Xgboost

xgboost思想：

1. xgboost的套路看陈天奇的论文和gbdt.pdf就好了，网络博客也很多

自己的思考：

1. xgboost怎样做回归？与分类有何区别？

xgboost做回归只需要更改损失函数即可，如交叉熵损失函数。只要损失函数可以二阶导

问题是：有些样本的预测值会在0-1之外，这个怎么处理？

1. 粗暴方法，小于0为0，大于1为1
2. 做归一化，（x-最小值）除以最大值减最小值
3. xgboost的每棵树的叶子节点的值怎么确定的？

叶子节点的值其实代表的是某一类样本的值，就是这一批样本在这个叶子结点上都等于这个值。然而这个值是在上一棵树训练好之后就确定下来的，因为gi,hi 都是上一棵树训练完成之后就知道了。问题是怎么得到树的结构了。树的结构就是本文下面总结的，用了一个特殊的信息增益公式。w值在树的结构确定了之后自然就得知了。

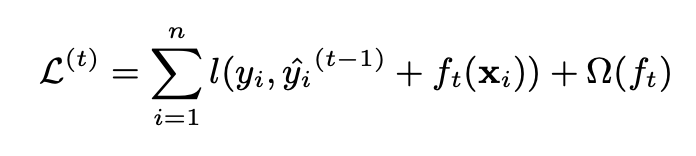
1. 第一棵树的叶子节点怎么确定？

因为上一棵树是没有的，所以可以计上一棵树的叶子节点全是0，跟计算第二棵树都是一样的套路。

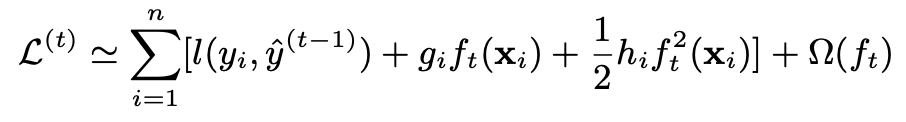
1. 加入了shrinkge的时候怎么确定叶子节点值？

shrinkge是学习率这个东西，因为每一棵树训练的目标是残差，而论文中或者资料中一般不会讲怎么加学习率。学习的目标残差等于y-学习率\*yi^即可。

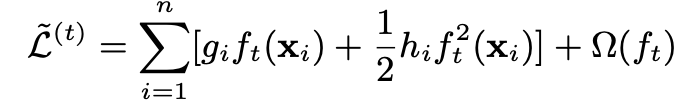
1. 首先是每一棵树都学习上一颗树的残差

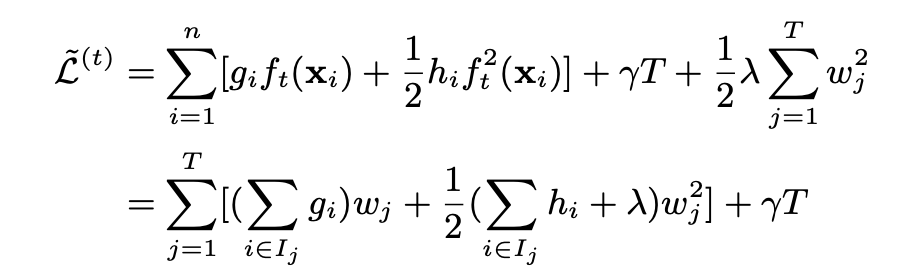


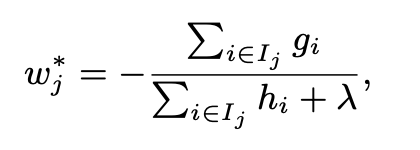
2.损失函数使用二阶泰勒展开式展开

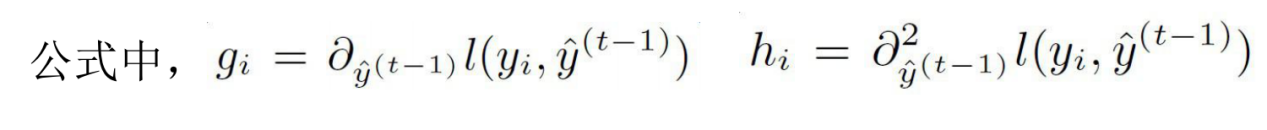


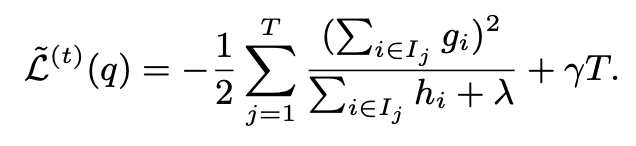
3. 因为每一颗树学习完成之后y（t-1）预测值是已知了。因此下一棵树的损失函数为如下所示

