به نام خدا

تمرین چهارم

درس یادگیری تعاملی

پاییز ۹۹

معین کافی mkafi.anaraki@gmail.com

فاطمه نورزاد ati.noorzad@gmail.com

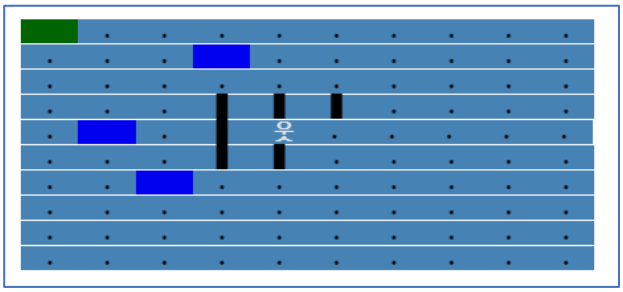
در این تمرین به بررسی الگوریتم های یادگیری تعاملی در محیط بازی پیاده سازی شده میپردازیم. در این محیط هدف این است که آدمک که وسط رودخانه است به جزیره ک با رنگ سبز مشخص شده برسد. برای رسین به این مقصد، او دو مانع در برابر خود میبیند.

اولین مانع که با رنگ سیاه نمایش داده شده است، سخره های بلندی هستند که او توانایی عبور از آن ها را ندارد. در صورتی که به این سخره ها برخورد کند، در جای خود باقی مانده و پاداشی به اندازه ۳ – میگیرد.

دومین این موانع، قسمت هایی هستند که با رنگ نیلی مشخص شده اند. آدمک در صورتی که به این نقاط برسد، برای ادامه دادن مسیر دچار مشکل میشود. چراکه این قسمت ها عمق زیادی دارند و او هم شناگر ماهری نیست. به همین علت اگر تلاش کند از این نقاط در هر جهتی خارج شود، پاداشی برابر با ۱۰- میگیرد.

عبور از سایر قسمت ها پاداشی ندارد. (چراکه وی را به هدف نزدیک کرده و از طرفی باعث خستگی او میشوند. این دوگانگی باعث صفر بودن پاداش خواهد شد.)

رسیدن به جزیره به تنهایی پاداشی ندارد. تنها در صورتی که بتواند در آن حرکت کند، پاداش ۱ را دریافت می نماید. (دقت کنید که این جمله به این معنی است که در صورت انتخاب هر عمل دلخواه پس از رسیدن به جزیره، پاداش داده میشود و تنها رسیدن به جزیره پاداش صفر دارد.)



شکل ۱. محیط بازی

برای استفاده از محیط پیاده سازی شده، کافی ست فایل قرار داده شده در Elearn را به کد های خود اضافه نمایید. به علاوه قسمتی به عنوان راهنما هم قرار داده شده که نشان میدهد چگونه میتوانید به فضای حالت و عمل دسترسی بیابید.

برای محیط بیان شده، هر یک از الگوریتم های زیر را پیاده سازی کنید:

- On-policy Monte-Carlo
- Off-policy Monte-Carlo
- Double Q-learning
- Tree Back-up
- SARSA
- Two-Step Expected SARSA

با استفاده از پیاده سازی بالا به سوالات زیر پاسخ دهید. دقت کنید که برای پاسخ خود دلیل کافی و شفاف ارائه کنید.

۱) برای مدل بیان شده، الگوریتم ها را از منظر پاداش برحسب تعداد episode ها مقایسه نمایید.
۲) از نظر سرعت یادگیری، کدامیک عملکرد بهتری دارد؟

قسمت امتیازی: آیا میتوانید با تغییر دادن پاداش ها به مقادیر معقول به همگرایی سریعتری برسید؟ منطق خود را برای این تغییرات ذکر کرده و آن را شبیه سازی کنید.

لطفا به نکات زیر توجه کنید:

- برای کد هایی که ضمیمه میکنید، حتما گزارش بنویسید. کد های ضمیمه شده بدون گزارش نمره ای نخواهند داشت. (این گزارش ها معیار تفاوت کد شما با مدل های مشابه موجود در اینترنت است.)
- حجم گزارش به هیچ عنوان معیار نمره دهی نیست. بنابراین به حد نیاز توضیح دهید.
- از پاسخ های روشن در گزارش خود استفاده کنید و تمام فرضیات خود را به طور شفاف بیان کنید.
- نمودار ها باید واضح بوده و بر هر محور برچسب مناسب داشته باشند. به علاوه دقت کنید که نمودار مربوط به هر الگوریتم را با رنگ متفاوت نمایش دهید.
- مشورت و هم فکری برای سوالات مانعی ندارد، اما در صورت مشابهت بین فایل های تحویل داده شده، مطابق قوانین درس برخورد میشود.

سلامت و موفق باشید :)