

با توجه به پیشرفت روز افزون مدل‌های یادگیری عمیق و همچنین نتایج قابل توجه ادغام این مدل‌ها با کاربردهای یادگیری تقویتی، تمرین چهارم به بررسی این الگوریتم‌ها و مدل‌ها خواهد پرداخت. حل کردن مسائل یادگیری تقویتی با استفاده از شبکه‌های عصبی پیشینه‌ای قدیمی دارد و با توجه به پیشرفت سخت‌افزارهای محاسباتی در دو دهه اخیر، سرعت توسعه مدل‌های عمیق برای مسائل یادگیری تقویتی افزایش قابل ملاحظه‌ای داشته است. استفاده از شبکه‌های عصبی این امکان را به ما می‌دهد که از مسئله را با استفاده از یک مدل end to end حل کنیم. با توجه به این نکته کسب مهارت کار کردن با مدل‌های یادگیری عمیق و حل مسائل یادگیری تقویتی با استفاده از این مدل‌ها از مهارت‌های ضروری در زمینه یادگیری تقویتی می‌باشد. این تمرین مقدمه‌آشنایی شما با این مسائل را فراهم می‌کند و طبیعتاً کسب مهارت‌های بیشتر در این زمینه نیازمند مطالعه و تمرین بیشتر خواهد بود.

## سوالات تحلیلی

1. یک مورد از فواید و معایب الگوریتم‌های Deep RL را توضیح دهید.
2. یک مورد از مزایای الگوریتم Deep Q-Learning نسبت به الگوریتم Q-Learning را بیان نمایید.
3. یک مورد از دلایل استفاده از بافر تجارب را نام ببرید.

## محیط مورد استفاده در سوال پیاده‌سازی

در این تمرین شما با کتابخانه [Gymnasium](#) کار خواهید کرد. Gymnasium یک API استاندارد برای یادگیری تقویتی و مجموعه‌ای متنوع از محیط‌های مختلف است. قبلاً به عنوان Gym شناخته می‌شد. در این [لینک](#) توضیح ساده‌ای از نحوه‌ی استفاده از محیط‌های آن داده شده است. در این تمرین شما با محیط [Acrobot](#) آشنا خواهید شد و کار خواهید کرد. این سیستم از دو پیوند تشکیل شده است که به صورت خطی به هم متصل شده اند تا یک زنجیره را تشکیل دهند که یک انتهای زنجیره ثابت است. اتصال بین دو پیوند فعال می‌شود. هدف اعمال گشتاورها بر روی مفصل فعال شده برای چرخاندن انتهای آزاد زنجیره خطی بالای یک ارتفاع معین در حالی که از حالت اولیه آویزان شدن به سمت پایین شروع می‌شود.



دو پیوند آبی که توسط دو مفصل سبز به هم متصل شده‌اند. مفصل بین دو پیوند فعال می‌شود. هدف این است که با اعمال گشتاور بر روی محرک، انتهای آزاد پیوند بیرونی را برای رسیدن به ارتفاع هدف (خط افقی سیاه بالای سیستم) بچرخانید.

برای استفاده از این محیط با توجه به حجم زیاد محاسبات توصیه ما این است که با استفاده از محیط گوگل کولب، از قطعه کد زیر جهت نصب و import کردن کتابخانه‌ها مورد نیاز استفاده کنید.

```
!pip install gymnasium[classic-control]
```

```
import gymnasium as gym
```

```
env = gym.make('Acrobot-v1')
```

```
observation, info = env.reset()
```

```
for _ in range(1000):
```

```
    action = env.action_space.sample()
```

```
    observation, reward, terminated, truncated, info = env.step(action)
```

```
    if terminated or truncated:
```

```
        observation, info = env.reset()
```

```
env.close()
```

جهت آشنایی بیشتر با نحوه کارکرد این محیط می‌توانید از [ریبازیتوری گیت‌هاب](#) خود کتابخانه استفاده کنید.

## بخش اول - آشنایی با محیط مسئله

با توجه به توضیحات و لینک فراهم شده محیط را در مود Acrobot-v1 اجرا کنید.

1. اکشن‌ها، استیت‌ها و پاداشی که عامل از محیط دریافت می‌کند را شرح دهید. برای مثال توضیح دهید که استیت عامل نشان دهنده چه خصیصه‌هایی از محیط است.
2. در مورد پیوستگی و گسسته بودن استیت‌ها و اکشن‌های این محیط تحقیق کنید.

