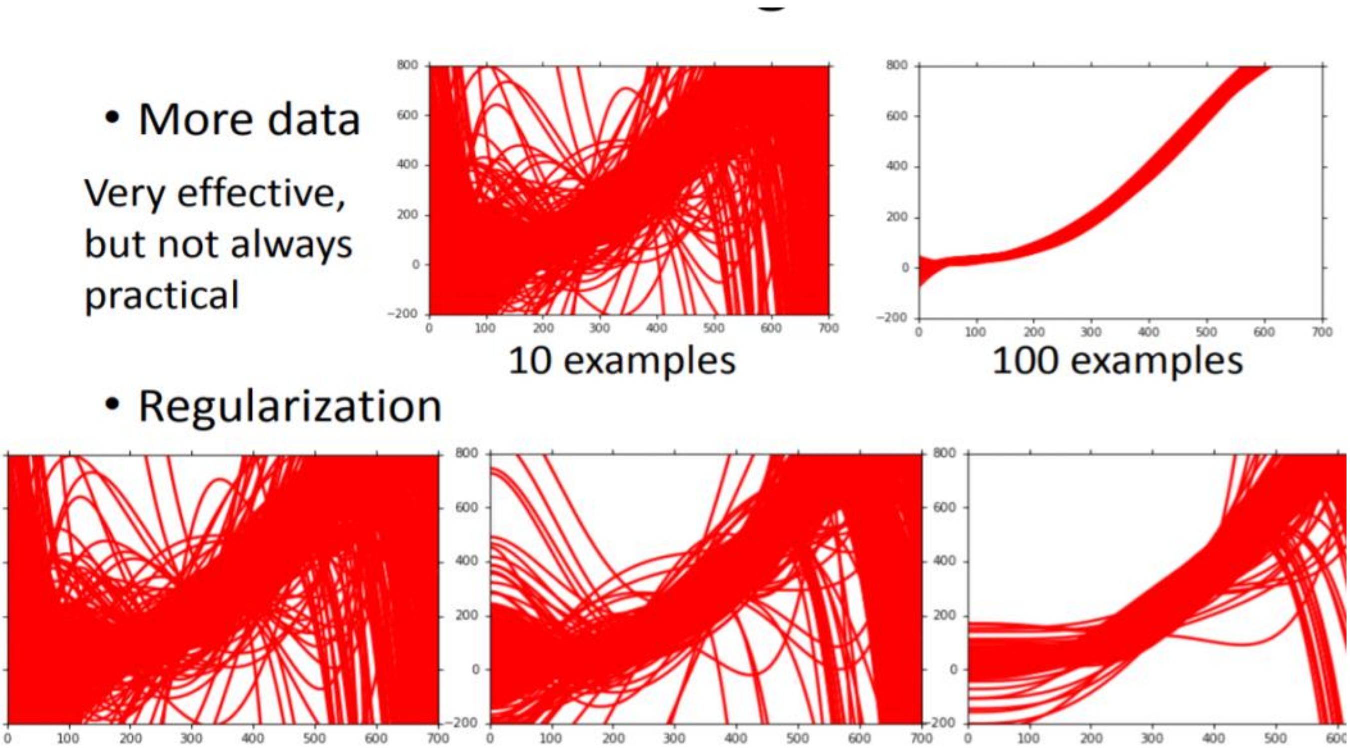


如果不能拟合 training data 就是underfitting，如果training set上表现很好但是testing set上表现很差那就是overfitting。

处理underfitting就是 使用更复杂的模型，使用更多的feature

What to do with large variance?

2018年12月24日 16:39



处理large variance(过拟合)：增加更多的数据(可以使用数据增强)和正则化（会影响Bias）

Model Selection

2018年12月24日 16:48

我们希望Bias和Variance可以达到一种平衡状态。

Cross Validation最后会用全部的training set训练一遍。