# 《数据挖掘与机器学习》

|  |
| --- |
| 【实验名称】 实验：回归算法 |
| 【实验目的】 |
| 1.了解回归算法理论基础  2.平台实现算法  3. 编程实现分类算法 |
| 【实验原理】 |
| 逻辑回归（Logistics regression）是根据现有数据对分类边界线建立回归公式，以此进行分类。逻辑回归在线性回归模型的基础上，通过引入Sigmoid函数，将线性回归的输出值映射到（0，1）范围。使用阈值将结果转换成0或1，能够完成两类问题的预测。 |
| 【实验描述】 |
| 前进银行搜集了贷款用户的数据，包括用户年龄、欠款额(百元)、月收入(元)和是否逾期的信息，数据在下表中。使用五个用户的数据建立一个逻辑回归模型。通过6号用户的个人信息，判断用户贷款20万元后是否会逾期。  屏幕剪辑 |
| 【实验分析】 |
| 与线性回归类似，需要拟合出w1、w2、w3、b这几个系数，其中w1是年龄的系数，w2是欠款额的系数，w3是收入的系数，b是偏移向量，“逾期否”特征作为分类使用的标签。当模型根据前五位用户的数据构造出模型之后，可以判断出编号6的用户的逾期情况。 |
| 【实验步骤】 |
| 阶段一：建立逻辑回归模型STEP1：加载数据集。STEP2：绘制数据的散点图，查看数据分布情况。STEP3：定义Sigmoid函数、损失函数，使用梯度下降确定模型参数。STEP4：初始化模型，并对模型进行训练。STEP5：根据得到的参数，绘制模型分类线。STEP6：绘制损失函数变化曲线。阶段二：对信用卡逾期进行预判STEP1：读取数据。STEP2：绘图显示数据。STEP3：建立逻辑回归模型。STEP4：返回预测结果。STEP5：使用参数绘制分类结果图，判断出编号6的用户的逾期情况。 |