

人工智能实验报告

|  |  |  |  |
| --- | --- | --- | --- |
| 题目 | 实验五：机器学习算法逻辑回归实现 | | |
|  |  | | |
|  |  |  |  |
| 学院 |  | 年级 |  |
| 专业 |  | 学号 |  |
| 姓名 |  | | |
| 完成时间 |  | | |
| 指导老师 | 刘坤 | | |

2025 年 月 日实验题目

**1. 实验目的：**

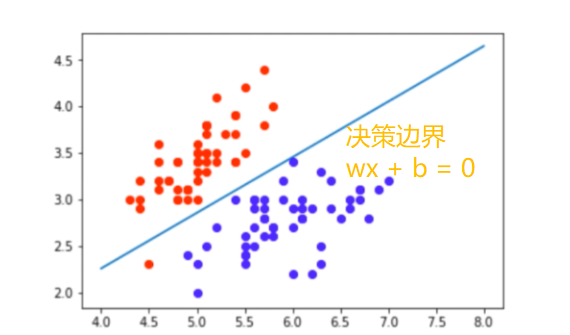
逻辑回归（Logistic Regression），又叫逻辑斯蒂回归，是机器学习中一种十分基础的分类方法，由于算法简单而高效，在实际场景中得到了广泛的应用。本次实验中，我们将探索逻辑回归的原理及算法实现，并使用 scikit-learn 构建逻辑回归分类预测模型。为了深入理解线性回归的原理、掌握其应用方法，并探究其在实际问题中的表现，我们进行实验。

**2. 实验原理**

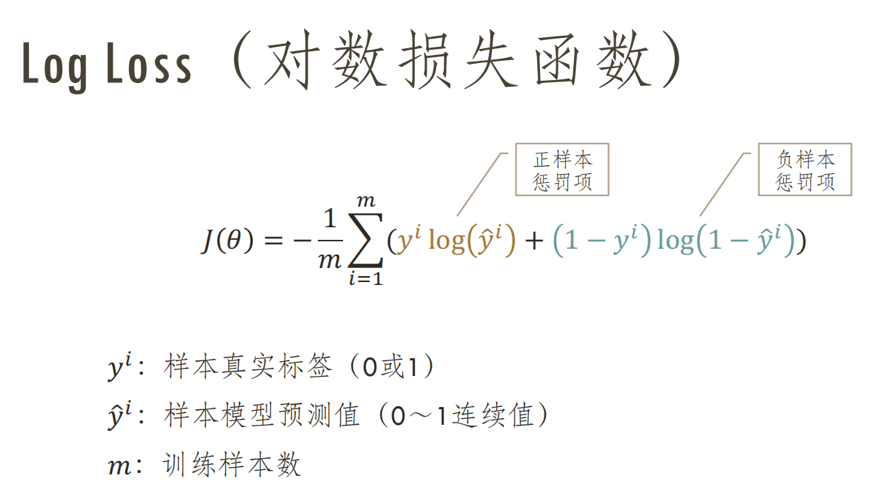
通过在线性回归模型中引入Sigmoid函数，将线性回归的不确定范围的连续输出值映射到（0，1）范围内，成为一个概率预测问题。逻辑回归的目标函数：

我们输入数据特征x，乘以一一对应的模型权重w后求和，通过输出层神经元激活函数σ（sigmoid函数）将（wx + b)的计算后非线性转换为0~1区间的概率数值后输出。学习训练（优化模型权重）的过程是通过梯度下降学到合适的模型权重[W]，使得模型输出值Y=sigmoid(wx + b)与实际值y的误差最小。

sigmoid函数是一个s形的曲线，它的输出值在[0, 1]之间，在远离0的地方函数的值会很快接近0或1。对于sigmoid输出作为概率的合理性，可以参照如下证明：

逻辑回归模型本质上属于广义线性分类器（决策边界为线性）。这点可以从逻辑回归模型的决策函数看出，决策函数Y=sigmoid(wx + b)，当wx+b>0，Y>0.5;当wx+b<0，Y<0.5，以wx+b这条线可以区分开Y=0或1（如下图），可见决策边界是线性的。

学习目标: 逻辑回归是一个经典的分类模型，对于模型预测我们的目标是：预测的概率与实际正负样本的标签是对应的，Sigmoid 函数的输出表示当前样本标签为 1 的概率，可以表示为. 当前样本预测为0的概率可以表示为。对于正样本y=1，我们期望预测概率尽量趋近为1 。对于负样本y=0，期望预测概率尽量都趋近为0。也就是，我们希望预测的概率使得下式的概率最大（最大似然法）。我们对 P(y|x) 引入 log 函数，因为 log 运算并不会影响函数本身的单调性。则有：  
希望 log P(y|x) 越大越好，反过来，只要 log P(y|x) 的负值 -log P(y|x) 越小就行了。那我们就可以引入损失函数，且令 Loss = -log P(y|x)，得到损失函数为：

我们已经推导出了单个样本的损失函数，是如果是计算 m 个样本的平均的损失函数，只要将 m 个 Loss 叠累加取平均就可以了：

这就在最大似然法推导出的lr的学习目标——交叉熵损失（或对数损失函数），也就是让最大化使模型预测概率服从真实值的分布，预测概率的分布离真实分布越近，模型越好。可以关注到一个点，如上式逻辑回归在交叉熵为目标以sigmoid输出的预测概率，概率值只能尽量趋近0或1，同理loss也并不会为0。

**3. 实验步骤**

需求：Python语言，依赖库sklearn, numpy.

逻辑回归模型虽然简单，但整体建模流程并不简单。逻辑回归的建模流程如下：

一、数据预处理：1.1.输入输出线性化1.2.输入变量归一化

二、模型训练：2.1.使用逐步回归进行建模；2.2.使用正则化进行训练(可选)

三、模型评估：3.1.模型评估；3.2.模型系数检查

**4. 实验代码**

**5. 实验结果**

**6. 实验总结**