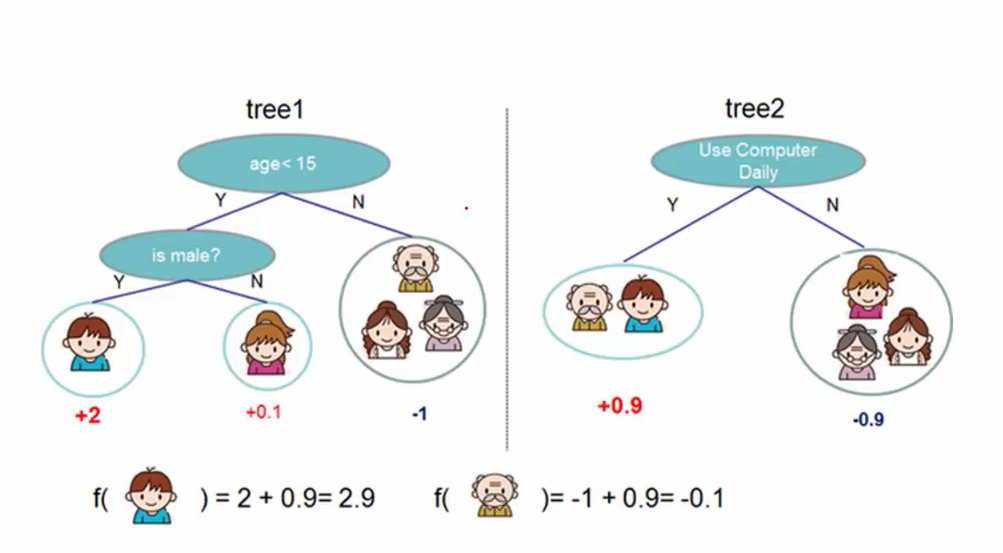
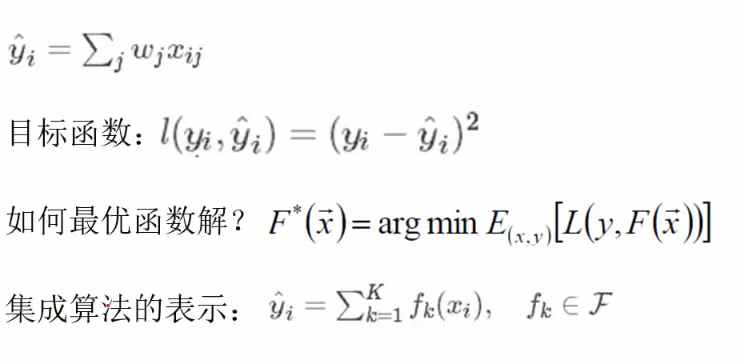
Xgb报告

1. Xgb的由多个弱分类器集成的一个分类算法，基础分类器就是决策树，我们看到一般单个决策树深度很小，不同的决策树可能用了不同的特征将同一个人分到了不同的叶子节点，然后每一个叶子节点都是一个权值的



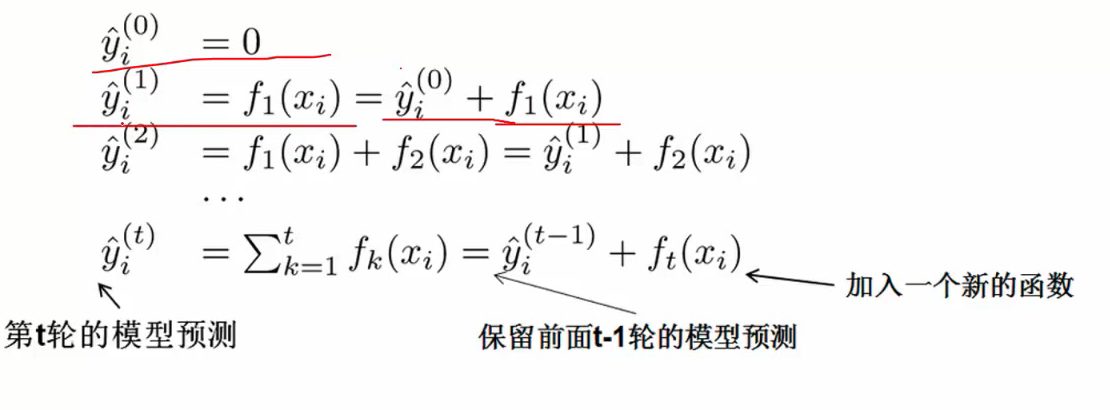
比如说图中这个分类任务是预测这个是不是喜欢玩游戏，我们有一系列特征，权值为正代表喜欢，比如第一个树这个小男孩得2分，第二个树得0.9分，综合得分2.9，

