|  |
| --- |
| 西南民族大学学生实验报告  教学单位：计科学院 实验室名称：BS226 实验时间：2019年10月16日 |
| 实验项目名称：实验3—Logistic回归 实验成绩： 教师签名： |
| 实验项目报告内容（1、实验目的；2、实验主要过程与结果；3、分析讨论）。  **一、实验目的**  学会使用Logistic回归进行简单的分类。  **二、实验主要过程与结果**   1. 基于最优化方法的最佳回归系数确定    1. 训练算法：使用梯度上升找到最佳参数   定义loadDataSet()函数，打开文本文件textSet.txt并进行逐行读取，函数如图1所示；定义sigmoid()函数，进行归一化处理，函数如图2所示；定义gradAscent()函数，通过梯度上升算法找到最佳回归系数，拟合出Logistic回归模型的最佳参数，如图3所示；  在主函数中加入如下语句，进行测试：  dataArr, labelMat = loadDataSet()  print(gradAscent(dataArr, labelMat))  运行结果如图4所示。    图 1    图 2    图 3    图 4   * 1. 分析数据：画出决策边界   定义plotBestFit()函数，画出数据集和Logistic回归最佳拟合直线，函数如图5所示；  在主函数中加入如下语句进行测试：  dataArr, labelMat = loadDataSet()  weights = gradAscent(dataArr, labelMat)  plotBestFit(weights.getA()) # 将weight矩阵转换成数组  运行结果如图6所示。    图 5    图 6   * 1. 训练算法：随机梯度上升   定义stocGradAscent0()函数，如图7所示，在主函数中加入如下语句进行测试：  dataArr, labelMat = loadDataSet()  weights = stocGradAscent0(np.array(dataArr), labelMat)  plotBestFit(weights)  运行结果如图8所示。    图 7    图 8  使用该算法时，回归系数经过大量迭代才能达到稳定值，并且仍然有局部的波动现象，因此，需要对算法进行改进，改进后的stocGradAscent1()函数，如图9所示，运行结果如图10所示。    图 9    图 10   1. 示例：从疝气病症预测病马的死亡率    1. 测试算法：用Logistic回归进行分类   定义函数classifyVector()，以回归系数和特征向量作为输入计算对应的sigmoid值，如图11所示；定义函数colicTest()打开测试机和训练集，对数据进行格式化处理并训练，如图12所示；定义函数multiTest()来调用colicTest()函数10词，求结果的平均值，如图13所示；在主函数中调用函数multiTest()，运行结果如图14所示。    图 11    图 12    图 13    图 14  **三、分析讨论**  1. 在1.2画出决策边界时，运行时需要将weights矩阵转换成数组，若不转换成数组，则会报错。  ValueError: x and y must have same first dimension, but have shapes (60,) and (1, 60)  这是由于当weights为矩阵时x的shape为60， y的shape为1， 60；  而weights是数组时，x的shape为60， y的shape也为60，维度相同的数据才能作图。 |

注：实验报告的内容及格式可由学院根据学科专业特点确定；全校各专业必须使用学校统一封面。