

# H1 MNIST

机器学习概论lab3

Author:@Rosykunai

Date2024 年 11 月

## MNIST

Deadline

实验环境

提交要求

成绩评定

## H2 Deadline

2024.12.15 23:59

## H2 实验环境

Conda (推荐):(Python=3.9)

```
conda activate ml24
pip install -r requirements.txt
```

本次实验需要用到pytorch,如果你想要安装GPU版(需要N卡)的pytorch,参考下面的方法:

换源:

```
conda config --add channels
https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkgs/free/
conda config --add channels
https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkgs/main/
conda config --add channels
https://mirrors.tuna.tsinghua.edu.cn/anaconda/cloud/pytorch/
conda config --add channels nvidia
conda config --set show_channel_urls yes
```

安装GPU版的pytorch:

```
conda install pytorch torchvision torchaudio pytorch-cuda=12.4 # CUDA版本
```

上述配置完成后,运行 `load_resources.py`,最后会输出你的pytorch使用的设备。

\*没有N卡不影响完成本次实验,我们只会用一些小模型做推理,CPU运行时间稍长一点(<3min)

## H2 提交要求

上传一个 `学号+姓名+LAB3.zip` 文件, 包含:

- `results/{datetime}`
  - `gmm`
    - `config.json`
    - `gmm.safetensors`
  - `pca`
    - `config.json`
    - `pca.safetensors`
  - `config.yaml`
- `submission.py`
- `report.pdf`

## H2 成绩评定

- Code(40%): 见MNIST.pdf
- Performance(30%):

我们在原始数据空间(784维)中使用 `davies_bouldin_score` 度量聚类的性能:

$$grade = 30 * \frac{DB_{sklearn}}{DB_{yours}}$$

\*如果你的模型性能超越了 sklearn, 你会在这部分得到满分。

- Report(30%):
  - 记录实验流程 (2%)
  - 记录你调试超参数的过程 (5%)
  - 报告最好的聚类 and 生成结果(输出的图片) (2%)
  - 回答问题 (20%): 见 GPU Performance.pdf
  - 反馈 (1%): 见 GPU Performance.pdf
- Warning:
  - 请独立完成实验, 不得抄袭
  - 如果助教根据你提供的超参数无法复现你所报告的结果, 你不在 Performance 部分得到任何分数