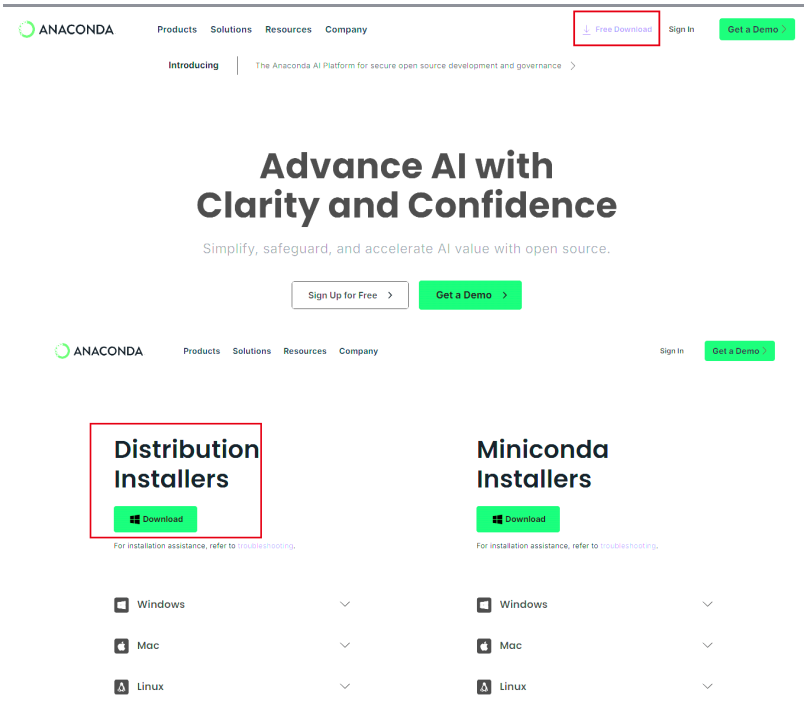
配置环境

作者：Doc.
日期：2025.7.11

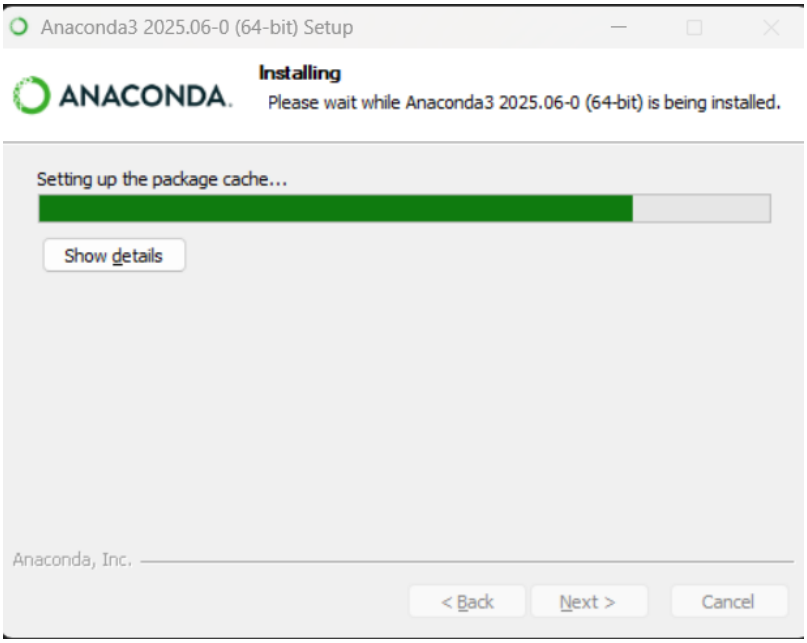
安装conda和配系统环境变量

第一步 下载安装包

下载地址为：https://www.anaconda.com/

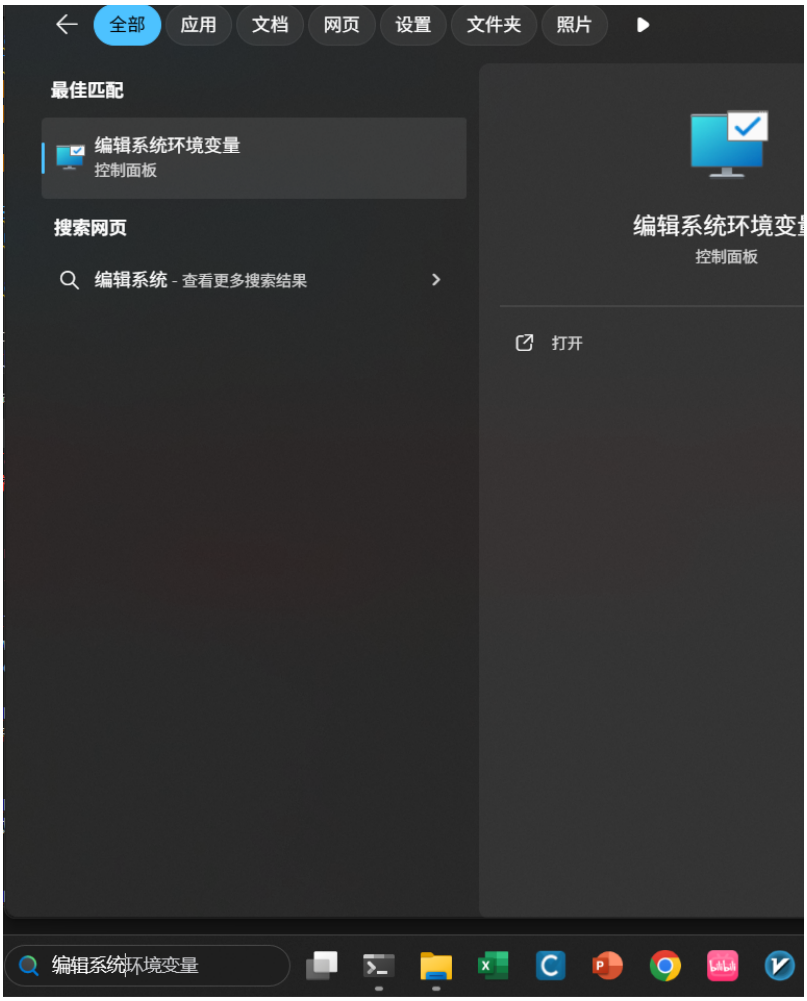


下载后安装：



第二步 把刚才安装好的conda配置到系统环境变量里面

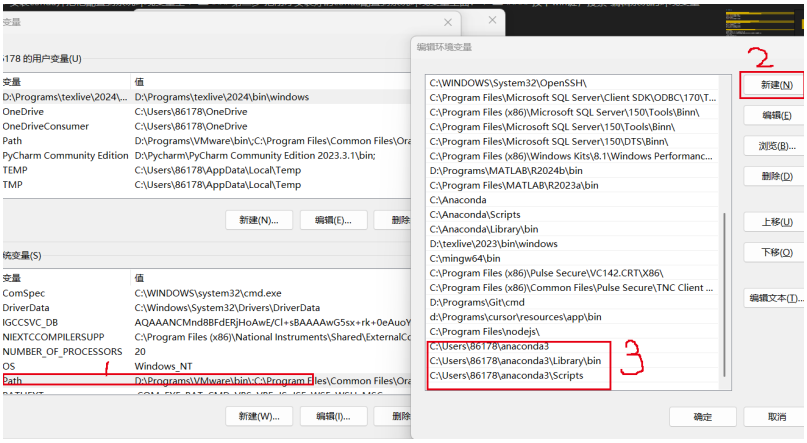
按下win键，搜索"编辑系统的环境变量"


全部
应用
文档
网页
设置
文件夹
照片
最佳匹配
编辑系统环境变量
控制面板
搜索网页
编辑系统 - 查看更多搜索结果
编辑系统环境变量
控制面板
打开
编辑系统环境变量

点击环境变量：


计算机名
硬件
高级
系统保护
远程
要进行大多数更改，你必须作为管理员登录。
性能
视觉效果，处理器计划，内存使用，以及虚拟内存
设置(S)...
用户配置文件
与登录帐户相关的桌面设置
设置(E)...
启动和故障恢复
系统启动、系统故障和调试信息
设置(T)...
环境变量(N)...
确定
取消
应用(A)

点击path，进入后新建三个环境变量，注意修改路径为自己刚下载的anaconda路径!



第三步 检查是否成功安装

输入 `conda --version`

如果得到了：

```
C:\Users\86178>conda --version
conda 25.5.1
```

则恭喜你，成功了!