总结出来就是，一个人的得分就可以表示为不同树的对应节点的特征值和权值的线性表示：WX。现在我们可以得到一个目标函数，然后可以进行优化了

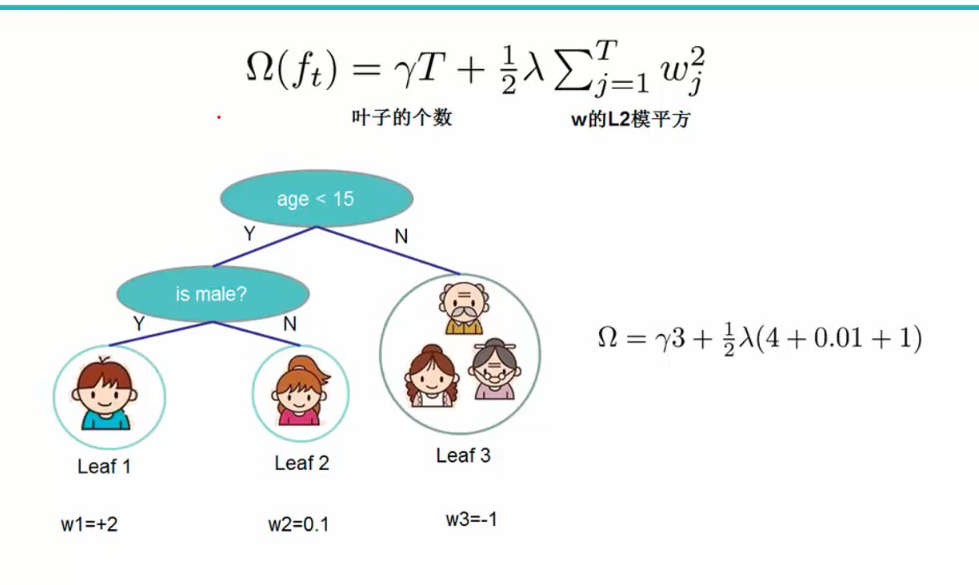


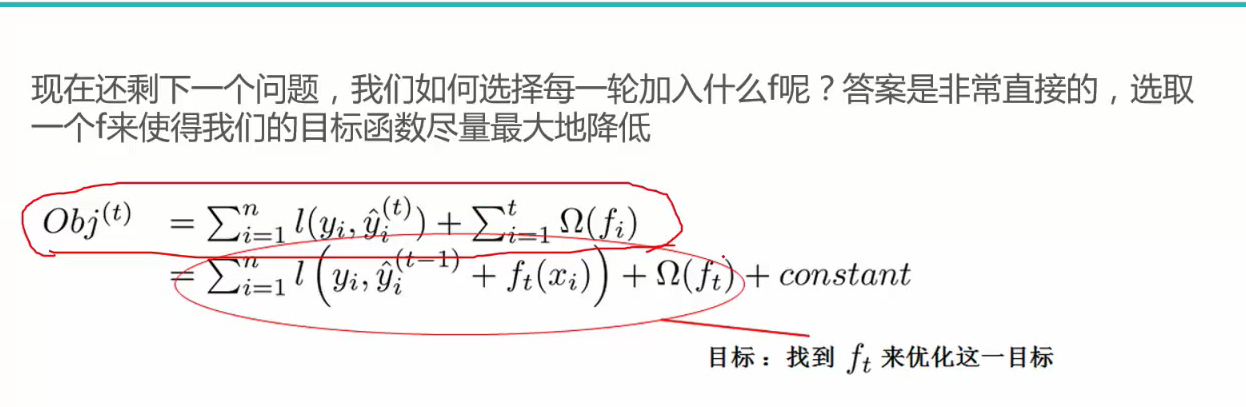
每一个样本都有一个损失值，最优解是所有样本的损失均值，相当于整体的期望求得一个最小解。

1. 既然是集成算法，它其实就是不断的添加新的树上去然后和之前的所有树集成到一起看效果，但是新树的增加是有条件的，它是期望每一次新加一个树之后呢对之前的效果要有所提升的，