## بخش دوم - پیاده سازی الگوریتم Deep RL مبتنی بر ارزش

در این بخش هدف پیاده‌سازی الگوریتم [1] Deep Q-learning می‌باشد.

1. الگوریتم DQN را بدون استفاده از کتابخانه stable\_baselines3 پیاده‌سازی کنید. برای پیاده‌سازی ترجیحا از کتابخانه Pytorch استفاده کنید. استفاده از کتابخانه tensorflow نیز بلامانع است. عامل شما باید تسک Acrobot را از محیط Gymnasium یاد بگیرد.
2. 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.
3. از چند اپیزود تست عامل پس از یادگیری ویدیو (render) ارائه دهید. برای رندر گرفتن توصیه می‌شود جهت درگیر نشدن با مشکلات گوگل کولب، وزن‌های مدل خود را ذخیره کنید و این کار را به صورت local روی کامپیوتر خود انجام دهید.
4. پارامترهای مورد استفاده خود را در گزارش در یک جدول بیان کنید.

## بخش سوم - بهبود روش های Deep RL مبتنی بر ارزش

به عنوان نمره امتیازی، می‌توانید از دو مورد زیر یک مورد را انتخاب کرده و پیاده‌سازی کنید.

- الگوریتم DQN را با حالت Prioritized Experience Replay پیاده‌سازی کنید. توضیحات مربوط به این الگوریتم را می‌توانید در مقاله [2] مطالعه کنید. همچنین بیان کنید اضافه کردن این ویژگی چه سودی برای الگوریتم DQN بدون این ویژگی دارد.
- الگوریتم Dueling Deep Q-Networks را پیاده‌سازی کنید. توضیحات مربوط به این الگوریتم را می‌توانید در مقاله [3] پیدا کنید. همچنین بیان کنید این الگوریتم چه مزیتی بر الگوریتم DQN کلاسیک دارد.



در هر دو حالت باید مسئله بخش دوم را حل و نتایج را با نتایج بخش دوم مقایسه کنید. برای مقایسه می‌توانید از نمودارهای مناسب استفاده کنید. آیا اختلاف به صورت significant می‌باشد؟ برای ادعای خود می‌توانید از تست‌های آماری استفاده کنید.

## توجه

در صورتی که تا به حال تجربه کار با کتابخانه Pytorch برای پیاده‌سازی شبکه‌های عمیق با هدف استفاده برای یادگیری تقویتی را نداشته‌اید، می‌توانید به مجموعه ویدیوهای آموزشی زیر رجوع کنید:

1. [https://www.youtube.com/watch?v=AOypla\\_8RXg](https://www.youtube.com/watch?v=AOypla_8RXg)
2. <https://www.youtube.com/watch?v=kPa6hU9prg4>
3. <https://www.youtube.com/watch?v=ubSsUJbLkwM>
4. <https://www.youtube.com/watch?v=dsNtkT7LF8M>
5. <https://www.youtube.com/watch?v=Px3dQiv6R7E>

## مراجع

1. V. Mnih et al., "Playing Atari with deep reinforcement learning," arXiv [cs.LG], 2013
2. T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," arXiv [cs.LG], 2015.
3. Wang, Ziyu, et al. "Dueling network architectures for deep reinforcement learning." International conference on machine learning. PMLR, 2016.

## 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 جمعه 29 دی ماه خواهد بود.
- انجام این تمرین به صورت یک نفره می باشد. اما بحث و گفت و گو در دیسکورد مانعی ندارد.
- نمودارها حتماً باید title، label axis و grid داشته باشند و مقادیر به صورت گویا نمایش داده شود.
- حجم توضیحات گزارش شما هیچ گونه تاثیری در نمره نخواهد داشت و تحلیل و نمودارهای شما بیشترین ارزش را دارد.
- سعی کنید از پاسخ های روشن در گزارش خود استفاده کنید و اگر پیش فرضی در حل سوال در ذهن خود دارید، حتماً در گزارش خود آن را ذکر نمایید.
- لطفاً گزارش و کد تمرین را در فایل هایی که از طریق google Doc و google colab با شما به اشتراک گذاشته شده است، وارد نمایید.
- در صورت وجود سؤال و یا ابهام می توانید در channel مربوط به این تمرین با دانشجویان دیگر مطرح نمایید و یا برای ارتباط با دستیاران آموزشی از طریق ایجاد یک thread در همان channel دیسکورد، سؤال خود را مطرح نمایید.