

机器学习：Python环境准备

李老师 2017-08-10

V1.0.0.0

The World's Local Training Provider

1

环境准备

2

机器学习toolkit

3

训练数据train data

前言

- Python平台是非常优秀的机器学习平台，但过去发展发生过稳定兼容差异问题（eg: 2.7与3.X部分语法不兼容；32位与64为运行环境差异），所以本着稳定兼容第一的原则配置运行环境
- 以下Python机器学习环境准备侧重培训学习
- 目前Python运行环境稳定性与兼容性最好的是2.7版本32位的
- 以下的环境安装仅是指导作用，详细的配置自行参考官方文档和本地实际
- Python基础学习参考：<http://www.runoob.com/python/python-tutorial.html>
- 开发环境建议：每个学员个人电脑安装完整的运行环境与开发环境；一是开发的友好性更好，运行效率高；二是在案例练习时，多人使用共享环境容易造成各种资源并发冲突；

1.1.1环境准备： 运行环境： 安装

- 硬件配置建议： cup i7+、 mem 8gb+、 ssd 120 gb+
- 操作系统Win7 pro 64位 / Win10 pro 64位
- Python安装包下载： <https://repo.continuum.io/archive/Anaconda2-4.4.0-Windows-x86.exe>
- 请下载2.7版本32位的安装
- 环境资源文件请安装在ssd硬盘分区上
- 环境资源权限建议为管理员，要求有读写文件执行各种操作的权限
- 例如本地安装根目录： D:\Anaconda2

1.1.2.环境准备： 运行环境： 配置环境变量

- 配置系统环境变量： PYTHON_HOME， 路径指向： D:\Anaconda2
- 配置系统Path变量， 在已有的后面增加如下3个路径： D:\Anaconda2; D:\Anaconda2\scripts; D:\Anaconda2\Library\bin
- windows命令行检查python、 pip、 conda、 jupyter、 cython命令是否正常
- Jupyter配置请参考官方文档， 主要配置登录密码与路径指向

1.2.环境准备：开发工具

- JetBrains PyCharm，官网下载最新旗舰版64位（强烈推荐）
- Jupyter，不建议共享环境使用，适合个人演示练习目的(本地安装)

二.机器学习toolkit（1）

- 安装机器学习包采用conda命令，请自行参考官方网站
- Python中央仓库地址：<https://pypi.python.org/pypi>
- 由于安装环境时，多数安装包已经安装，故以下更多的包可以更新，并安装或更新到与运行环境兼容的最新版本
- 若安装过程中有VC环境错误，请安装VCForPython27程序包，微软官方网站下载对应32位版本
- 基本包安装参考网站：<https://scipy.org/index.html>
- numpy
- matplotlib
- sympy
- pandas
- scipy
- pillow

二.机器学习toolkit（2）

- 机器学习包：scikit-learn (sklearn)

三.训练数据train data

- 训练数据，下载后请自行解压，目录路径必须是英文或者数字，且与Python环境在同一机器上
- 电影数据： <https://grouplens.org/datasets/movielens/100k/>