

实验二 Logistic Regression

一、说明

- 实验采用 jupyter notebook, 请填写完代码后提交完整的 ipynb 文件
- 文件命名规则: 班级_姓名_ML2018_HW1.ipynb, 如计科 1701_张三_ML2018_HW2.ipynb
- 提交方式: 采用在线提交至:
<http://pan.csu.edu.cn:80/link/9F4CE22006E26B34845D7DB15BA69A3D>
- 实验提交截止日期: 2018.10.21 24:00

二、实验内容

本次实验我们使用 `pima-indians-diabetes` 数据集, 利用随机梯度下降算法对数据集进行 logistic 回归模型进行训练。回归模型分析实际问题, 所研究的变量往往不全是区间变量而是顺序变量或属性变量。在 `pima-indians-diabetes` 数据集中, 每一行数据包括八个特征, 一个标签 (0 或 1), 特征分别如下:

- #1 怀孕次数
- #2 口服葡萄糖耐量试验中血浆葡萄糖浓度
- #3 舒张压 (mm Hg)
- #4 三头肌组织褶厚度 (mm)
- #5 2 小时血清胰岛素 ($\mu\text{U}/\text{ml}$)
- #6 体重指数 ($\text{kg}/(\text{身高}(\text{m}))^2$)
- #7 糖尿病系统功能
- #8 年龄 (岁)
- #label 是否患有糖尿病

利用这些数据, 我们将利用随机梯度下降算法训练一个逻辑回归模型, 并测试模型的准确度。

利用牛顿法实现逻辑回归模型的参数求解，理解牛顿法的数学原理，并用代码实现牛顿法求解。

用 **Softmax Regression** 来处理多分类问题，数据特征向量为二维，标签（0-3）表示种类。用随机数据作为测试集，来看训练后的模型对测试集上的数据的泛化能力。

三、实验目标

- 掌握随机梯度下降算法原理及用 python 实现该算法。
- 理解逻辑回归模型及其中的数学原理。
- 掌握 python 基本数据操作
- 实现牛顿法优化参数
- 实现 Softmax Logistic

四、实验操作步骤

1. 启动 jupyter notebook

参考实验一，打开文件 `LogisticRegression.ipynb`。

2. 完成实验任务

- **任务 1** 将获取的数据集种的数据类型由 `str` 转换为 `float`
- **任务 2** 找出数据集每个特征的最大值和最小值
- **任务 3** 对数据进行归一化处理
- **任务 4** 对数据进行加权求和
- **任务 5** 利用随机梯度下降算法对参数更新
- **任务 6** 在训练集上对模型进行训练，并利用测试集进行测试，返回预测后的值。
- **任务 7** 求出预测的准确度
- **任务 8** 画出数据分布图
- **任务 9** 利用牛顿法对参数进行更新
- **任务 10** 绘制出决策边界

- 任务 11 利用梯度下降法对参数进行更新，获得优化后的参数。