##### QG工作室数据挖掘小组实验报告

|  |  |  |
| --- | --- | --- |
| 实习生：李炜乐 | 导师：张润鹏 | 日期：2019 年 7 月23日 |

|  |
| --- |
| **实验名称：Adaboost算法** |
| **已完成内容：**   1. **实现了Adaboost分类集成算法** 2. **实现了Adaboost回归集成算法** |
| **未完成内容：暂无** |
| **未完成原因：暂无** |
| **需要帮助：暂无** |

|  |  |
| --- | --- |
| **实验总结** | |
| **知识点总结：**   1. **Adaboost算法是其中算法实现最难的一个，该算法的思想是进行学习。Adaboost先对数据集进行权重的何划分，刚开始的权重划分都一样，然后依靠决策树进行建树预算，对其错误的数据集增加更大的而权重，对其正确的数据集降低权重，然后再建树的过程中对权重更大的数据进行分类。以至于效果最优。** 2. **Adaboost分类算法建树的时候需要更改较多代码，其中需要将权重D加入建树的过程，我采用将基尼指数进行权重代替以及样本的权值也进行权重代替。** 3. **Adaboost回归算法也同样需要不少修改不少代码，在这里我将权重加入到方差计算的式子中去。** 4. **因为时间原因Adaboost分类回归算法都完成的比较仓促，对于两个算法，都有一个致命的问题，它的权重会有一定的更改，但是错误率并没有改变。对于回归算法我预测是因为数据集的不稳定，导致预测的数据都差不多为一个数据，对于其权重变化并不大，不能导致目标函数的减少，对于分类问题，我认为是因为再利用权重建树的时候不能很好的体现出权重的作用。目前还在解决问题中** 5. **Adaboost分类算法实验adult数据集，第一次错误率为4%左右，而后没有太大的改变。回归算法使用森林火灾数据集，误差在10000左右很大，原因cart树不能进行很好的预测** | |
| **遇到问题**：对于决策树进行权重的带入的理解还不够深刻，可能因为没有很好的利用到权重导致不能进行很好学习 | **解决过程：目前进行了多种方法进行更改，但是效果都不太明显还在进一步更改。** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **导师评价** | | | | |
| **实验分数** | **知识掌握情况** | **代码编写能力** | **建议** | **评价日期** |
|  |  |  |  |  |