##### 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：Extratree算法** |
| **已完成内容：**   1. **实现了extratree分类算法** 2. **实现了extratree回归算法** |
| **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. **Extratree算法属于集成算法中的一类，跟随机森林比较相似。极度随机树讲究一个极度随机。与随机森林相比，extratree也有着列取样，即数据集属性的随机选取。都属于构建多颗决策树然后进行预测。** 2. **ExtraTree分类树算法在建树过程，选其中的一个特征的随机一个属性进行结点的分支，对于每一个结点都讲究一个随机性。结果取其森林预测的数值中出现次数最多的值** 3. **ExtraTree回归树算法在建树的过程，随机选其其中一个属性，并且随机选取该属性中的最大值和最小值中的某个值。结果取其多颗决策树预测后的均值。** 4. **ExtraTree分类树的数据集为adult数据集，运行的准确率83.5%左右，运行时间为4.25s左右** 5. **ExtraTree回归森林的数据集森林火灾数据集，运行的误差比为-30左右，运行时间为83s** | |
| **遇到问题**：对于ExtraTree回归树并不能很好的预测出很好的预测值，预测值于实际值相差较大 | **解决过程：对于其他数据集拟合性能挺好的，这个数据集的结果较多为0值，预测可能一定程度会影响了数据的预测** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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** |
|  |  |  |  |  |