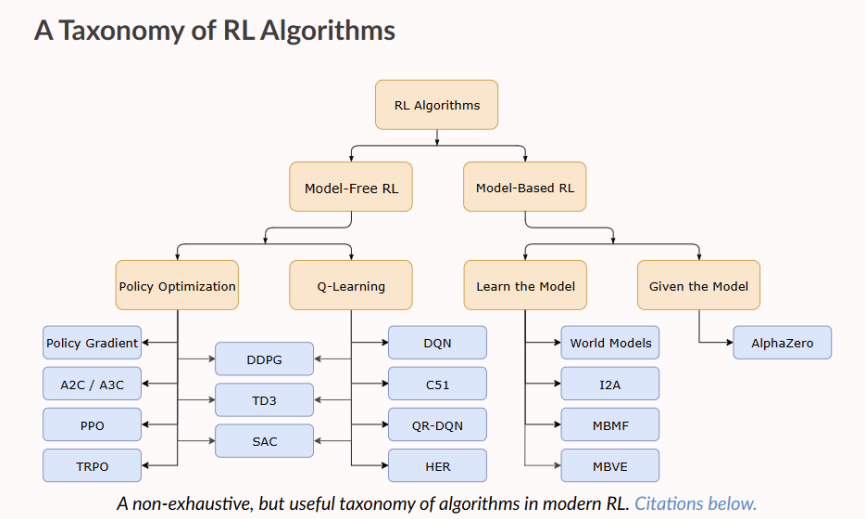
به نام او...

تمرین چهارم: Deep RL

## بخش اول - سوالات تحلیلی(15 نمره)

در این بخش به بررسی چند سوال تحلیلی از روش های مبتنی بر Deep RL میپردازیم.

**سوال اول-** الگوریتم های Deep RL یا مبتنی بر یادگیری Value هستند یا یادگیری مستقیم Policy. این دو دسته را با هم مقایسه کنید. به طور کلی مادامی که میتوان value را یاد گرفت، چه نیازی به یادگیری مستقیم Policy است؟ آیا دسته سومی که از دو رویکرد استفاده کند وجود دارد؟ اگر بله توضیح دهید.(5 نمره)




تصویر از وب سایت Open AI spinning up

**سوال دوم -** یکی از الگوریتم هایی که از ایده DQN استفاده میکند، Dueling DQN است، درباره این الگوریتم تحقیق کنید و تفاوت اصلی آن را DQN شرح دهید. مهم است که درباره نحوه عملکرد advantage function توضیح دهید. (5 نمره)

**سوال سوم -** الگوریتم A2C را بررسی کنید، در این الگوریتم چطور از ایده advantage استفاده شده است؟ این روش چه تفاوتی با الگوریتم Dueling DQN دارد که در قسمت قبل با آن آشنا شدید؟(5 نمره)

## بخش دوم - محیط پیاده سازی(5 نمره)

شما قبلا در hands-on 3 با کتابخانه GYM آشنا شدید، در این تمرین قرار است با محیط Cart Pole - v1 کار کنید.

**سوال چهارم-** در لینک مورد نظر به بررسی این محیط بپردازید، درباره Action space، Observation Space و ریوارد این محیط اطلاعات کامل را ارائه دهید.(3 نمره)

**سوال پنجم-** چرا استفاده از الگوریتم های Classic RL برای این محیط خوب نیست؟(2 نمره)

---

## بخش سوم - پیاده سازی الگوریتم(60 نمره)

در این بخش هدف پیاده سازی الگوریتم DQN و Duelling DQN میباشد.
ترجیحا از pytorch برای پیاده سازی های خود در این بخش استفاده کنید.

