

实验二：KD 树数字识别

实验目的

1. 完成 KD 树算法，并利用实现的算法完成数字识别任务
2. 对所建模型进行分析评判。

实验数据

MNIST 数据库是由 Yann 提供的手写数字数据库文件，下载地址为：[MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges](#)。(其中包含数据集格式等详情)，主要包含了 60000 张的训练图像和 10000 张的测试图像



实验内容

基础内容：

1.基于 kd 树的 knn 的实现

熟悉 kd 树的原理与构建过程。

2.MNIST 数据集分类

根据提供的代码，理解代码的计算流程，将代码补充完整.(也可以选择自己另写一个完成，这里不做限制)，并完成对模型的训练和分类，最终输出的准确率理论上与 sklearn 的结果是一致的（需要补充的代码在代码中用 blank 标出，共有 4 处）

3.对 knn 进行超参数的搜索

对 knn 分类器进行超参数的搜索，选取你的最优的超参数下的 knn 分类器的优化结果

选作内容

1. 优化 kd 树

尝试使用不同的策略来构建 kd 树，使得在分类阶段可以有更快的分类效率。

作业提交时间：

2023.11.13，周一，23:59

提交内容： 报告（pdf）和代码（zip），请确保你的代码清晰可读、可复现、无 bug、无特殊环境依赖，无法复现 的代码会极大影响你的得分。

评分依据：

实验报告占整个实验分数的 60%，其中实验报告中的实验分析占整个实验分数的 30%；
代码部分占整个实验分数的 40%

很好的完成基础内容部分的工作即可以拿到满分。

最终的实验得分计算方法为：

$$\text{Min}(\text{基础内容得分} + \text{选作内容得分}, 100\%)$$

本次实验占课程总成绩的 6 分

代码语言：不限