

Machine learning: Homework #1

Due on April 20, 2020

Problem 1

自己动手实现 KNN 算法。语言限定为 python。数据集使用 MNIST 数据集。

数据集介绍：MNIST 手写数字数据集是具有 60,000 个示例的训练集和 10,000 个示例的测试集。它是 NIST 提供的更大集合的子集。数字已经过尺寸标准化并以固定尺寸的图像为中心。对于想要在真实数据上尝试学习技术和模式识别方法，同时在预处理和格式化方面花费最少的人来说，它是一个很好的数据库。数据集链接：<http://yann.lecun.com/exdb/mnist/>

KNN 结果展示方式：

- 输入若干测试图片，输出对应每张图片 k 近邻的图片。
- 绘制 knn 算法的训练 misclassification rate 曲线，并做出分析
- 可以自由发挥（人脸数据集等等）

注意：你可能需要选择 k

最终结果（算法测试以及训练曲线）汇总在一个文档（pdf/doc/docx/ppt etc.）中即可，待检查。
实验报告模板已上传

如果你不熟悉 python: [廖雪峰 python3 教程](#)

如果你不熟悉 numpy: [python-numpy 教程](#)