**人工智能上机实验报告**

|  |  |  |
| --- | --- | --- |
| 课程名称：人工智能 | 班级：计科1902 | 实验日期：2021.11.25 |
| 姓名： 徐超信 | 学号：1935010102 | 指导教师： 韩焘 |
| 实验序号：一 | | 实验成绩： |
| 一、实验名称  **机器学习范式实验-以逻辑回归为例** | | |
| 二、实验目的及要求  1、熟悉机器学习的全过程；  2、明确调参过程及其意义；  3、熟练使用逻辑回归算法解决具体问题。 | | |
| 三、实验环境  WIN10+anaconda的虚拟python环境 | | |
| 四、实验内容  工具包的参数意义参照课件Chp3\_LogisticRegression.pptx中80-89页  根据以下步骤完成实验过程。   1. 配置好anaconda的虚拟python环境，安装csv，numpy，sklearn等安装包。 2. 看懂模板代码，问题：初始设定的数据集中，样本的各个属性是什么？样本的预测目标是什么？测试集变量是什么，一共多少个条目？训练集变量是什么，一共多少个条目？ 3. 执行跑通模板代码（忽略红字warning），并尽可能指出输出的报告中各个指标所代表的意义。（注：数据中男生性别为1，女生性别为0，）为体现调参结果的优劣，以下实验均只比较和记录macro-average f1-score作为评价标准 4. 模板代码有红字提示，解释其意义，并修复该warning。（后续沿用修复后的结果） 5. 依据模板样式修改代码，增加算法在训练集上的预测结果，并print对应的报告结果。 6. 参数中C是什么意义，设置C=1e-2,C=1和C=100在训练集和测试测试集上有什么结果，最优结果选择什么？选择三者中最优的参数进行后续实验。 7. 根据模板代码的样式，将数据集的属性改为{height，weight}，其训练和测试结果如何？ 8. 正则项采用L1范数，其训练和测试结果如何？ 9. 将数据集的属性改为全部可以用属性，其训练和测试结果如何？ | | |
| 五、实验步骤，调试过程，实验结果等  详细记录实现过程中，程序修改过程，及在调试过程中出现的问题及解决方法，  记录程序执行的结果。  记录共有23283条  参数中C是什么意义: 交叉熵损失函数系数  Warning:迭代计算的次数过小,导致来不及收敛,可以通过将max\_iter调大,使得其能够收敛成功.  设置C=1e-2,C=1和C=100在训练集和测试测试集上有什么结果，最优结果选择什么？选择三者中最优的参数进行后续实验。  C=1e-2:    C=1    C=100    将数据集的属性改为{height，weight}，其训练和测试结果如何:   正则项采用L1范数，其训练和测试结果如何？ 需要迭代更多次才收敛     将数据集的属性改为全部可以用属性，其训练和测试结果如何？ | | |
| 七、总结  对上机实践结果进行分析，给出你的上机的心得体会及改进意见。 | | |
| 八、附录  附上在测试集上你认为的表现最优逻辑回归算法代码。 | | |