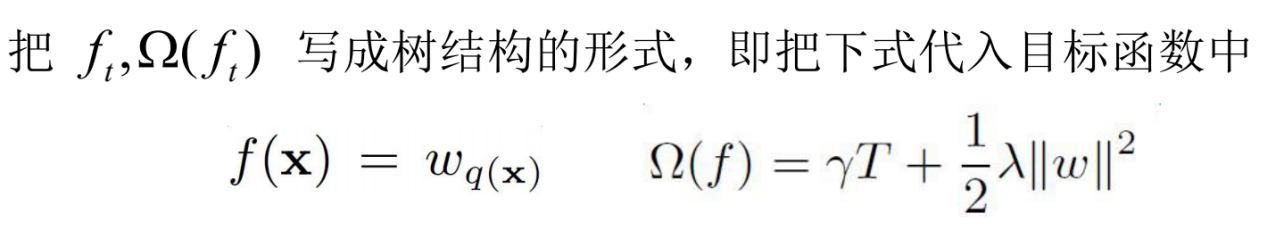
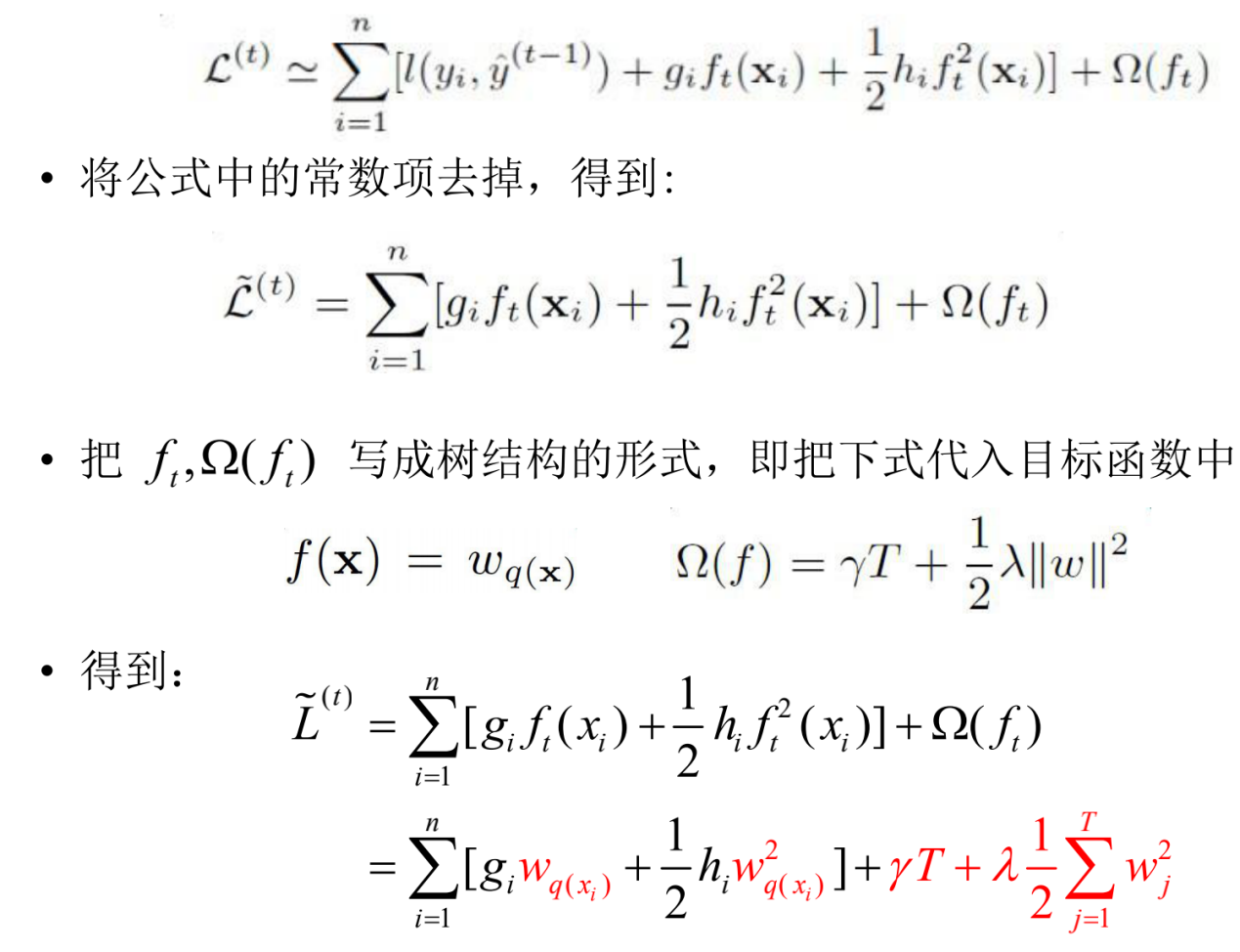


