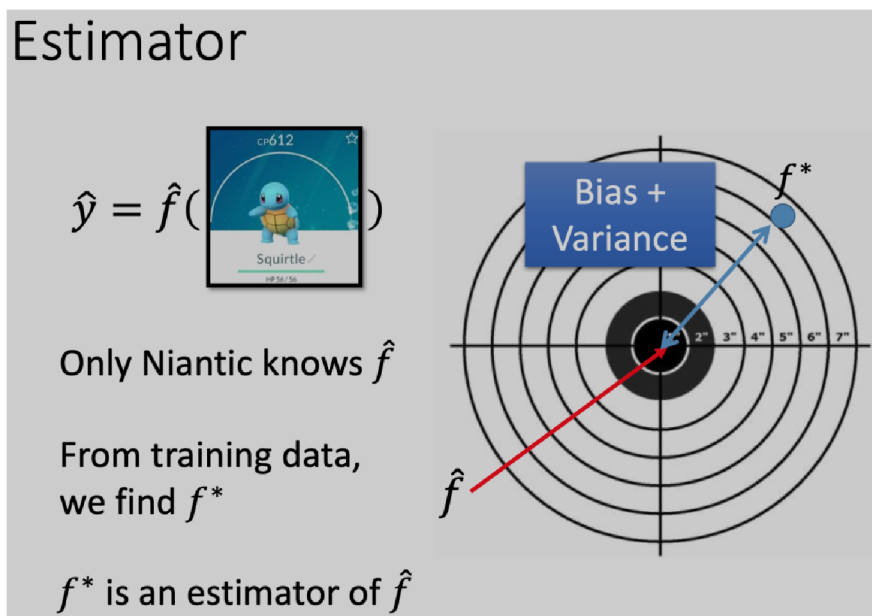


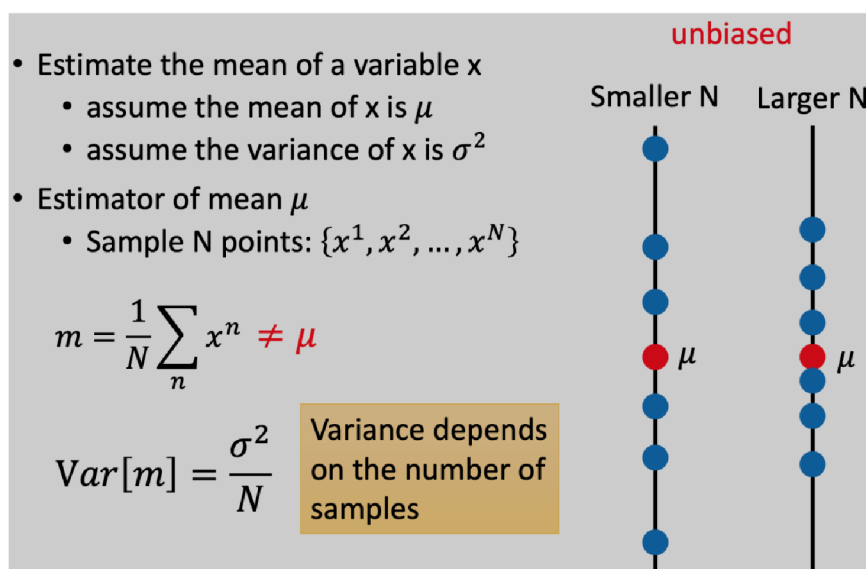
模型预测不准？这两项指标你评估了吗？

原创 多杰平措 多杰平措OPL 2020-04-09 22:25:53 手机阅读 囧

根据【跟我一起学ML20（二）：Regression】一文，我们知道无论我们设计的模型有多么的优秀，在testing data上使用我们的模型总是会有误差，但是你知道如何进一步减小这个误差吗的方法吗？