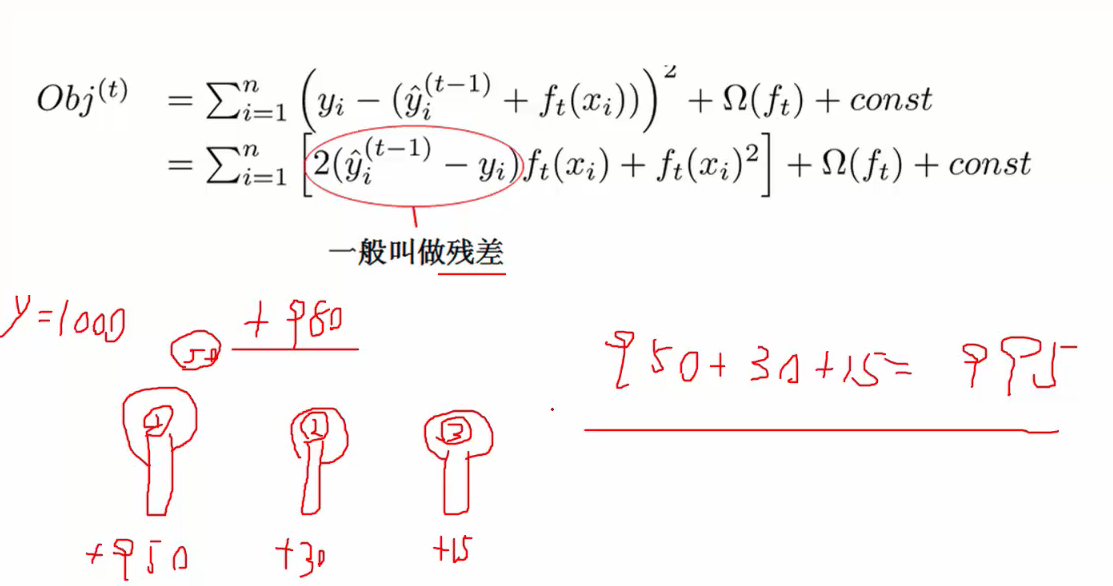


1. 由于决策树存在过拟合的风险，所有给最终的目标函数添加了一个惩罚项，这个惩罚项和叶子节点的个数有关。



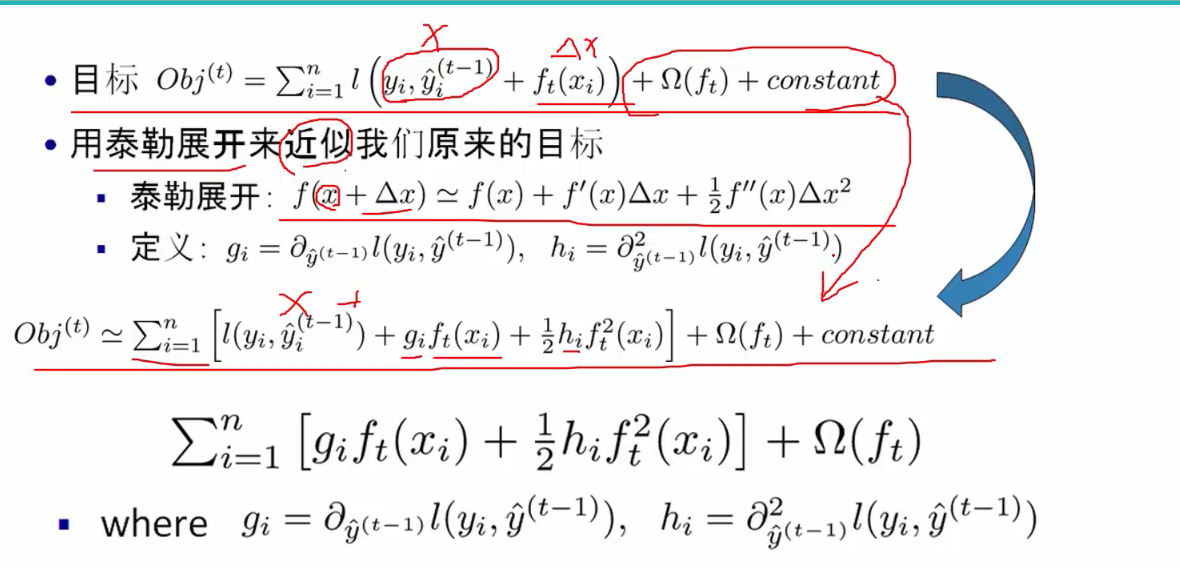


Xgb就是梯度提升的，



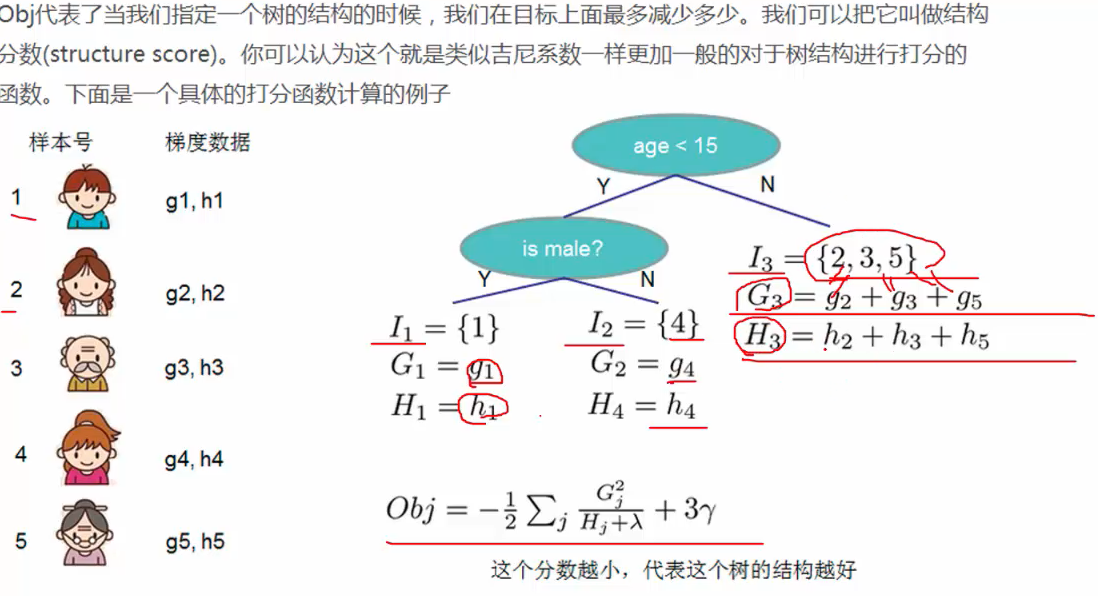
比如真实值是1000，t-1次预测结果是950，那这50便是预测残差，那新加一个t轮的树，需要更接近预测结果，它在提升30，这样以此类推，所以核心就在新增树很根据之前的残差来构建新的决策树

