

第三次编程

主讲老师：江 瑞

学生：姓名（学号）

1 题目介绍

鸢尾花卉数据集是一个多重变量分析数据集，在机器学习领域被广泛使用。

通过花萼长度 (sepal length)、花萼宽度 (sepal width)、花瓣长度 (petal length)、花瓣宽度 (petal width) 4 个属性预测鸢尾花卉属于山鸢尾 (Setosa)、杂色鸢尾 (Versicolour)、维吉尼亚鸢尾 (Viginica) 三个种类中的哪一类。样本数量 150 个，每类 50 个。

2 编程要求

1. 参看课件，推导用随机梯度下降法求解一元 Logistic 回归的过程。
2. 编程实现该随机梯度下降算法，根据花瓣长度 特征，对数据集中的山鸢尾 和维吉尼亚鸢尾 使用 Logistic 回归进行二分类。
3. 使用 Accuracy、Sensitivity、Specificity、Recall、Precision、F1、auROC 等指标评价你的分类方法。
注：需要分类阈值的指标，取 0.5 作为分类阈值。
4. 使用 scikit-learn 中的LogisticRegression 分类器求解该二分类问题，与自己实现的效果进行对比。

3 提交说明

为方便起见，脚本 code.py 中已经为你实现了数据加载、特征提取和数据集划分的相关代码，你可以根据自己的需要使用、改进。

请提交完成的代码文件及实验报告。

本次作业截止日期：2022 年 11 月 23 日（周三）晚 12 点（两周后）