# 问题

本项目研究机器学习中的逻辑回归(Logistic Regression)模型的最优参数。在一个二分类问题中，对于一个样本, 逻辑回归模型假设该样本属于正类的概率为

属于负类的概率为

给定一个数据集，其中。求在给定数据集S下，似然函数值最大的参数和

# 建模

该问题的变量为和。

似然函数

对于逻辑回归模型

故

故优化目标为

对目标函数取负对数，则问题等价于

本项目将公式作为目标函数，对该无约束优化问题进行求解。

# 求解

下面使用梯度下降法对问题进行求解。

对于点，目标函数在该点的偏导数分别为

其中,

故梯度下降法的框架为

|  |
| --- |
| 梯度下降法 |
| 输入: 初始点  数据集S=  终止精度，Wolfe-Powell线搜索参数 |
| 1. 若,停止, 得解, 否则转步2 2. 计算下降方向 3. 使用Wolfe-Powell线搜索确定步长，使得 4. 令, ,   k=k+1, 转步2 |

# 实验

本项目使用的数据集来源于UCI数据集中的iris数据集(<https://archive.ics.uci.edu/ml/datasets/iris>)，在该数据集中有三种鸢尾花，在实验中只使用其中的两种: Setosa和Versicolour. 两类鸢尾花共100个样本，包含4个特征值。

源码中实现了armijo和Wolfe-Powell线搜索(linear\_search模块)，无约束下降算法的一般框架(gradient\_method模块)并实现了LogisticRegression的最速下降算法（logistic\_regression模块）。main模块加载数据集并用LogisticRegression模型进行训练。

经过最速下降算法之后，参数变为w=(5.05, 27.41, -45.68, -19.16), b=4.47，目标函数（公式7）的最小值为1.24e-11。在程序的训练样本中，该参数将所有的样本正确分类，准确率为100.0%