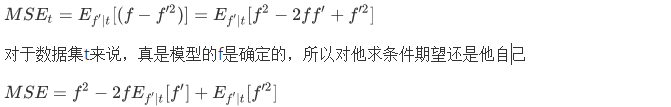
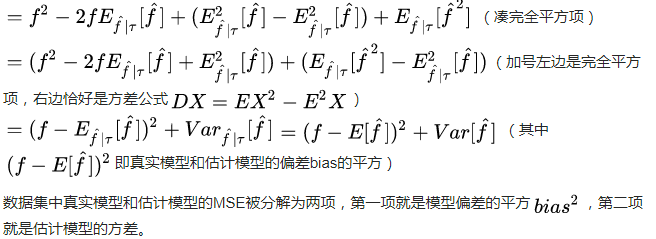
**偏差，方差，误差的来源：**

Error = Bias2 + Variance(+Noise)

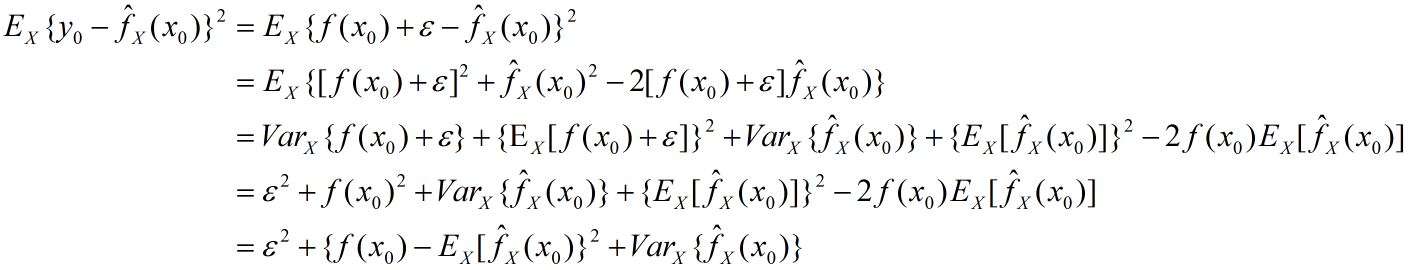
对于给定的数据集合t，有唯一的真是模型f，但是我们是不知道的。我们用不同的模型f’取逼近f，得到的均方误差MSE是不一样的(其他误差的推到要复杂)，可以把这个MSE和f‘都看作给定数据集合t下的随机变量

那么我们有

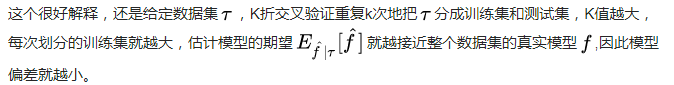




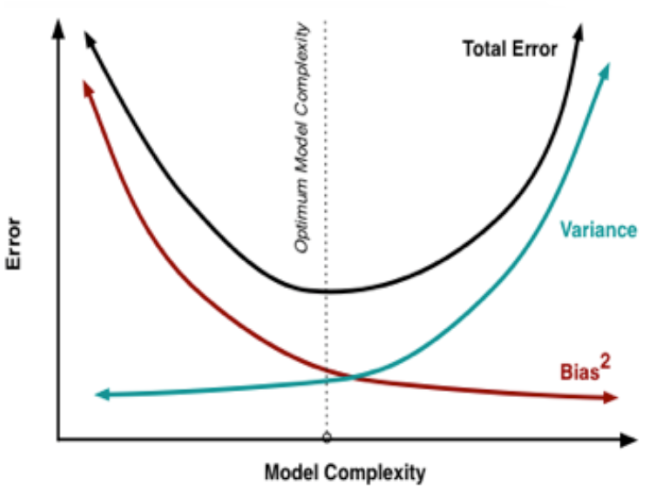
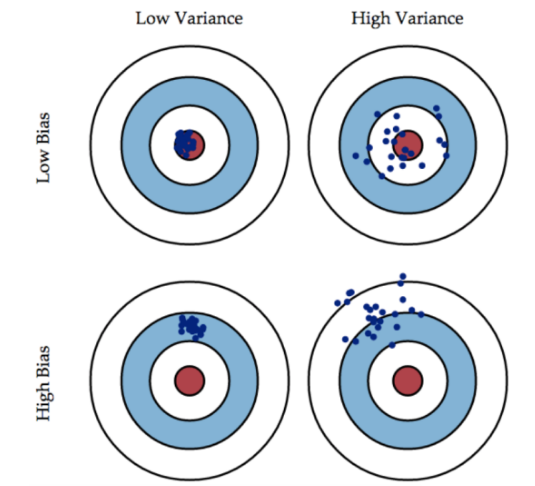
可以再看一个带误差的推导公式