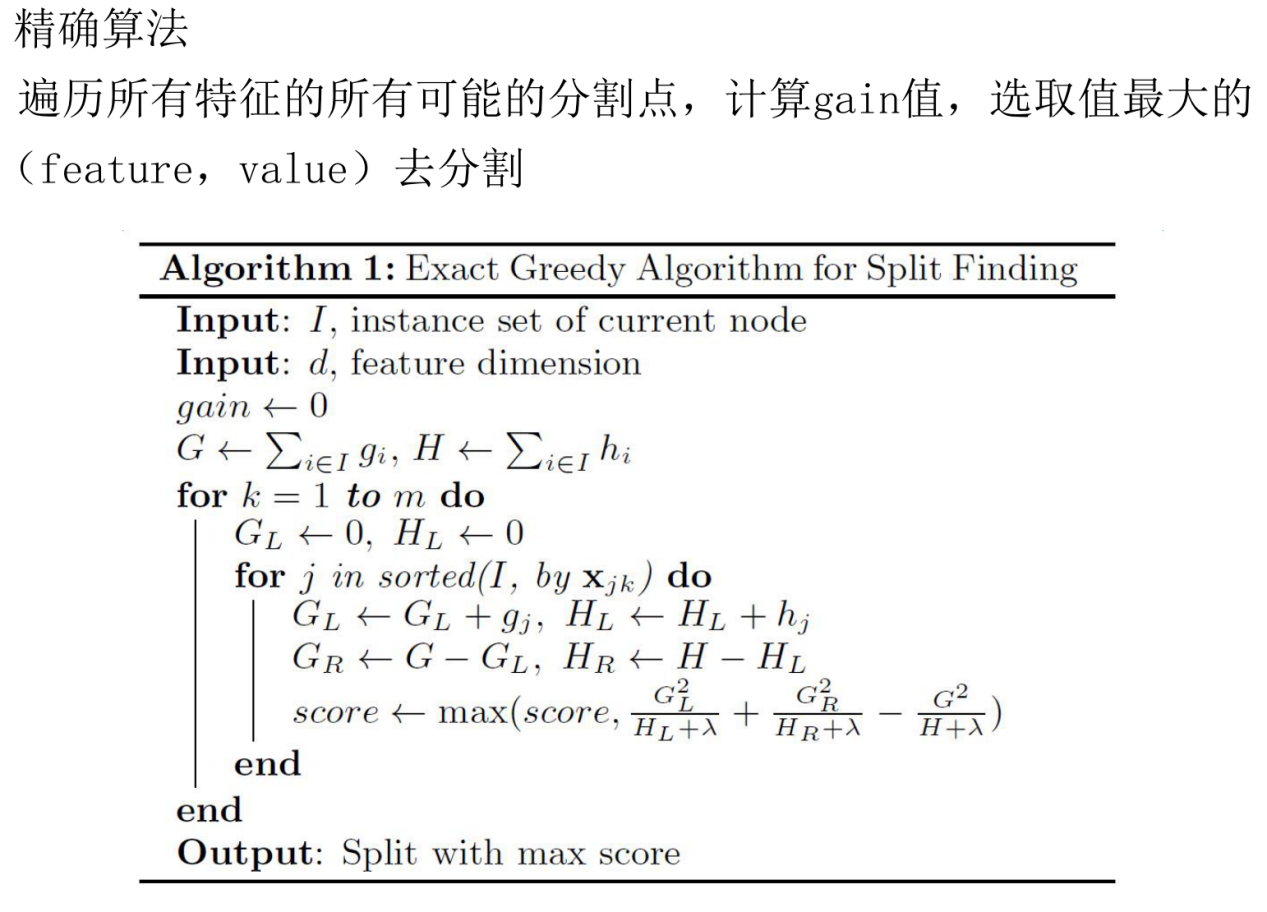
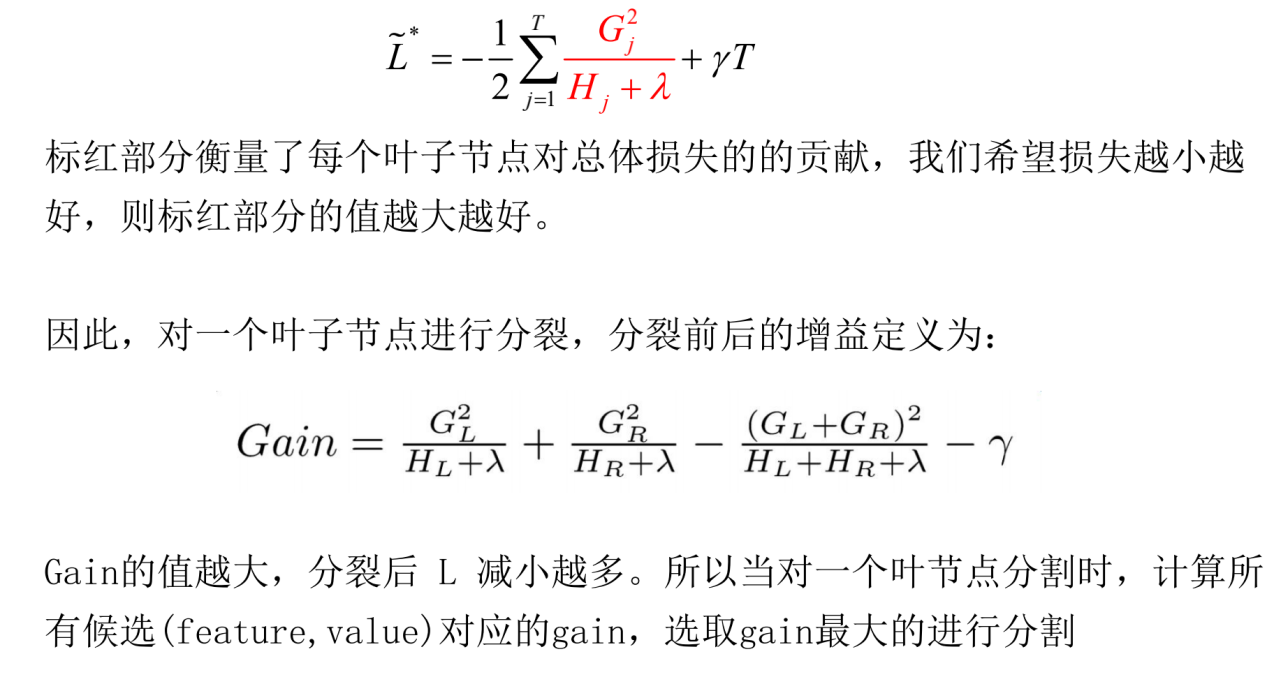
