##### 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 21 日 |

|  |
| --- |
| **实验名称：Logistic算法** |
| **已完成内容：**   1. **完成了Logistic算法代码的编写** 2. **将Logistic算法使用到adult数据集中** |
| **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. **Logistic回归算法总共包括三个方面：使用随机梯度上升的方法计算局部最优的参数，然后将其带入logistic公式进行计算法，根据取值划分数据集的类别** 2. **Logistic的原理是根据Logistic进行划分数据集，利用随机梯度上升的方法进行得到较优的参数。** 3. **经过多次验证Logistic的错误率在20%到25%之间，运行为8.23s时间** | |
| **遇到问题**：在adult数据中存在某一些数据集的值过大，可能会很大程度上影响了预测，导致错误率极大增加 | **解决过程：采取了数据归一化的处理，在一定程度上明显的减少了错误率** |

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