任务内容：

**GBDT算法梳理**

GBDT也是集成学习Boosting家族的成员，但是却和传统的Adaboost有很大的不同。回顾下Adaboost，我们是利用前一轮迭代弱学习器的误差率来更新训练集的权重，这样一轮轮的迭代下去。GBDT也是迭代，使用了前向分布算法，**但是弱学习器限定了只能使用CART回归树模型**，同时迭代思路和Adaboost也有所不同

1前向分布算法

GBDT中采用了前向分布算法，我们知道Adaboost是一系列弱学习器模型的线性组合，主要涉及到了基函数，基函数的参数，基函数的系数。

参考李航统计学习方法第二版 162 页。

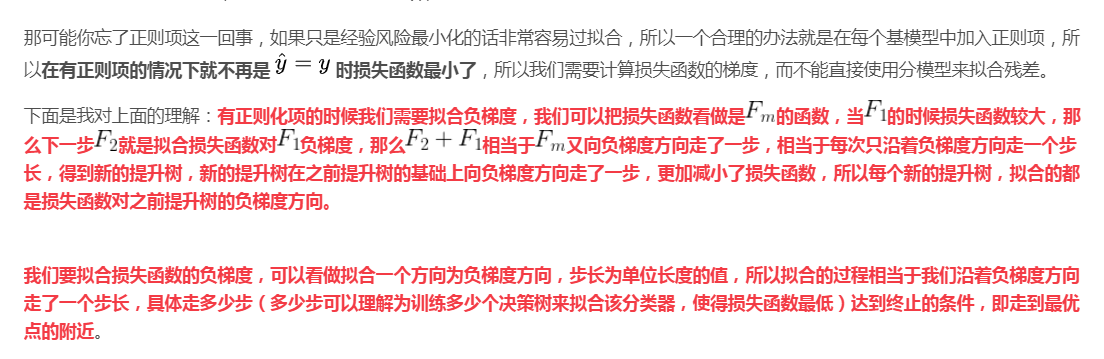
原来的函数最小化的损失函数是一个复杂的优化问题，其计算步骤复杂，时间长。

而前向分布算法在于优化这一问题的求解，原来的损失函数中的基函数（弱分类器）以及每个基函数的系数都是M个的，而最终又有N个基函数，因此原来的优化方法的计算复杂度是M\*N的。而通过前向分布算法，可以有效降低计算的复杂度，而达到逼近原来优化函数的特点。

具体的公式见 李航统计学习方法第二版 163页

2负梯度拟合

当损失函数为平方根或者是指数损失函数时，每一步的优化都很简单，但是对于一般的损失函数而言，每一步的优化不是那么简单，因此提出了负梯度优化的方法，其关键是利用损失函数的负梯度在当前模型的值作为回归问题提升树中的残差的近似值，从而拟合出一个回归树。



