

人工智能基础算法 第二次作业

2024 年 9 月 25 日

题目内容：

MNIST 是经典的手写数字数据集，每张图像大小均为 28×28 像素，对图像中的数字进行膨胀与腐蚀等变换操作，即可得到粗体和细体两种类型的手写数字。本次作业任意选择其中的 2000 张，将其均分为两部分，分别记为数据集 A 和数据集 B。

请使用 Python, Matlab, R 或其他编程语言完成任务并撰写简要报告，提交作业时需要包含所有源码以及报告。

作业要求：

1. （15 分）试推导 Ridge Regression 的解，可参考课件 LN3 的 14 页、26 页，请将推导过程附在报告中。
2. （10 分）利用线性回归，实现对手写数字的粗体和细体进行分类。测试并报告以 A 为训练集、B 为测试集的分类准确率。请根据原理自行编程，若直接使用现有集成工具包，只得 40% 的分数。
3. （15 分）利用 Ridge 回归，在训练集 A 中使用 10 折交叉检验法，测试并报告采用不同正则化系数 λ_i 的分类准确率，进而得到最佳超参数 λ ，并在测试集 B 下测试分类准确率。请根据原理自行编程，若直接使用现有集成工具包，只得 40% 的分数。
4. （20 分）调用 Lasso 回归，在训练集 A 中使用 10 折交叉检验法，测试并报告采用不同正则化系数 γ_i 的分类准确率，进而得到最佳超参数 γ ，并在测试集 B 下测试分类准确率。

5. （20 分）调用 SVM，在训练集 A 中使用 10 折交叉检验法，测试并报告采用不同超参数 C_i 的分类准确率，进而得到最佳超参数 C ，并在测试集 B 下测试分类准确率。
6. （20 分）比较并讨论上述四种方法的结果。

说明：

1. 作业附件 morpho_mnist 文件夹中提供了训练集 A 和测试集 B 的原始数据和标签（.npy 格式）。
2. 提交作业时，请将作业要求 2-5 写成四个不同的文件，并根据“序号_回归方法”命名。
3. 线性回归和 Ridge 回归请根据原理自行实现，否则将扣去相应的分数；Lasso 回归和 SVM 可直接调用现有集成工具包。
4. 使用 10 折交叉检验时，注意采用合适的遍历超参数策略，且交叉检验只能在训练集上实现。
5. 提交作业时请将全部代码及.pdf 格式的报告压缩为同一个.zip 文件，**严禁抄袭，否则本次作业将会被记为 0 分！**