1. 数学原理

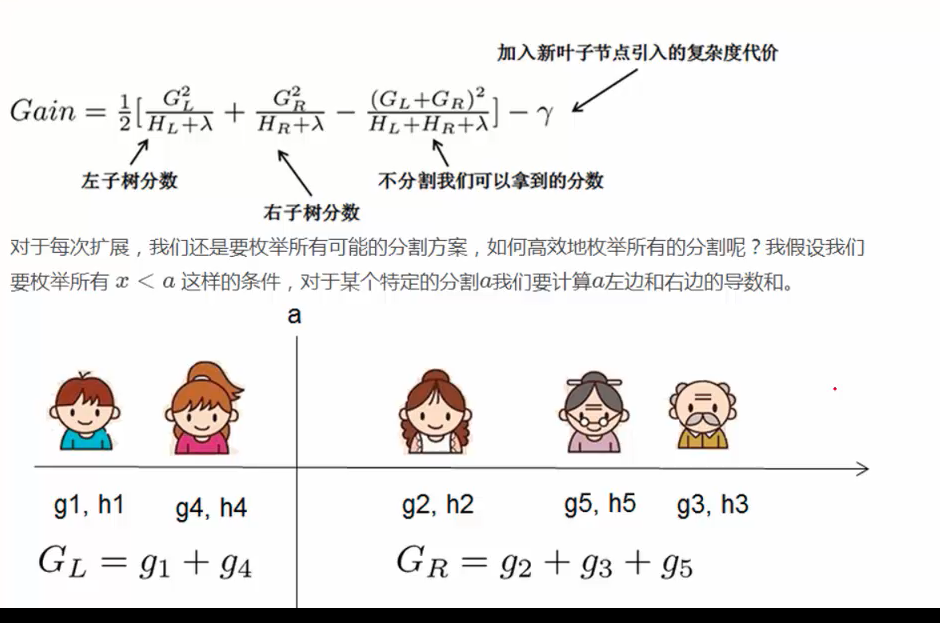


我们来看目标函数其实可以抽象为f（x+dx）微分的表现形式，又可以进行泰勒展开，对于dy=Adx我们近似的用函数的一阶导g和二阶导h的和来表示，

最终目标函数可以转化为这样的形式



在xgb中决策树的生成策略就不是信息增益或是基尼系数了，而是用这个优化函数因为最终的优化函数与叶子节点的个数有关



分割前后做一个差值，评估这一次分割的效果，只不过评价方法变了