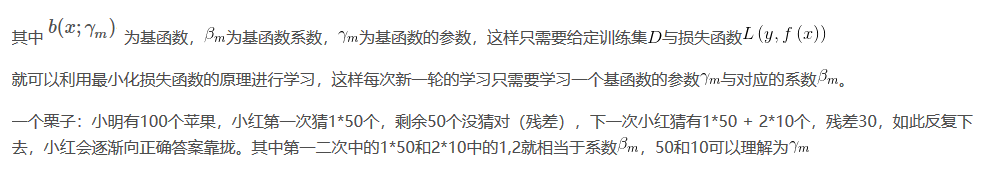
**GBDT算法梳理**

1. 前向分布算法
2. 负梯度拟合
3. 损失函数
4. 回归
5. 二分类，多分类
6. 正则化
7. 优缺点
8. sklearn参数
9. 应用场景

前向分布算法：

在这里插入图片描述



**负梯度拟合**

在GBDT的迭代中，假设我们前一轮迭代得到的强学习器是 *Ht*−1(*x*)Ht−1(x) H\_{t-1}(x) Ht−1​(x) , 损失函数是*L*(*f*(*x*),*Ht*−1(*x*))L(f(x),Ht−1(x)) L(f(x),H\_{t−1}(x)) L(f(x),Ht−1​(x)) , 我们本轮迭代的目标是找到一个CART回归树模型的弱学习器 *ht*(*x*)ht(x) h\_t(x) ht​(x) ，让本轮的损失 *L*(*f*(*x*),*ht*(*x*)+*Ht*−1(*x*))L(f(x),ht(x)+Ht−1(x)) L(f(x),h\_t(x)+H\_{t−1}(x)) L(f(x),ht​(x)+Ht−1​(x)) 最小，那么GBDT是如何实现让本轮的损失最小的呢？

针对这个问题，大牛Freidman提出了用损失函数的负梯度来拟合本轮损失*L*(*f*(*x*),*ht*(*x*)+*Ht*−1(*x*))L(f(x),ht(x)+Ht−1(x)) L(f(x),h\_t(x)+H\_{t−1}(x)) L(f(x),ht​(x)+Ht−1​(x))的近似值，进而拟合一个CART回归树。

损失函数

常用的GBDT损失函数：

1.对于分类算法，其损失函数一般有对数损失函数和指数损失函数两种:  
(1)如果是指数损失函数，则损失函数表达式为  
在这里插入图片描述  
其负梯度计算和叶子节点的最佳残差拟合参见Adaboost原理篇。  
(2)如果是对数损失函数，分为二元分类和多元分类两种，  
对于二元分类  
在这里插入图片描述  
2.对于回归算法，常用损失函数有如下4种:  
（1）均方差，这个是最常见的回归损失函数了  
在这里插入图片描述  
2）绝对损失，这个损失函数也很常见  
在这里插入图片描述  
对应负梯度误差为：  
在这里插入图片描述  
3）Huber损失，它是均方差和绝对损失的折衷产物，对于远离中心的异常点，采用绝对损失，而中心附近的点采用均方差。这个界限一般用分位数点度量。  
4）分位数损失。它对应的是分位数回归的损失函数。

GBDT的正则化主要有三种方式。  
（1）第一种是和Adaboost类似的正则化项，即步长(learning rate)。定义为ν,对于前面的弱学习器的迭代  
Ht(x)=Ht−1(x)+ht(x)  
如果我们加上了正则化项，则有  
Ht(x)=Ht−1(x)+αht(x),0<α≤1  
对于同样的训练集学习效果，较小的ν意味着我们需要更多的弱学习器的迭代次数。通常我们用步长和迭代最大次数一起来决定算法的拟合效果。

（2）第二种正则化的方式是通过子采样比例（subsample）。取值为(0,1]。注意这里的子采样和随机森林不一样，随机森林使用的是放回抽样，而这里是不放回抽样。如果取值为1，则全部样本都使用，等于没有使用子采样。如果取值小于1，则只有一部分样本会去做GBDT的决策树拟合。选择小于1的比例可以减少方差，即防止过拟合，但是会增加样本拟合的偏差，因此取值不能太低。推荐在[0.5, 0.8]之间。

使用了子采样的GBDT有时也称作随机梯度提升树(Stochastic Gradient Boosting Tree, SGBT)。由于使用了子采样，程序可以通过采样分发到不同的任务去做boosting的迭代过程，最后形成新树，从而减少弱学习器难以并行学习的弱点。  
　　  
（3）第三种是对于弱学习器即CART回归树进行正则化剪枝。在决策树原理篇里我们已经讲过，这里就不重复了。

优缺点

GBDT主要的优点有：  
（1）可以灵活处理各种类型的数据，包括连续值和离散值。  
（2）在相对少的调参时间情况下，预测的准备率也可以比较高。这个是相对SVM来说的。  
（3）使用一些健壮的损失函数，对异常值的鲁棒性非常强。比如 Huber损失函数和Quantile损失函数。

GBDT的主要缺点有：  
由于弱学习器之间存在依赖关系，难以并行训练数据。不过可以通过自采样的SGBT来达到部分并行。

sklearn参数

loss:损失函数度量，有对数似然损失deviance和指数损失函数exponential两种，默认是deviance，即对数似然损失，如果使用指数损失函数，则相当于Adaboost模型。

criterion: 样本集的切分策略，决策树中也有这个参数，但是两个参数值不一样，这里的参数值主要有friedman\_mse、mse和mae3个，分别对应friedman最小平方误差、最小平方误差和平均绝对值误差，friedman最小平方误差是最小平方误差的近似。

subsample:采样比例，这里的采样和bagging的采样不是一个概念，这里的采样是指选取多少比例的数据集利用决策树基模型去boosting，默认是1.0，即在全量数据集上利用决策树去boosting。

warm\_start:“暖启动”，默认值是False，即关闭状态，如果打开则表示，使用先前调试好的模型，在该模型的基础上继续boosting，如果关闭，则表示在样本集上从新训练一个新的基模型，且在该模型的基础上进行boosting。