**寒假第三周读书报告**09118223 吴亦珂

读书进度：第十章基本结束

1. **问题与解答**

1.（别人提出）为什么提升树模型中没有权值？

讨论结果：回归提升树是一个拟合残差的过程。与AdaBoost不同的是：AdaBoost使用多个基学习器来拟合训练集数据（可以看成都具有相同的目标函数），而基学习器训练的结果有好有坏。基于训练结果好的给一个较大比重的设定，来分析各个基学习器的比重。而回归提升树是一个不断减少残差的过程，这是与AdaBoost最大的不同。因此其不需要给每一个树设定一个权值。

2.（别人提出）在梯度提升构建回归树的算法中，(2)b步骤不是已经获得了吗？为什么不采用这个而是保留了叶结点区域但重新用另一种方式获得？

讨论结果：在（2）b步骤当中，仅仅是获得了第m棵树的叶结点区域，但没有获得。是在（2）c步骤中获得的。

3.（别人提出）AdaBoost如何拓展至多分类问题？

讨论结果：书本上虽然只介绍了AdaBoost用于二分类，但其也可以用于多分类问题，通常可以采用多次使用二分类的方法来达成这个目的，如采用“one versus all”或者“one versus rest”的方法来达成。

1. **下周计划安排**

下周讨论本书第九章，EM算法是机器学习中很重要的一种算法，因此应该好好研读。

1. **读书收获**

第八章主要介绍了AdaBoost算法。

1. 基本思想：提高被前一轮弱分类器错误分类样本的权值，降低被正确分类样本的权值，使得那些没有被正确分类的数据受到更大关注，同时加大分类误差率小的弱分类器的权值，使其在表决中起较大的作用；减小分类误差率大的弱分类器的权值，使其在表决中起较小作用。
2. 具体算法：
3. 初始化训练数据的权值分布

IMG_256

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

IMG_256

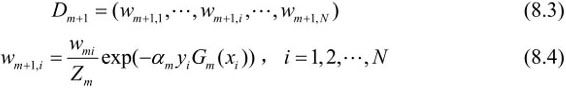
计算Gm(x)在训练数据集上的分类误差率：

IMG_256

计算Gm(x)的系数

IMG_256

更新训练数据集的权值分布

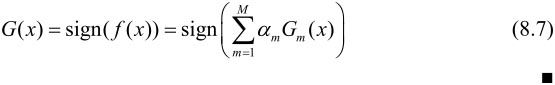


Zm是规范化因子：

IMG_256

3.构建基本分类器的线性组合

IMG_256

4. 对AdaBoost算法的解释说明AdaBoost算法是前向分步加法算法的特例，当模型是由基本分类器组成的加法模型，损失函数是指数函数时，二者等价。

1. 提升树算法

采用分步前向算法，并且每一步拟合当前模型的残差。

1. 梯度提升

利用损失函数的负梯度在当前模型的值作为提升树算法中的残差的近似值，其余基本一致。