تمرین هوش مصنوعی و یادگیری تقویتی
تیر ۱۴۰۲

هدف این تمرین آشنایی با الگوریتم‌های یادگیری تقویتی و کاربرد آن‌ها در حل مسایل دارای محیط و اقدام‌های گسسته است.
توضیحات شبیه‌ساز مورد استفاده در فیلم پیوست آورده شده‌است.

بخش اول) پرسش‌های مفهومی

با استفاده از دانش خود به سوالات زیر پاسخ دهید:

۱. همانطور که می‌دانید در میزان همگرایی و پایداری در روش‌های off-policy و on-policy متفاوت است. علت این تفاوت را شرح دهید و موارد استفاده‌ای که برای هر کدام مناسب است را معرفی کنید.

۲. در روش‌های model-free تابع ارزش مستقیما از تجربه‌ها بدست می‌آید. مشکلاتی که این موضوع ایجاد می‌کند را شرح دهید. برای حل این مشکلات راه‌ها پیشنهادی خود را ارائه دهید.

۳. در بعضی مسائل ممکن است پاداش یک حرکت به صورت دفعی و بلافاصله بدست نیاید. (پاداش در مراحل بعدی یا در نهایت به صورت تدریجی ظاهر شود.) مشکلات این موضوع را بررسی کنید. پیشنهادات خود را برای حل این مشکل بسط دهید.

۴. در یک روستا هر ساله در فصل برداشت محصول مسابقه‌ای برگزار می‌شود که همه می‌توانند در آن شرکت کنند و با احتمال P برندهٔ یک جایزه شوند. اما اخیرا در این روستا بیماری خطرناکی شایع شده است که منجر به مرگ یا فلج شدن می‌شود. علائم این بیماری پنهان است و کسی نمی‌داند که بیمار هست یا نه. احتمال سرایت بیماری از یک فرد بیمار به یک فرد سالم با مقدار زیر متناسب است.

$$N_S^a \times N_H, N_s : number\ of\ sick\ people, N_H : number\ of\ healthy\ people, a > 1$$

- برای این مسئله یک مدل MDP طراحی کنید, action space, states, states transitions, rewards را مشخص کنید.
- برای حل این مسئله چه الگوریتمی‌هایی پیشنهاد می‌کنید.
- الگوریتم value-iteration را برای این مسئله با فرض جمعیت ۱۰ نفره روستا به همراه فرض‌های دلخواه خود تا ۳ مرحله پیش ببرید.

بخش کامپیوتری

برنامه‌ریزی حرکت یک تاکسی خودران

در این بخش می‌خواهیم با استفاده از چند الگوریتم یادگیری تقویتی، برنامه‌ریزی حرکت یک تاکسی خودران را به گونه‌ای انجام دهیم که با کمترین تعداد حرکت بتواند تعداد مشخصی مسافر را از مبداهای مشخص سوار کند و به مقصدهای مشخص برساند. برای این کار از محیط شبیه‌سازی gym-taxi-v3 استفاده می‌کنیم. برای نصب این کتابخانه می‌توانید از روش گفته شده در اینجا استفاده کنید. ویدیوی آموزشی پیوست نیز در مورد آشنایی با این محیط شبیه‌سازی به شما کمک می‌کند. برای سادگی می‌توانید از کد موجود در فایل تمرین استفاده کنید و الگوریتم‌های زیر را به آن اضافه کنید.

در این بخش می‌توانید از کد موجود در فایل تمرین استفاده کنید و الگوریتم‌های زیر را اضافه کنید.

- الگوریتم Q-Learning
- یکی از روشهای on-policy یا off-policy الگوریتم Monte Carlo

برای شروع مسأله بعد از نصب محیط و ساخت محیط حتماً در زمان reset مقدار seed را برابر ۳ رقم آخر شماره دانشجویی خود قرار دهید:

$$
\begin{gathered}
import\ gym \\
env = gym.make('Taxi-v3') \\
env.seed(seed = 111) \\
init_{state} = env.reset()
\end{gathered}
$$

برای هر الگوریتم با ۲ ضریب کاهنده متفاوت و ۲ حالت برای نرخ یادگیری (کاهنده و ثابت) در هر مورد ۱۰ بار و هر بار حداقل ۲۰۰۰ اپیزود مسئله را تکرار کنید و پاداش متوسط را محاسبه کنید

الف) محیط شبیه‌ساز را بررسی کنید.(Action space, Observation space)

ب) تعدادی از state ها غیرقابل دسترس هستند. آن‌ها را پیدا کنید و علت این موضوع را بنویسید.

پ) پایداری و همگرایی را برای هرکدام از الگوریتم‌ها بررسی کنید. (تعداد اپیزود موردنیاز برای همگرایی, پاداش متوسط) هایپرپارامترها را برای بهترشدن نتیجه تنظیم کنید

ج)(امتیازی) از نتیجه نهایی render بگیرید. (فیلم را درون فایل نهایی قرار دهید.)

ه)(امتیازی) برای حل مسأله از الگوریتم Deep Q-Learning استفاده کنید.

چند توضیح:

- برای یادگیری مفاهیمی که در تمرین مطرح شده و احتمالا تدریس نشده‌اند از منابع موجود در اینترنت استفاده کنید.
- برای انجام بخش‌های مختلف تمرین می‌توانید از کتابخانه‌های آماده‌ای مانند numpy، matplotlib، pandas، sklearn و seaborn استفاده کنید.
- **تحویل گزارش این تمرین ضروری است و به تمرین بدون گزارش نمره‌ای تعلق نمی‌گیرد**. حجم گزارش معیاری برای ارزیابی نخواهد بود و لزومی به توضیح جزئیات کد نیست؛ اما از آنجا که برای این تمرین از کتابخانه‌های موجود استفاده می‌کنید لطفا تمامی پارامترهای تنظیم‌شده در هر قسمت از کد را گزارش کرده و فرض‌هایی را که برای پیاده‌سازی‌ها و محاسبات خود به کار برده‌اید ذکر کنید. از ارائهٔ توضیحات کلیشه‌ای و همانند بَرداری از منابع موجود بپرهیزید.
- **در فرایند ارزیابی گزارش، کدهای شما لزوما اجرا نخواهد شد. بنابراین همهٔ نتایج و تحلیل‌های خود را به‌طور کامل ارائه کنید.**
- **شباهت بیش از حد گزارش و کدها باعث از دست دادن نمرهٔ تمرین خواهد شد. همچنین گزارش‌هایی که در آنها از کدهای آماده استفاده شده باشد پذیرفته نخواهند شد.**
- گزارش شما باید به صورت تایپ شده و با فرمت pdf ارائه شود و کدهایی که به همراه گزارش تحویل می‌دهید باید قابل اجرا باشند. در انتها تمامی فایل‌های لازم را در یک فایل zip یا rar بارگذاری و ارسال کنید.
- **نمره هایی این تمرین به ارائه آن وابسته می‌باشد که زمان‌بندی آن متعابقا اعلام می‌شود.**
- پرسش‌های خود را از طریق ایمیل یا تلگرام از دستیار آموزشی مربوطه بپرسید:

	ایمیل	تلگرام
علی اسکندریان	alieskandarian@ut.ac.ir	alieskandarian79@