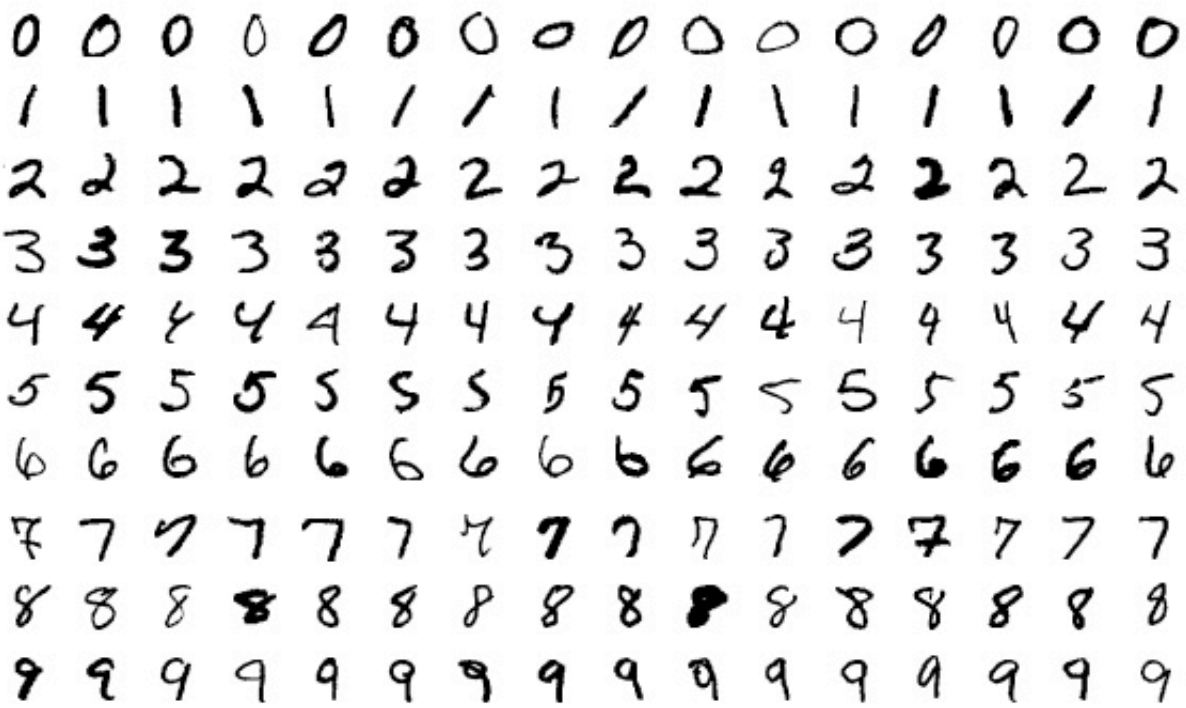
Regression experiments on MNIST dataset

The MNIST (Modified National Institute of Standards and Technology) dataset is a large database of handwritten digits, widely used for training various image processing systems and machine learning models. It was created by "re-mixing" samples from the original NIST dataset and has become a benchmark for evaluating the performance of image classification algorithms.



Key Parameters The MNIST dataset can be downloaded from this link: https://jbox.sjtu.edu.cn/l/G1gcEU

- MNIST contains 60,000 training images of handwritten digits and 10,000 test images.
- The dataset consists of grayscale images of size 28x28 pixels.
- MNIST is widely used for training and testing in the field of machine learning, particularly for image classification tasks.

Assignment Content

- Self-study Section 8.6 "Classification" from the textbook, focusing on linear classifiers (linear discrimination).
- For the MNIST dataset, from the training data, consider the following digit pairs: 4 and 9, 4 and 6, 0 and 1, 2 and 7. For each pair, use the linear classifiers from Section 8.6 to classify them. Optimize the classifier on the training data and test its performance on the test data. Report the best results.

- Incorporate the regularization methods from Chapter 6 to improve the model and see whether the results can be further enhanced.

Group Organization

- Each group can have up to 5 members.
- It is recommended that group members have clear roles and document their contributions, such as modeling, programming, and report writing.