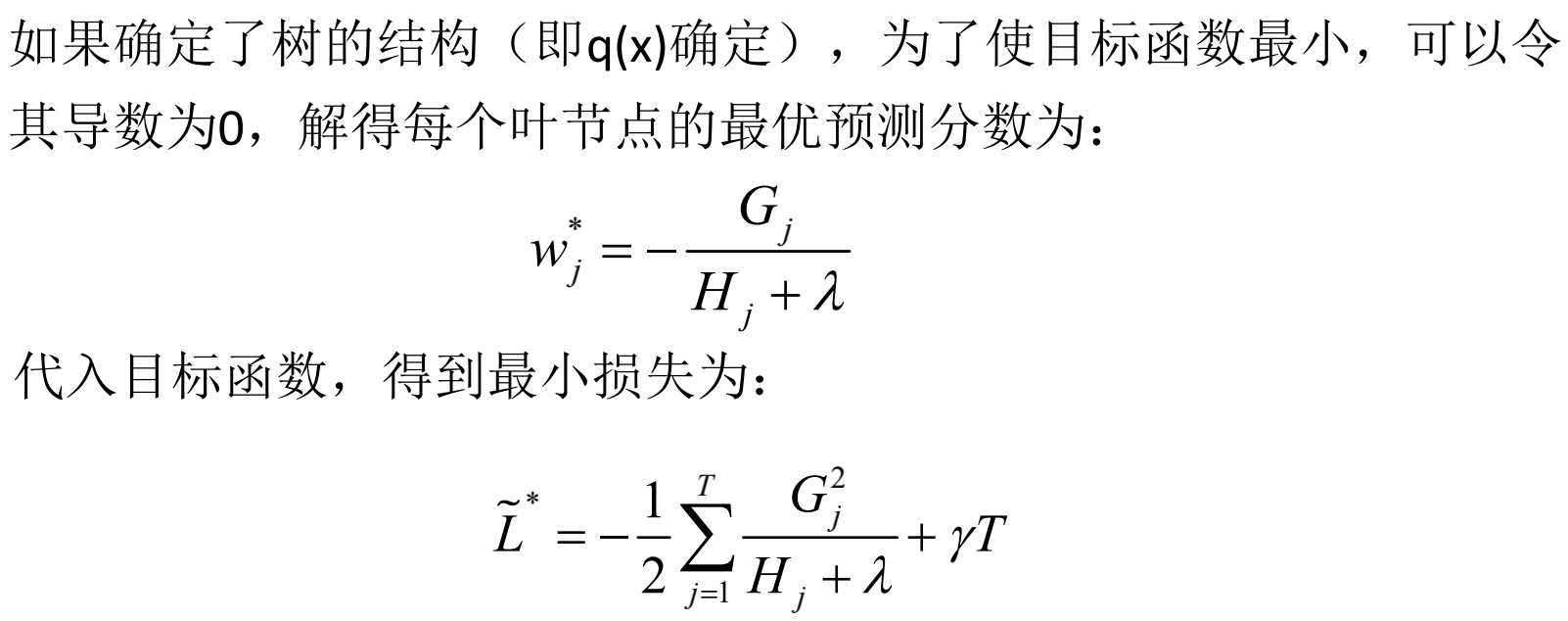
