

一、author:

- name: \*\*\*  
- email: 2018212114@mail.hfut.edu.cn

二、代码参考:《机器学习实战》(Peter Harrington)

三、环境:

- OS: Linux  
- 语言: Python3.5  
- python 依赖:  
    torch1.3.1  
    torchvision0.4.2  
    numpy1.7.3  
    pandas0.24.2

四、文件结构

- train : 训练集数据 (图片)  
- test : 测试集数据(图片)  
- train2 : 训练集数据 (txt 文件)  
- test2 : 测试集数据 (txt 文件)  
- results : 训练和测试 log  
- weights : 模型参数(支持向量)  
- src : 代码存放  
    - SVM.py : 支持向量机算法实现(参考《机器学习实战》)  
    - predictDigits.py: 训练模型并用测试集测试  
    - transformData.py 和 splitDataSet.py 为数据预处理脚本 (运行需更改路径)

五、注:

- 1)运行 predictDigits.py 文件即可对测试集数据进行预测;
- 2)对数据进行预处理的 python 脚本文件需更改文件路径后, 方可运行;
- 3)train2 和 test2 存放的是将图像数组转换为 0,1 矩阵存储的 txt 文件, 因此只要维度为 28\*28, jpg 和 txt 文件都可作为数据。
- 4)由于 SVM 中使用高斯核, 矩阵维度较高, 受内存限制代码不支持过多数据, 经测试 2000 张以内即可得到较好效果。