Assignment Submission

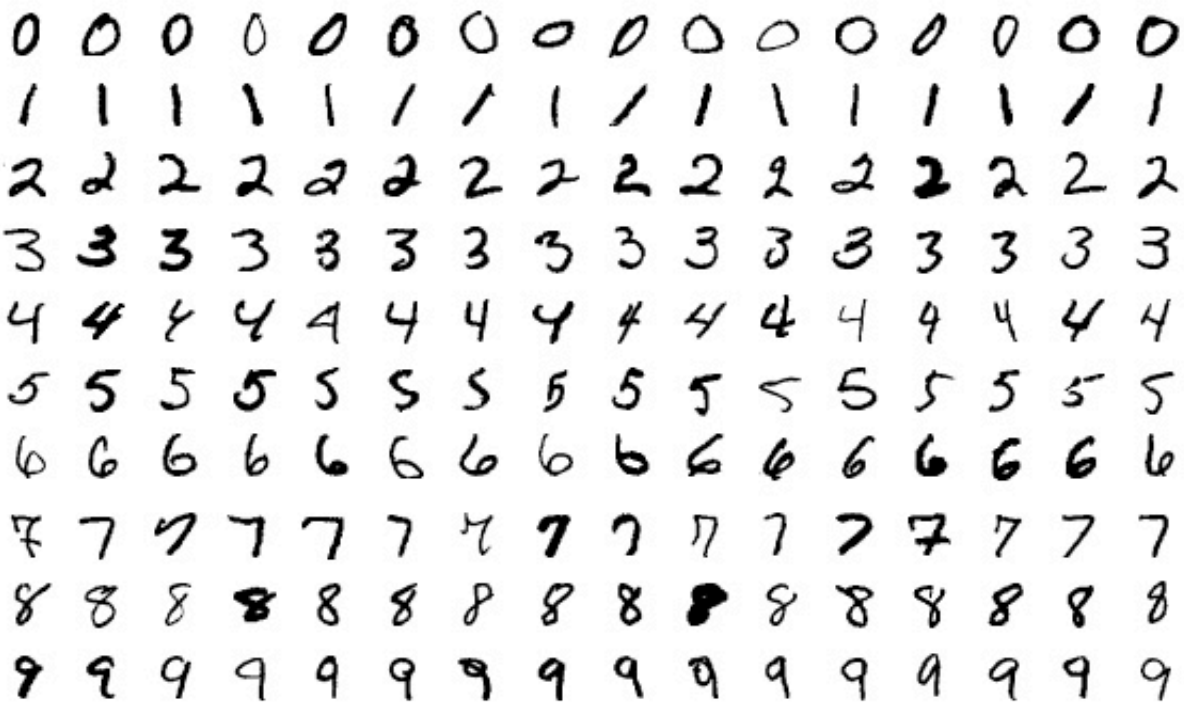
- Each group must write a report (in PDF format). The first page of the report should include the following information: Group Member Names and Each Member's Contribution.
- The group must submit executable source code (with necessary documentation), data, and the report as a zip/rar file uploaded to the system. (Only one person from each group needs to upload the submission.)

Notes

- There are many existing models for classifying the MNIST dataset with high accuracy. You may refer to related results, but it is strictly forbidden to directly use existing results (e.g., you cannot directly use results from sklearn).
- Plagiarism is strictly prohibited.

中文描述

MNIST（修改后的美国国家标准与技术研究院）数据集是一个大型手写数字数据库，常用于训练各种图像处理系统和机器学习模型。它是通过对 NIST 原始数据集中的样本进行 "重新混合 "而创建的，已成为评估图像分类算法性能的基准。



主要参数 MNIST数据可从此链接https://jbox.sjtu.edu.cn/l/G1gcEU 获取

- MNIST 包含 60,000 张手写数字训练图像和 10,000 张测试图像。
- 数据集由大小为 28x28 像素的灰度图像组成。
- MNIST 广泛用于机器学习领域的训练和测试，尤其是图像分类任务。

作业内容

- 自学教材8.6节“Classification”中的线性分类器（linear discrimination）内容。
- 对于MNIST数据集，从训练数据中，分别考虑数字组合：4和9，4和6，0和1，2和7。对于每一个数字组合，选择8.6节中的线性分类器对他们进行分类，在训练数据上优化分类器，并在测试数据上测试该分类器的表现，选择最优的结果汇报。
- 可结合第6章的正则化方法对模型进行改进，看是否可以进一步提高结果。

小组组成

- 每组最多5人。
- 建议组内成员分工明确，并做好分工记录。比如建模，编程和撰写报告等。

作业提交

- 每组需撰写报告（PDF格式），在报告的首页需要注明一下信息：***成员名***，以及***每个人的贡献***。
- 每组讲可运行的源程序（请提供必要的documentation），数据以及报告打包为zip/rar文件上传至系统。（每组仅需一人上传）

注意事项

- 对于MNIST数据的分类模型众多且精度很高，可以参考相关结果，但严禁直接使用已有结果（比如不能直接使用skitlearn的结果）。
- 严禁互相抄袭。