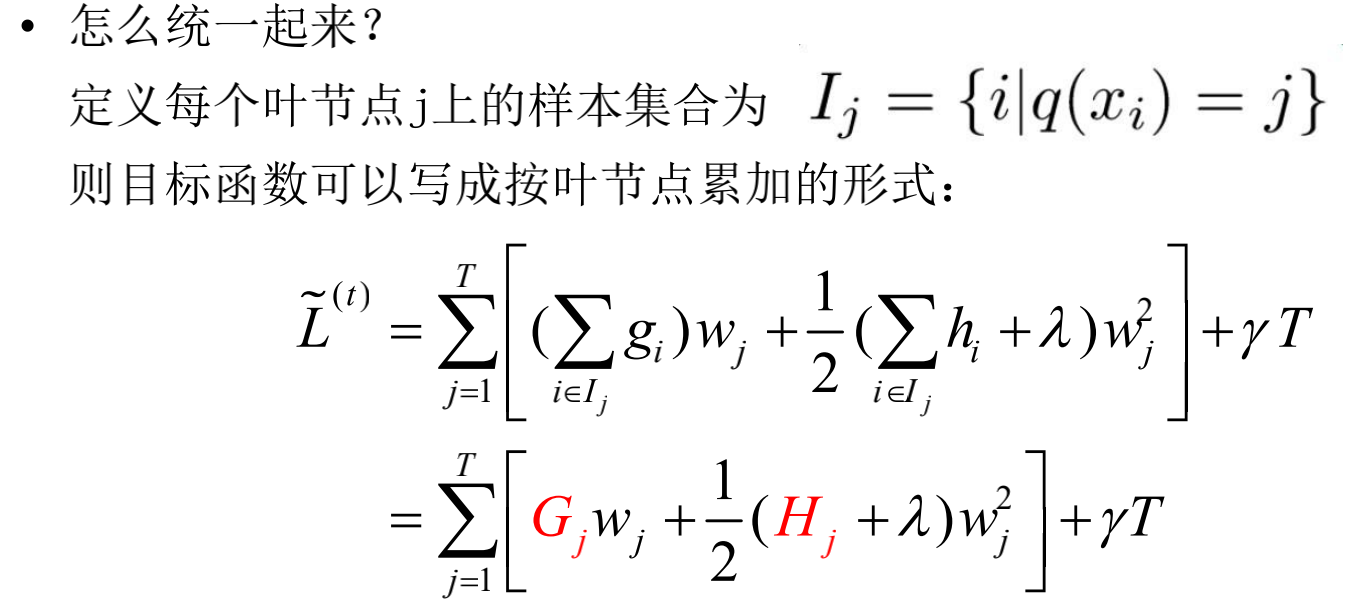
1. 因为要求的参数是叶子节点，做变换
2. 因为是二元一次方程，故可求最大值及其参数的值



