

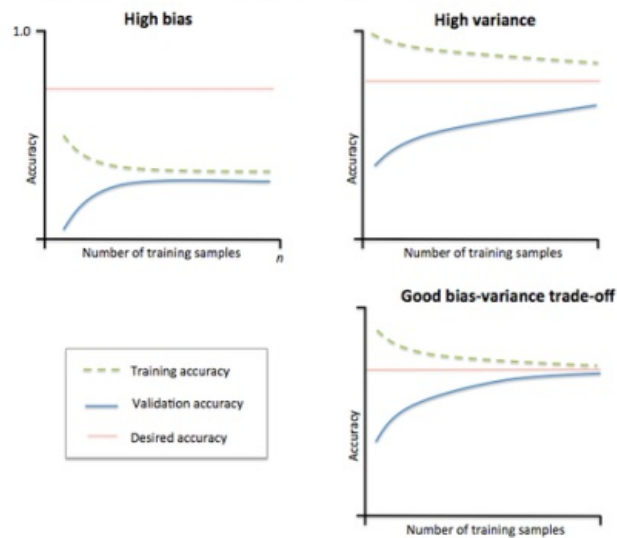
六、工作流程与模型调优

一、判断是否过拟合

欠拟合（高偏差）--->

过拟合（高方差）---->

□ 模型状态验证工具：学习曲线



学习曲线两条线分别是训练集上的准确率和测试集上的准确率，

过拟合在于两者差距太大，且比较大

欠拟合在于两者差距不大，都较小

正常的情况下学习曲线两者都比较大，且相差不大

二、如何解决过拟合和欠拟合

□

三、Bagging（可以简单的解决过拟合的方法）

□ Bagging

① 用一个算法

- 不用全部的数据集，每次取一个子集训练一个模型
- 分类：用这些模型的结果做vote
- 回归：对这些模型的结果取平均

② 用不同的算法

- 用这些模型的结果做vote 或 求平均

四、Boost（容易过拟合）

□ Adaboost

❶ 考得不好的原因是什么？

- 还不够努力，练习题要多次学习
 - 重复迭代和训练
- 时间分配要合理，要多练习之前做错的题
 - 每次分配给分错的样本更高的权重
- 我不聪明，但是脚踏实地，用最简单的知识不断积累，成为专家
 - 最简单的分类器的叠加