

第三次编程

主讲老师：江 瑞

学生：姓名（学号）

1 题目介绍

MNIST 数据库 (Modified National Institute of Standards and Technology database) 是美国国家标准与技术研究院收集整理的大型手写数字数据库，在机器学习领域被广泛使用。数据库中的每张图片由 28×28 个像素点构成，每个像素点用一个灰度值表示，原始数据中将 28×28 的像素展开为一个一维的行向量（每行 784 个值）。图片标签为 one-hot 编码：0-9。

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，以每张图中白色像素点的比例为特征，对数据文件中的数字 0 和数字 1 使用 Logistic 回归进行二分类。
- (3) 使用 Accuracy、BER、MCC、Sensitivity、Specificity、Recall、Precision、F1、auROC、auPRC 等指标评价你的分类方法。
- (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。

为方便起见，你可以直接使用我们准备好的 MNIST 数据集 `mnist.mat`，也可以自行下载。

脚本 `code.py` 中已经为你实现了数据加载和特征提取的相关代码，你可以根据自己的需要使用、改进。

3 提交说明

请提交完成的代码文件及实验报告。

本次作业截止日期：2021 年 11 月 25 日（周四）晚 12 点