



Tutorial for conventional Machine learning (for Cast usage)

How to set up

Programming language

Mat or python ???

Matlab:下限高上限低

Python:下限低上限高

IDE

Matlab or Pycharm or Vscode ???

Matlab整体很方便上手，但是运行速度缓慢而且对第三方兼容性差，因此当大家走出校园或者在做一些软件方向的前沿研究时候会碰到很多问题。

pycharm和vs code可以运行python语言，pycharm整体需要自己手动设置的东西可能少一点，vs code被称为宇宙IDE，其出色的社区extension功能为大家带来了很多很有功能性的插件，而且在环境改变上vs code会更加迅速，因此这里推荐大家使用vs code正式课堂演示也是使用vs code作为展示。

How to download and setup vscode?

建议大家参考一些视频教学或者问GPT，这里给出一些资源：

B站搜索：【教程】VScode中配置Python运行环境

B站视频对应文档：<https://www.lanzouh.com/iG5tn03iqwwh>

YouTube：<https://www.youtube.com/watch?v=D2cwvpJSBX4>

Tips:如果你的电脑室Unix系统则对应搜索自己电脑操作系统的视频教学或者文档

Environment set up

考虑到Python代码的使用最好不要完全依赖标准库，丰富的第三库和优秀的社区文化是python的特点，所以需要大家先正确安装pip

B站搜索:Python教程——手把手教你用pip安装第三方库，新手小白必看的菜鸟教程！

Youtube: <https://www.youtube.com/watch?v=oyjOBhwSI-8>

Optional: Conda

Doanlaod some packages

如果是只用pip的话:

```
pip install -r requirement.txt
```

如果是conda想新建环境的话:

```
conda create env -f environment.yaml
```

Let's dive into some ML models !!!

经典的ML模型我们在通常会先pip install scikit-learn，也就是我们常说的sklearn包，里面包含了大量的ML所需要的模型和一些评估函数，整体上手非常简单，但是sklearn一般比较适配pandas包（个人感觉），对数据要求也比较严格，所以我们在开始使用模型前往往需要做简单的数据处理，清洗数据或者自动填充，下面会介绍一些数模甚至是实际项目和研究中常用的模型

Traditional ones

- 可解释性强
- 训练速度快
- 对小数据集很友好
- 在经典结构化数据任务中表现依然很可靠

1. Linear Models
2. SVM, Support Vector Machine

- 3. Decision Tree
- 4. Ensemble Models

Novel ones

- 1. Boosting series
- 2. Neural Network

Some practice

Platform we can use

kaggle这类竞赛平台以及一些机构组织的学习机会，比如Datawhale、freecodecamp

The routine

(详见授课)