

什么是GBDT(梯度提升树)



虽然GBDT同样由许多决策树组成，但它与随机森林有许多不同。

其中之一是GBDT中的树都是回归树，树有分类有回归，区分它们的方法很简单。将苹果单纯分为好与坏的是分类树，如果能为苹果的好坏程度打个分，那它就是回归树。



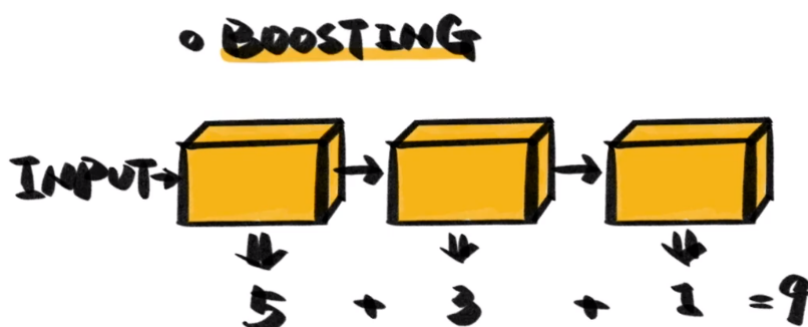
另一个不同的是GBDT中的每棵树都建立在前一棵树的基础上。

以苹果打分为例，我们会先训练一棵树大体预测一下苹果们的分数，再去训练一棵树去预测它们与真实分数间的差距，如果两者相加仍与真实分数存在差距，我们再训练第三棵树预测这部分差距，重复这个过程不断减少误差，将这些树的预测值加起来，就是苹果的分数。

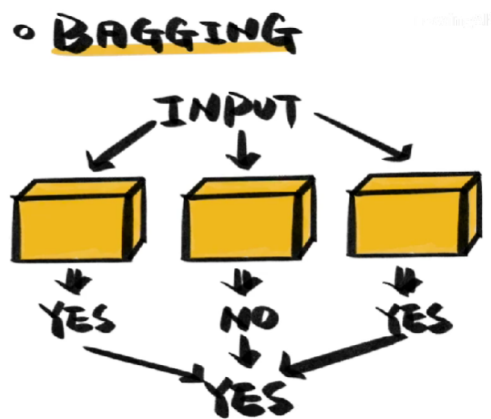
除了苹果，被评分的还可以是网页、电影、商品。通过预测关联程度、点击率或是用户的喜好程度来排序，GBDT在搜索、广告、推荐系统等领域有着广泛应用，能处理标签、数值等各类数据，解释性强，这些都是GBDT的优点。

不过由于树与树之间的相互依赖，需要较长的训练时间。运用多个模型共同解决问题，GBDT自然属于集成学习。

像这种一个模型依赖于上一个模型，共同逼近正确答案的方法被称为Boosting提升，也就是GBDT中的B。



与随机森林类似，模型间相互独立共同投票出结果的方法，则被称为Bagging(装袋)。



还有一种Stacking(堆叠)，是在一多个模型的基础上放置一个更高层的模型。将底层模型的输出作为它的输入，由它给出最终的预测结果。

