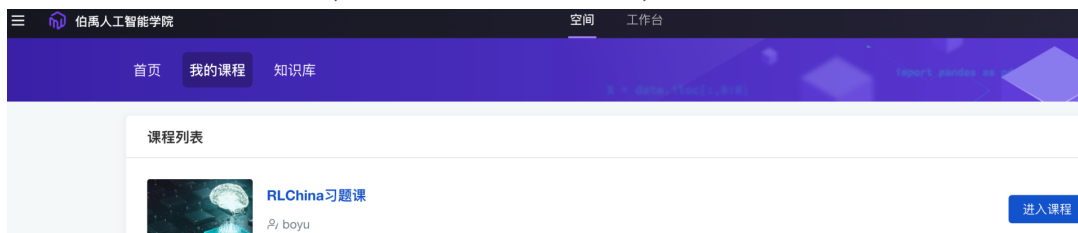


[RLChina习题课]和鲸平台

1. 通过链接：<https://www.heywhale.com/org/boyuai/register?source=href&key=297cedf5a38b9391b64ec220> 进入，填写邮箱，收到邮件后，注册和鲸平台

2. 进入“空间”-“我的课程”，可以看到RLChina习题课，点击“进入课程”



3. 在课程里能够看到4个课程文件，也就是4个notebook，分别是马尔可夫决策过程、动态规划算法、时序差分算法和DQN算法

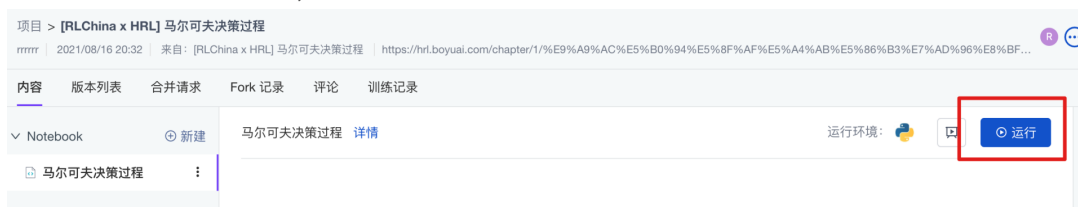
4. 选中一个notebook后，点击Fork按钮，就可以将notebook复制到自己的工作台



5. Fork完之后的notebook可以在“工作台”-“项目”处看到



6. 选择一个notebook后，点击“运行”按钮



7. 选择“计算资源”和“镜像”，点击运行

选择计算资源

计算资源

CPU 4核16G CPU 长时间版本

剩余：20小时

4核16G资源，工作区大小2G，单次连续使用时长48小时

CPU 2核 8G 基础资源

伯禹教育训练营专用资源，单次使用时长120min

申请更多资源

镜像

d2l-pytorch镜像

d2l-pytorch项目使用

没有合适的镜像？去定制

取消

运行

8. 等待镜像加载完成后

启动器

马尔可夫决策...

文件 编辑 查看 运行 Kernel 帮助

Markdown

d2l-pytorch镜像

获取 kernel 中...

In []:

马尔可夫决策过程

简介

9. 然后就可以在线运行代码了

启动器

马尔可夫决策...

文件 编辑 查看 运行 Kernel 帮助

Markdown

d2l-pytorch镜像

Python 3

马尔可夫决策过程

简介

马尔可夫决策过程 (Markov Reward Process) 是强化学习的基础。要学习好强化学习我们首先得明确掌握马尔可夫决策过程的基础知识。我们通常说的强化学习中的环境一般就是一个马尔可夫决策过程。与多臂老虎机不同，马尔可夫决策过程包含了状态信息以及状态之间的转移机制。如果我们要用强化学习去解决一个实际问题，我们第一步要做的事情就是能够把这个实际问题抽象为一个马尔可夫决策过程，也就是明确马尔可夫决策过程的各个组成要素。在本节

10. 左侧可以找到当前目录文件夹，下载和上传文件在这里进行

文件树

启动器

马尔可夫决策...

文件 编辑 查看 运行 Kernel 帮助

Markdown

d2l-pytorch镜像

Python 3

马尔可夫决策过程

简介

马尔可夫决策过程 (Markov Reward Process) 是强化学习的基础。要学习好强化学习我们首先得明确掌握马尔可夫决策过程的基础知识。我们通常说的强化学习中的环境一般就是一个马尔可夫决策过程。与多臂老虎机不同，马尔可夫决策过程包含了状态信息以及状态之间的转移机制。如果我们要用强化学习去解决一个实际问题，我们第一步要做的事情就是能够把这个实际问题抽象为一个马尔可夫决策过程，也就是明确马尔可夫决策过程的各个组成要素。在本节内容中，我们将从马尔可夫过程出发，一步一步进行介绍，最后引出马尔可夫决策过程。

马尔可夫过程 (Markov Process)

随机过程 (Stochastic Process)

随机过程是概率论的“动力学”部分。概率论的研究对象是静态的随机现象，而随机过程的研究对象是随时间演变的随机现象。例如，天气随时间的变化，城市交通随时间的变化。随机过程中，随机现象在某时刻 t 的取值被称为状态 S_t ，所有可能的状态组成状态集合 \mathcal{S} 。于是，随机现象研究的便