دانشگاه صنعتی شریف
دانشکده‌ی مهندسی کامپیوتر

هوش مصنوعی

پاییز ۱۴۰۰

استاد: محمدحسین رهبان
گردآورندگان: محمد محدی، حمیدرضا کامکاری
بررسی و بازبینی: محمدرضا یزدانی‌فر

تمرین هفتم سری اول — Markov Decision Processes — مهلت ارسال: ۹ دی

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص‌شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمارین تا سقف سه روز و در مجموع ۲۵ روز، وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال‌شده پذیرفته نخواهند بود. همچنین، به ازای هر روز تأخیر غیر مجاز ۱۰ درصد از نمره تمرین به صورت ساعتی کسر خواهد شد.
- هم‌کاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

سوالات (۱۰۰ نمره)

۱. (۱۰۰ نمره) فرض کنید دارید بازی packman در یک صفحه 2×2 به صورت زیر انجام می‌دهید و تنها یک روح در صفحه قرار دارد. هر سری در هر خانه که هستید می‌توانید یکی از دکمه‌های بالا_پایین_چپ_راست را بزنید و کامپیوتر به احتمال 90 درصد شما را در آن جهت هدایت می‌کند(البته اگر دیواری در آن جهت وجود داشته‌باشد سر جای خود می‌ماند). همچنین به احتمال 10 ممکن است کیبورد کار نکند و سر جای خود باقی بمانید.

روح نیز در هر مرحله سعی می‌کند خودش را به شما نزدیک کند. اگر چند انتخاب برای نزدیک کردن داشت یکی را با احتمال مساوی انتخاب می‌کند.

توجه کنید هر مرحله از بازی که زنده بمانید $+1$ امتیاز دریافت می‌کنید.

مسئله را با استفاده از MDP مدل‌سازی کنید و از حداکثر ۳ تا state استفاده کنید. روی شکل به نحو خوبی State و Action به همراه Transition Probability ها را نشان دهید. توجه کنید ممکن است بسیاری از حالات متقارن باشند.

راهنمایی: برای حالات متقارن کافیست یکی از آنان به همراه یال‌های خروجی و احتمال‌هایش را نشان بدهید.

فرض کنید که یک بازی packman با ابعاد بالا را با یک MDP مانند $M = (S, A, R, P, \gamma)$ مدل کرده‌ایم و γ عددی مثبت و کوچکتر از یک است. می‌گوییم در این بازی سیاست π_1 از سیاست π_2 بهتر است $(\pi_1 \geq \pi_2)$ اگر داشته باشیم:

$$\forall s \in S : V^{\pi_1}(s) \geq V^{\pi_2}(s)$$

که V^{π} مقدار value function است به شرط آن که بازیکن، سیاست π را اعمال کند.

(آ) ثابت کنید که در M، حداقل یک سیاست وجود دارد که از هر سیاست دیگری بزرگ‌تر باشد.

(ب) ثابت کنید که در M، حداقل یک سیاست deterministic وجود دارد که از هر سیاست دیگری بزرگ‌تر باشد.

توضیح: سیاست deterministic ، سیاستی است که بازیکن در یک state ،یک action مشخص را با احتمال ۱ انجام دهد. در مقابل این سیاست‌ها، سیاست‌های stochastic قرار دارند.