**K-fold Cross Validation中的k值会影响偏差和方差**



**Bias，Variance**



**Bias（准）**：

描述的是根据样本拟合出的模型的输出预测结果的期望与样本真是结果的差距，也就是说这是一种拟合度或者精确度的表现，针对的单个模型来说的。要想Bias好（即low bias），那就需要复杂化模型，增加模型的参数，这样容易过拟合。

**Varience（确）**：

描述的是样本训练出来的模型在测试集合上表现，即一种泛化性能或者稳定性的表现，针对的是多个模型来说的。要想在varience上表现好，那么就需要简化模型，减少模型的参数，但这样很容易欠拟合，导致高bias。

**当然了，我们还可以从另外的一个角度取理解这两者之间的关系：**

我们训练模型的目的是为了让模型在测试数据集合上拟合的效果更好一些，也就是Error(test)比较小，但是在实际的问题中我们是不知道test data的，也根本不知道test的内在规律是什么。所以我们会通过一下的两个步骤来实现Error的最小化

分为两步：

1. 让Error(train)尽可能小

2. 让Error(train)尽可能小<=>Error(test)

那么可以根据三段论，因为A小，而且A=B，这样B就很小

最小化Error(train)🡪把模型复杂化，参数搞得很多(比如适用十元的线性回归模型肯定要比二元的线性回归模型的性能好呢) 🡪low bisa🡪让Error(test)=Error(test)? 🡪让我们的模型没有偏见🡪对局部数据不敏感🡪简单模型🡪low varience

**本质上的理解：**

从上面可以看出，我们的B和V是不可兼得的。如果要降低模型的Bias，就会一定程度上提高Variance，反之亦然。造成这种现象的根本原因是我们总是试图适用有限的训练样本估计无限的真实数据。

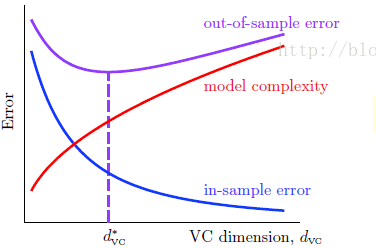
当我们更加相信这些数据的真实性，而忽视了对模型的先验知识，就会尽量保证在训练样本上的准确度，这样可以减少模型的Bias。但是，这样学习到的模型很可能会失去一定的泛化能力，导致过拟合，降低模型在真实数据上的表现，增加模型的不确定性。

相反，如果我们更加相信了模型的先验知识，在学习模型的过程中对模型增加更多的限制，就可以降低模型的Variance，提高模型的稳定性了，但是也会使得Bias增大。

**从VC维上理解Bias和Variance：**

针对特定的数据集合，我们由于为了提高在训练数据集合上的准确度，那么就要把测试集合上的数据考虑周全，那么可能一些不太靠谱的数据点也会被我们考虑进来，这些点很有可能带来新的vc纬，随着新的vc纬的增加，相应的模型也就会变得复杂起来，根据我们的vc纬的理论，bias肯定就很小了，那么test误差就会变得很大。所以为了更好的泛化性能，我们要降低模型的复杂度。

其实也是很容易理解，在训练开始的时候，我们bias很大，应为模型还没有来得及学习呢，也就是与真实模型的差距很大。然而此时variance却很小，因为训练数据集还没有来得及对模型产生影响呢。



随着训练的进行，bias变小了(可以暂时理解为in-sample error)，因为模型学到东西了。如果训练的时间太长，那么就会导致除了学习到真实模型的信息外，我们还学习到了许多具体的，只是针对我们使用的训练数据集合的信息。但是不同的训练数据集之间的某些特征是不同的，噪声也是不一致的，这就导致了我们的模型在很多其他的数据集合上的无法获取得到很好的效果，即过拟合了（可以看到我们的in-sample error和out-sample error之间的差距越来越大了，就是随着时间的推移，我们的模型变得复杂了，例如：神经网络就会这样）

**关于k-fold cross validation对Bias和Variance的影响**

首先是Variance，对于一个数据集合，假设只有一个点是异常的，如果只去训练一个模型，那么这个点会对整个模型带来影响，使得模型学习出来后有很大的Variance。但是如果采用了k折交叉验证之后，只有一个模型会受到这个异常数据的影响，而其余的k-1个模型都是正常的，那么在平均之后，这个异常数据的影响就会被平摊下来，大大减少了variance。

相比之下，bias是直接建模的，只需要保证模型在训练样本上训练误差小就行了，而要达到这个目的，我们就必须把所有的数据一起训练，才能达到最优的解嘛，但是k折交叉验证目标函数破坏了这个情形，所以bias会增大。

好了，我们大概的了解了一下k折交叉验证中的b和v，那么，我们来理解一下这里的k对于b和v的影响有哪些吧

当k较大的时候，那么我们的训练数据集就会比较大，此时bias自然就会很小了，可以发现，我们的k-1个训练集合之间很大部分都是相互重合的，相关性比较高，那么，很容易导致variance变高。