使用conda创建一个虚拟环境，并在这个环境里装包

第一步 创建虚拟环境rl（你可以随意命名，可以把不是rl）

在任意打开一个终端

```
Windows PowerShell - conda

C:\Users\86178>conda create -n rl python==3.10
Retrieving notices: done
Channels:
 - defaults
Platform: win-64
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: C:\Users\86178\anaconda3\envs\rl

  added / updated specs:
    - python==3.10


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    openssl-1.1.1w             |       h2bbff1b_0         5.5 MB
    python-3.10.0              |       h96c0403_3        15.3 MB
    setuptools-78.1.1          |  py310haa95532_0         1.7 MB
    ucrt-10.0.22621.0          |       haa95532_0         620 KB
    vc-14.3                    |      h2df5915_10          19 KB
    vc14_runtime-14.44.35208   |      h4927774_10         825 KB
    vs2015_runtime-14.44.35208 |      ha6b5a95_10          19 KB
    wheel-0.45.1               |  py310haa95532_0         145 KB
    ---------------------------------------------------------
```

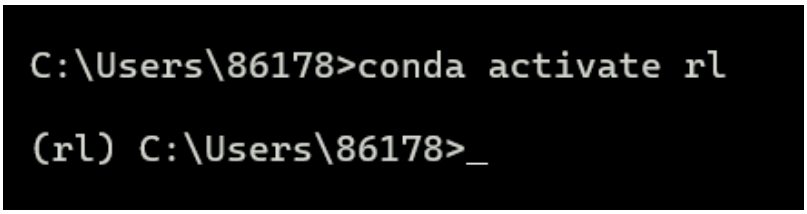
使用命令：

```
conda create -n rl python==3.10
```

这里为了在创建环境的时候指定了python解释器的版本，避免疏漏

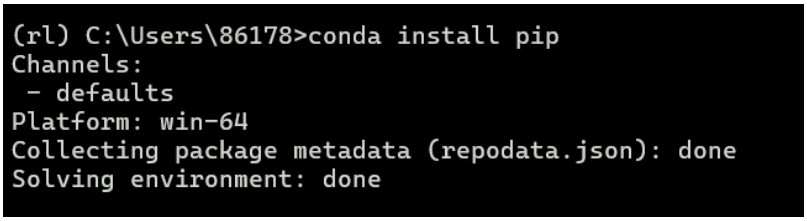
激活环境：

```
conda activate rl
```



至此你可以看到左边出现了小括号，指示着我们当前所处的环境是rl

使用conda命令安装包管理器pip，用来安装其余必要的库：



安装所有必要的包：

首先你需要用一次conda额外安装：

```
conda install -c conda-forge box2d-py
```

否则你将不能玩月球登录游戏呜呜呜

这里你可以选择两种方法去做：

方法一：

在你的环境里面命令行输入：

```
pip install torch gymnasium matplotlib
```

（没错，可以一行全写完）

以及记得：

```
pip install "gymnasium[toy-text]"
```

否则你将不能可视化游戏过程呜呜

方法二：

在命令行中切换到本路径，输入：

```
pip install -r requirements.txt
```

（按照我给好的包依行安装）

注意！！！！！！！！！ 库是在虚拟环境rl里面安装的，我们跑python代码所有的命令要在我们的环境里执行，否则没有解释器，也找不到相应的库

动手跑跑demo

注意，你需要先git clone（或者fork）：

```
git clone https://github.com/thuasta/summer-training-2025
```

在本部分中，为了避免路径的混淆，你需要始终保持在根路径下输入所有命令，即如下路径：

悬崖游走

```
python -m rl.runners.cliffwalking --algorithm dqn
```

最后一个参数可以改为sarsa, qlearning, dqn详情参照技术文档

摆锤平衡

```
python -m rl.runners.cartpole
```

或者

```
python ./rl/runners/cartpole.py
```

月球登录

```
python -m rl.runners.LunarLander
```

或者

```
python ./rl/runners/LunarLander.py
```

大家可以自行查阅一下这两个命令的区别

完成作业(选做)

策略梯度（Policy Gradient, PG）是强化学习中一种直接优化策略的方法，它面临一个显著问题：梯度估计的方差很大，导致训练不稳定、收敛缓慢。因此，研究并实现有效的方差减小技术，对于提升 PG 方法的实用性和效率具有重要意义。

作业目标：

- 理解策略梯度估计中的方差来源
- 任选一种经典的方差减小技术（baseline, GAE, reward normalization, advantage function 等）
- 将该技术应用于基本的策略梯度实现中
- 通过实验量化该方法对梯度估计方差的改善效果

FAQ

日后强化到一定境界，出现了顿悟时刻，请务必带本文档作者发rl顶会论文ww