**سوال ششم-** الگوریتم DQN را از پایه و بدون استفاده از کتابخانه های آماده پیاده سازی کنید، عامل DQN شما باید محیط Cart Pole-v1 را یاد بگیرد. (10 نمره)
**الف)** در نهایت پارامترهای مورد استفاده خود را در یک جدول بیان کنید.(5 نمره)
**ب)** پس از یادگیری عامل، نمودار پاداش کسب شده در طول یادگیری توسط عامل در چند ران مختلف را در گزارش خود قرار دهید. نمودار مورد نظر باید شامل بازه اطمینان 95 درصد باشد.(5 نمره)
**ج)** از چند اپیزود تست عامل پس از یادگیری ویدیو تهیه کنید، این ویدیو باید در فایل Zip نهایی ارسال شما موجود باشد.(5 نمره)

**سوال هفتم-** الگوریتم Duelling DQN را از پایه و بدون استفاده از کتابخانه های آماده پیاده سازی کنید، عامل Duelling DQN شما باید محیط Cart Pole-v1 را یاد بگیرد. (10 نمره)
**الف)** در نهایت پارامترهای مورد استفاده خود را در یک جدول بیان کنید.(5 نمره)
**ب)** پس از یادگیری عامل، نمودار پاداش کسب شده در طول یادگیری توسط عامل در چند ران مختلف را در گزارش خود قرار دهید. نمودار مورد نظر باید شامل بازه اطمینان 95 درصد باشد.(5 نمره)
**ج)** از چند اپیزود تست عامل پس از یادگیری ویدیو تهیه کنید، این ویدیو باید در فایل Zip نهایی ارسال شما موجود باشد.(5 نمره)

**سوال هشتم-** آیا تفاوتی در همگرایی دو الگوریتم مشاهده میکنید؟ کدام بهتر و سریعتر محیط را یاد گرفته است، چرا؟ نتایج را تحلیل کنید(10 نمره)

---

**بخش چهارم - استفاده از کتابخانه Stable baseline 3 شامل (20 نمره)**

در پروژه های واقعی معمولا پیاده سازی الگوریتم های از پیش آماده کمتر ضروری است، به همین دلیل برای آشنایی شما با یکی از کتابخانه های مطرح در زمینه Deep RL این بخش پایانی، در نظر گرفته شده است

**سوال نهم -** از کتابخانه گفته شده الگوریتم A2C را که بررسی کردید روی محیط Cart Pole -v1 اجرا کنید. عامل A2C شما باید محیط Cart Pole-v1 را یاد بگیرد. (5 نمره)

**الف)** در نهایت پارامترهای مورد استفاده خود را در یک جدول بیان کنید.(5 نمره)

**ب)** پس از یادگیری عامل، نمودار پاداش کسب شده در طول یادگیری توسط عامل در چند ران مختلف را در گزارش خود قرار دهید. نمودار مورد نظر باید شامل بازه اطمینان 95 درصد باشد.(5 نمره)

**سوال دهم-** نتایج A2C را با DQN و مقایسه کنید، کدام الگوریتم بهتر است، چرا؟ (5 نمره)

---

**منابع:**


https://gymnasium.farama.org/environments/classic_control/cart_pole/

https://spinningup.openai.com/en/latest/

https://stable-baselines3.readthedocs.io/en/master/index.html

https://huggingface.co/learn/deep-rl-course/en/unit0/introduction

https://github.com/rlcode/reinforcement-learning/tree/master

**Dueling Network Architectures for Deep Reinforcement Learning**

**Asynchronous Methods for Deep Reinforcement Learning**

## نکات تمرین

- استفاده از LLM ها در این تمرین مشکلی ندارد. اما در صورت استفاده لطفاً منبع و prompt خود را ذکر نمایید تا تقلب محسوب نشود.
- مهلت ارسال این تمرین تا پایان روز یکشنبه 23 دی ماه 1403 خواهد بود.
- دقت کنید که در تمرین شما باید تصمیماتی از جنس **طراحی** بگیرید، مانند اینکه از چه شبکه ای استفاده کنید، هایپر پارامتر ها چه باشد و ... . این تصمیمات بر عهده شماست و تمرین در نهایت تنها یک پاسخ درست ندارد و اگر به مقصود(همگرایی و یادگیری عامل) رسیده باشید یقینا نمره را دریافت خواهید کرد.
- انجام این تمرین به‌صورت یک‌نفره است. اما بحث و گفت‌وگو در پیامرسان درس مانعی ندارد.
- لطفاً گزارش و کد تمرین و ویدیو یادگیری عامل را در قالب یک فایل zip در سامانه ایلرن بارگذاری کنید.
- در صورت وجود سؤال و یا ابهام می‌توانید از ایمیل زیر با دستیاران آموزشی در ارتباط باشید

- ali.ramezani.96@ut.ac.ir
- reyhaneh.ahani@gmail.com