https://blog.csdn.net/youhuakongzhi/article/details/94488888

3损失函数

对于分类算法，其损失函数一般为对数损失函数和指数损失函数；

对于回归算法，常用的有四种，均方差、绝对损失、Huber损失、分位数损失

参见 https://www.cnblogs.com/pinard/p/6140514.html

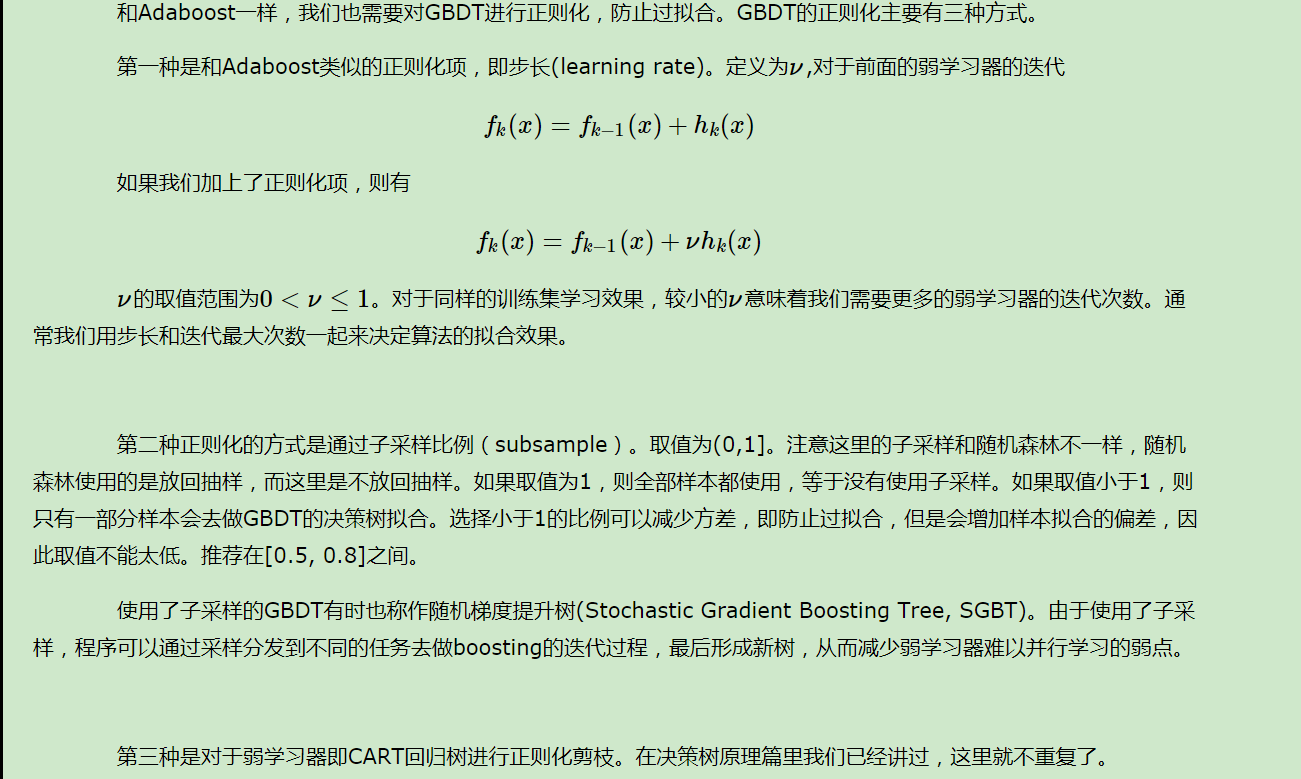
4回归

参见上面的网站

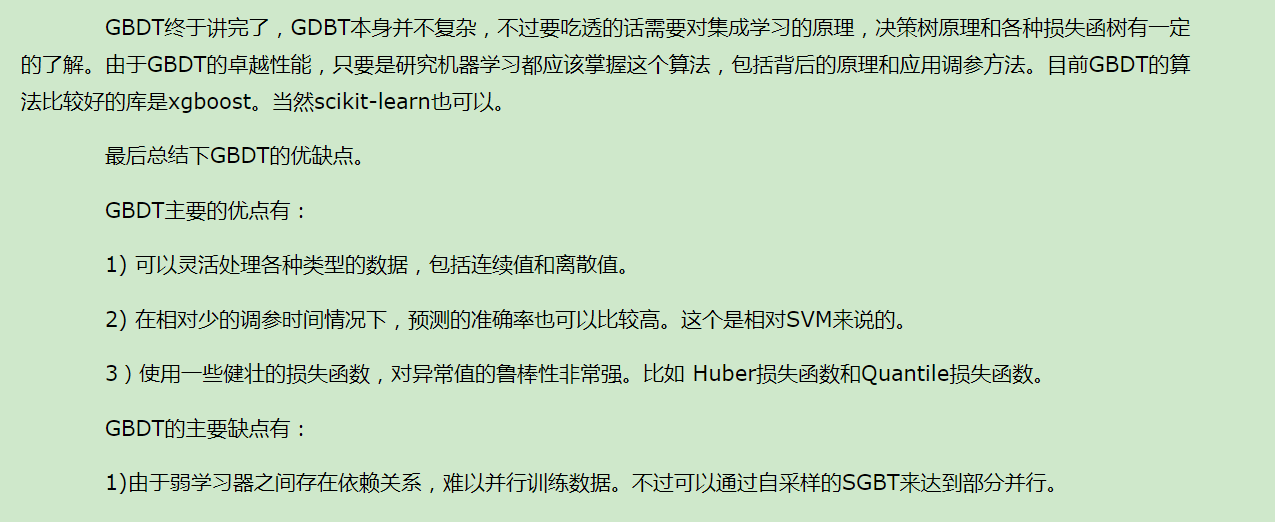
5二分类，多分类

参见上面的网站

6正则化



7优缺点



8 sklearn参数

http://blog.sina.com.cn/s/blog\_62970c250102xg5j.html

9应用场景

**搜索引擎排序应用 RankNet**