偏差与方差

偏差指的是算法的期望预测与真实值之间的偏差程度，反映了模型本身的拟合能力；

方差度量了同等大小的训练集的变动导致学习性能的变化，刻画了数据扰动所导致的影响。

Boosting

Boosting从优化角度来看，是用 forward-stagewise 这种贪心法去最小化 loss 函数,由于采取的是串行优化的策略，各子模型之间是强相关的，于是子模型之和并不能显著降低 variance，而每一个新的分类器都在前一个分类器的预测结果上改进，力求预测结果接近真实值，所以说 boosting 主要还是靠降低 bias 来提升预测精度。

Bagging

Bagging对样本重采样，对每一重采样得到的子样本集训练一个模型，最后取平均。由于子样本集的相似性以及使用的是同种模型，因此各模型有近似相等的bias和variance（事实上，各模型的分布也近似相同，但不独立）。

由于 $E[\frac{\sum X_i}{n}] = E[X_i]$ ，所以bagging后的bias和单个子模型的接近，一般来说不能显著降低bias。

另一方面，若各子模型独立，则有 $Var(\frac{\sum X_i}{n}) = \frac{Var(X_i)}{n}$ ，此时可以显著降低variance。若各子模型完全相同，则 $Var(\frac{\sum X_i}{n}) = Var(X_i)$ ，此时不会降低variance。bagging方法得到的各子模型是有一定相关性的，属于上面两个极端状况的中间态，因此可以一定程度降低variance。

为了进一步降低variance，Random forest 通过随机选取特征子集，进一步减少了模型之间的相关性，从而使得variance进一步降低。

参考资料

[为什么说bagging是减少variance，而boosting是减少bias?]