README

文件说明

手写体数据集包含两个文件：

1. mnist_train.csv
    - 训练集，包含 60,000 个训练样本。
2. mnist_test.csv
    - 测试集，包含 10,000 个测试样本。

.csv 文件结构说明：

- **第 1 行是表头**，介绍了文件中各列表示的信息；**其余的每一行对应一个样本**，各样本包含 785 列。
- **第 1 列是标签**，值为 0 或 1；**其余 784 列是手写体的像素值**，值的区间是 0 到 255。

TODO

1. 编写代码处理输入文件，构建好训练集与测试集。
    1. 训练集：训练集的构建应仅使用 mnist_train.csv 中的数据，不可包含 mnist_test.csv。
    2. 测试集：根据 mnist_test.csv 中的样本构建测试集。
    3. 验证集（可选）：可以自行从训练集中划分出一部分样本作为验证集，但不可使用测试集中样本在训练过程中进行验证。
2. 将构建好的训练集作为输入，根据 Assignment 3 的要求**手动**实现 K-Means 算法及 EM 算法训练的 GMM 模型，固定聚类簇类数为 10 类。
3. 在训练完成后针对测试集进行模型性能测试，使用聚类精度作为聚类性能的评价指标。
4. 根据 Assignment 3 的要求完成实验报告。