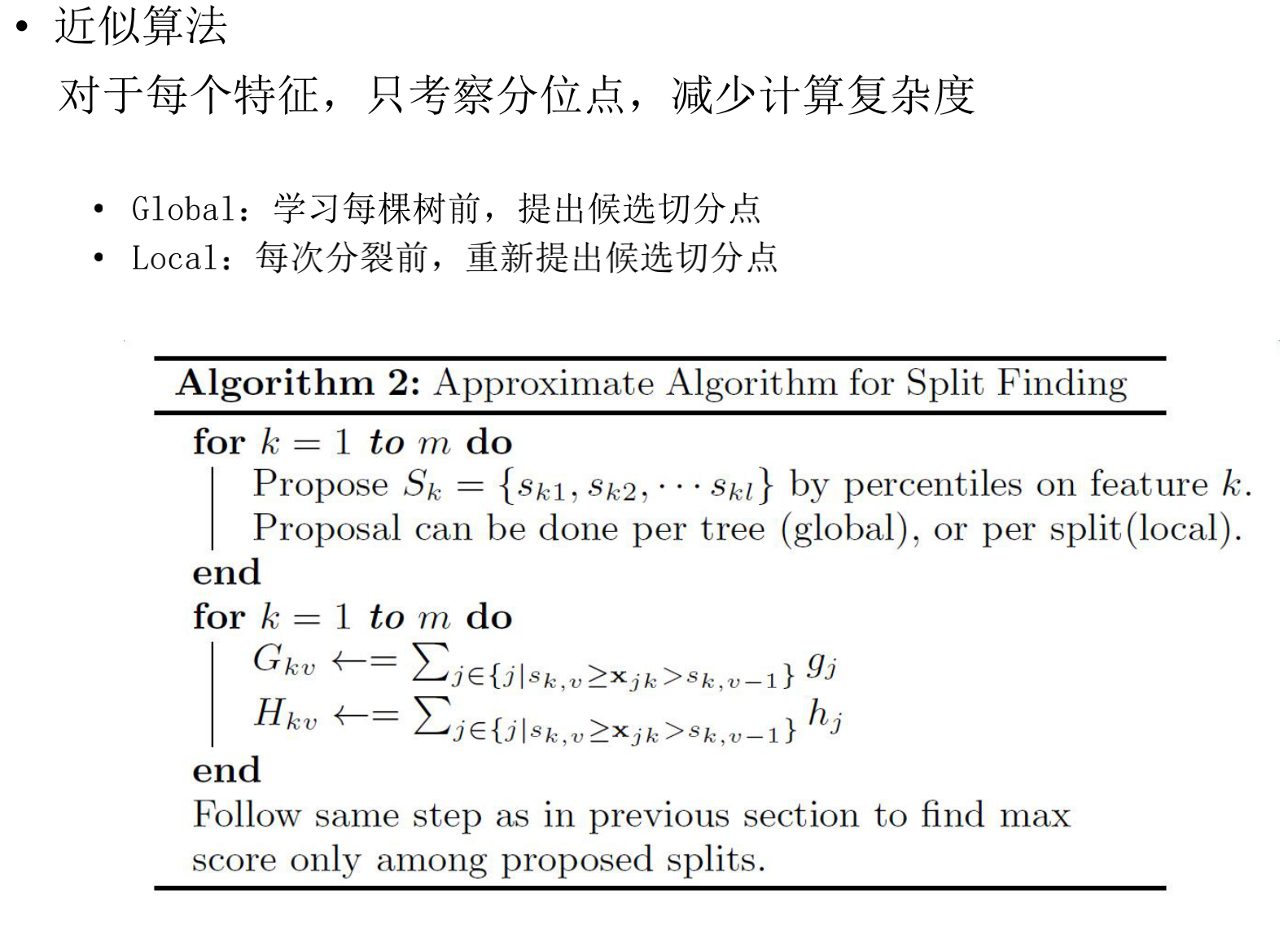
1. 将w代入损失函数即可得出。但是这是最优目标损失，但是问题是，做分类的时候基本无法做到这样，只能用这个函数来指导分裂，因为树分类是使用特征进行分裂。
2. 其中每次学习的分裂函数中包括上一颗树的叶子节点的值

