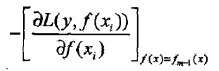
方骏-2020年4月25日-读书报告

1. 自己提出的问题的理解：
2. 提出的问题1：AdaBoost算法中的α是负值的话，那么对应的Gm不是也会很大的影响最后的f(x)吗？而且这样的话计算下一轮m+1的时候，误分类的点的权值小于正确分类的点的权值。

讨论后的理解：这里其实可以认为的是是一定大于0的，因为是有一定大于0，一个分类器的目标就是让尽量小，那么如果出现了一个是大于0.5的，让它对应的的输出结果全部取反，就可以得到一个正确率高的分类器了，所以分类器完全可以做到让结果小于等于0.5，这样的话就不会是负值了。

1. 提出的问题2： 最后的梯度提升算法是指什么？

讨论后的理解：如果损失函数是一般形式，不是这么直接的指数或者平方损失，那么优化可以利用最速下降法，即利用损失函数的负梯度在当前模型的值：



作为残差的拟合值，根据这个拟合一个回归树，然后通过拟合出的回归树拟合出最优的每一的每一个区域的c值，就可以得到这一次的基本模型了，最后相加就是回归树。

1. 别人提出的问题的理解：
2. 问题3：定理8.1如何说明AdaBoost算法可以在每一轮选取选取适当的Gm使得Zm最小，从而使训练误差下降最快？

自己的理解：这个定理说明了AdaBoost算法的误分类率上界是小于等于一个指数级的量的，那么在迭代计算的过程中，最小化这个就可以达到指数级下降的速率，这样的话训练误差的下降就非常快，可以到达尽快收敛的目标，这就是AdaBoost提升方法的优势。

1. 问题4： AdaBoost算法的迭代次数如何判断？

自己的理解：是通过误差精度要求来判断的，如果误差率小于给定的精度要求，那么算法就可以停止了，由于AdaBoost算法的误差率减小呈指数级速率下降，所以收敛速度非常快，因此不需要考虑精度太高导致收敛非常慢的情况。

1. 问题5：提升树与AdaBoost之间是什么关系？

自己的理解：很类似，都是属于提升方法，而且都是加法模型和前向分布算法，区别就是AdaBoost可以更广一些，因为基分类器可以选择很多，提升树的基分类器是决策树，而且如果损失函数不一样的，那么提升树就和AdaBoost的迭代方法是不同的，这个可以看后面的总结部分。总的来说，就是都属于提升方法，但是AdaBoost解决的范围更广一下。

1. 读书计划

1、本周完成的内容章节：《统计机器学习》第八章

2、下周计划：《统计机器学习》第九章

四、读书摘要及理解

1、提升方法是一种常用的统计学习方法，在分类问题中，通过改变训练样本的权重，学习多个分类器，并将这些分类器进行线性组合，提高分类的性能。提升方法的思路就类似于“三个臭皮匠顶个诸葛亮”，从弱学习算法出发，反复学习，得到一系列弱分类器，然后组合这些弱分类器，构成一个强分类器。这一系列弱分类器都是通过改变训练数据的权值分布来训练的。

2、AdaBoost算法是一种提升方法。首先初始化训练数据的权值分布：



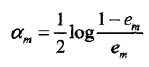
对于m=1,2,…,M，使用具有权值分布的训练数据集学习，得到基本分类器：



计算在训练数据集上的分类误差率：

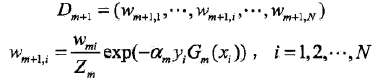


之后计算的系数：



这里的对数是自然对数。

然后更新训练数据集的权值分布：



这里，是规范化因子：

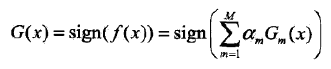


它使得真正是一个概率分布。

最后，构建基本分类器的线性组合：

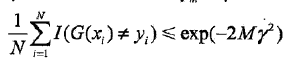


得到最终分类器：



这里值得说明的是总是大于等于0的，因为总是小于等于的，这里如果有分类器使得小于0.5的，那么完全可以让对应的的输出均取相反值，这样就可以得到高正确率的模型了，所以可以认为所有弱分类器的分类误差率都是小于等于0.5的。这样，得到的就是随着的减小而增大，所以分类准的权重也大。至于更新训练数据的权值，这里的下一轮权值由于是正值，所以对于误分类的点的计算公式中的指数都是正数，正确分类的都是负数，自然权重更大。

3、AdaBoost算法的训练误差是呈指数级下降的，所以AdaBoost往往结束很快，这也是它的优点。具体证明这里不再列出，但是结论就是：



可见误分类上界是指数级下降的，误差下降很快。

4、AdaBoost算法可以认为是模型为加法模型、损失函数为指数函数、学习算法为前向分布算法的二类分类学习方法。加法模型就是：



b是基函数，其他的均为参数和系数。

前向分布算法求解这一优化问题的想法就是：因为学习的是加法模型，所以想要从前向后每一步只学习一个基函数及其系数，逐步逼近优化目标函数就可以了，损失函数极小化是：



这样每一步只需要优化损失函数：



这样就是和AdaBoost类似的计算步骤了。

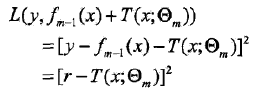
5、提升树模型就是利用了决策树为基函数的提升方法。提升树模型的加法模型为：



对于损失函数也是指数损失函数的二分类问题，只是在AdaBoost中将基函数改为决策树这么简单，所以这里考虑的是回归提升树，损失函数为平方误差。将输入空间划分为J个互不相交的区域，并且在每个区域上确定输出的常量c，那么树就可以表示为：

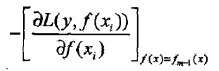


误差可以写做：



这样最小化误差就可以看作是拟合当前模型的残差，只要在这一次的树生成中，目标是可以拟合上一次模型的残差就可以了，即输入数据对应的输出是上一个模型的残差。这样就简单很多了。

如果损失函数是一般形式，不是这么直接的指数或者平方损失，那么优化可以利用最速下降法，即利用损失函数的负梯度在当前模型的值：



作为残差的拟合值，根据这个拟合一个回归树，然后通过拟合出的回归树拟合出最优的每一的每一个区域的c值，就可以得到这一次的基本模型了，最后相加就是回归树。