如果 f^* 是我们模型的预测值， f^{\wedge} 是真实值，那么 f^* 是对 f^{\wedge} 的一个估值。



Bias反映了 f^* 相对于 f^{\wedge} 的偏离程度， f^* 越接近于 f^{\wedge} ，则Bias越小；反之，Bias越大。

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

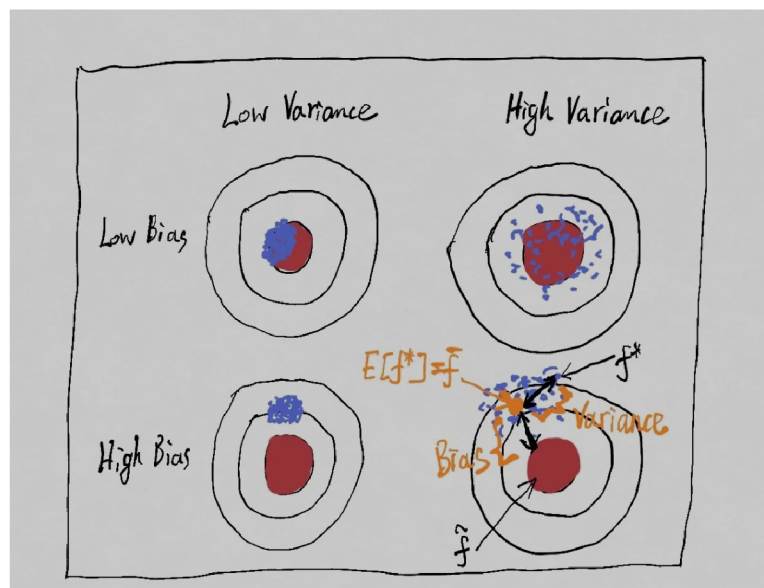
$$m = \frac{1}{N} \sum_n x^n \quad s^2 = \frac{1}{N} \sum_n (x^n - m)^2$$

Biased estimator

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

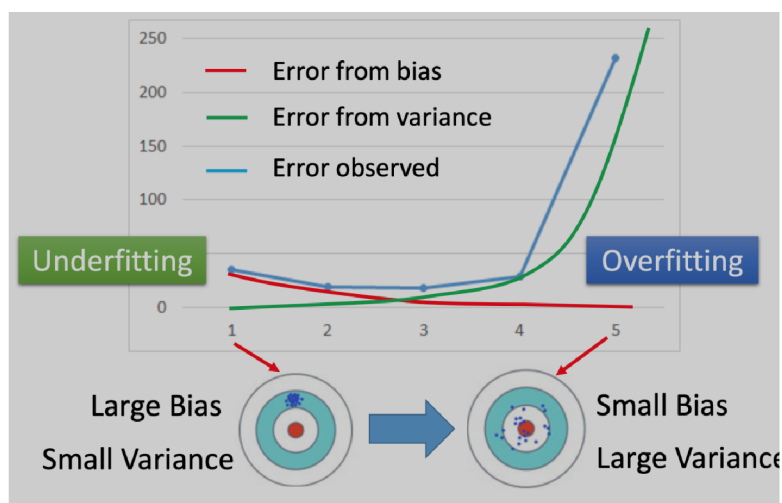
Increase N

Variance则反映了若干个 f^* 估值之间的聚合程度，如果若干 f^* 估值聚合度越高，则Variance越小；反之，Variance越大。



综上，Bias与Variance的关系如上图所示，我们最喜欢左上角的情况，最不喜欢右下角的情况。

- 如果你的模型在training data上面拟合的不是很好，换句话说你的模型预测结果与正确结果偏差较大，那么会造成Large Bias & Small Variance的情况；
- 如果你的模型在training data上面拟合的很好，但是在testing data上面拟合的很差，那么就会造成Small Bias & Large Variance的情况。



你可能会问，知道这些对模型预测精度的提升会有什么帮助吗？其实，一旦你了解到你的模型的Bias和Variance表现，你就知道了你的模型设计的是否得当以及后续的优化方向。让我们反过来思考：

- 如果你的模型是Large Bias & Small Variance类型的，就说明你的模型不是很贴合这个要解决的问题，那就要考虑重新设计新的模型来更好的解决这个问题。
- 如果你的模型是Small Bias & Large Variance类型的，就要考虑你的训练数据是否太少。那可以适当的增加训练数据，这样会在不伤害Bias的基础上减小Variance；或者进行Regularization，不过这种方法会伤害Bias，但是也会减小Variance，这个时候就要看你怎么样做trade-off了。

简而言之，如果你知道了Bias和Variance的关系，你就知道了后续的优化方向，从而在testing data上带来更精确的预测结果。

怎么样，是不是很简单，赶快pick它去优化你的模型吧~

什么？你的数据不够，也没有其他更多数据可以获取了？如此聪明的你肯定不会忘了还有K折交叉验证这个方法呢~

注：以上部分图片来自李宏毅老师机器学习2020课程slides。



END

欢迎关注公众号 **多杰平措OPL**，为您带来更多精彩内容~

