

一、实验名称：Logistic 回归

回归：意思是最佳拟合，Logistic 回归是我们首次接触优化算法，相对于之前的算法思想，我们需要从数据里面学习到一组能够最好的描述观测数据的参数。很明显，我们几乎不能从数据里面获得较好的参数，需要一步一步的向最合适的参数移动，这个学习的过程也就是最优化学习的过程。

二、任务描述：

本节通过 logistcs 回归算法实现分类任务，选择的任务为：预测患有疝病的马的存活问题（此处选择的数据样例仅供参考，也可以选择任何测试数据）。

三、实验内容

1. 准备数据
2. 处理缺失数据
3. 可视化展示
4. 算法实现

四、实验环境和编程工具

系统环境： Windows all/Linux

程序语言： Python

编译工具： PyCharm

第三方包： pickle, matplotlib, operator, numpy, scikit-learn

五、实验步骤

1. 读取并分析数据集
将数据集按比例分为训练集、测试集。
2. 数据可视化
3. 分别在训练集、测试集上进行数据可视化，探索数据分布情况
4. 测试算法
自己构建一个 logistic 回归分类算法（也可以直接从 scikit-learn 里面调用现成的算法），在训练集上训练，在测试集上，评估模型效果，计算模型准确率、错误率
5. 画出决策边界
将学习到的模型参数可视化为决策边界，直观的理解模型运行原理。