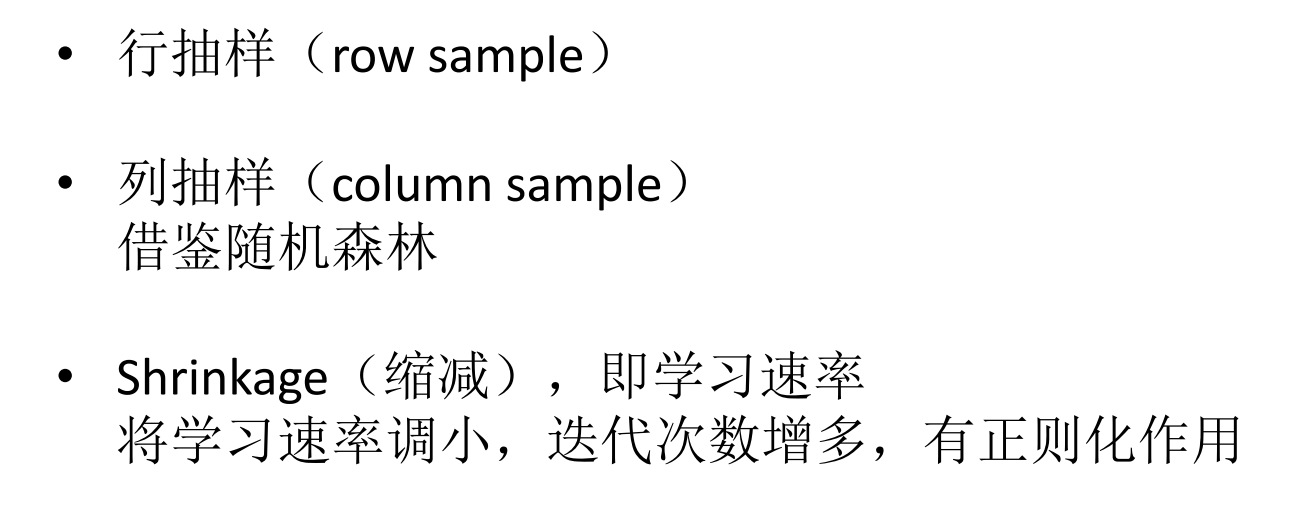


1. 
2. 



精确算法时间复杂度非常高。所以经常使用近似算法作为替代





缩减这个怎样用到xgboost中去？

拿回归树举例子，当第一棵树学习好的时候，往往会通过目标值与第一棵树叶子结点做差，

得到的残差拿来学习。然而第一棵树往往会学到的信息非常多，导致第二棵树开始的时候需要学习的信息已经非常少。这样就容易导致学习的不太好，往往会通过第一棵树的结果乘以一个缩减作为下一棵树学习的目标。

xgboost是回归树，如何应用到分类中去？