

Lab01使用说明

loan.ipynb

运行部分，有如下几个步骤(逐次运行即可)：

1. 从 `loan_data_set.csv` 读入数据
 2. 数据处理（编码、均值代替空值、标准化、分层随机取样、划分Xy）
 3. 调用 `Logistic.py` 中的逻辑回归模型训练，并评估在测试集上的准确率
 4. 更改参数进行训练测试、比较
-

Logistic.py

逻辑回归模型，它的成员及使用方法：

- 成员变量
 - `coe`：拟合系数。默认为 `FALSE`，需调用 `fit` 函数以对其赋值。
 - `threshold`：分类阈值。默认为 0.5，可调用 `verification` 函数对其赋新值。
- 成员函数：
 - `__init__(self, penalty="l2", gamma=0, fit_intercept=True):`
构造函数。用于定义和初始化模型（本程序只用了默认的构造方式）
 - `sigmoid(self, x):`
sigmoid 函数。回归函数 $\frac{1}{1+e^{-z}}$
 - `fit(self, X, y, lr=0.01, tol=1e-7, max_iter=1e7)`
拟合函数，输入训练集数据和训练参数，作 loss function 图并得到拟合系数 `coe`
 - `verification(self, x_validation_data, y_validation_data)`
在验证集上验证比较，得到分类阈值 `threshold`（可省略此步）
 - `predict(self, X, threshold=0.5)`
在拟合后，对测试集进行分类（1或0），默认分类阈值 0.5。
 - `evaluate(self, y_test, y_predict)`
评估，将分类预测和真实情况比较，返回分类准确率。

补充：类外编写了求损失函数和损失函数梯度的函数，以实现 `fit` 功能。