دانشگاه صنعتی شریف
دانشکده‌ی مهندسی کامپیوتر

هوش مصنوعی

بهار ۱۴۰۰

استاد: محمدحسین رهبان
گردآورندگان: نیما فتحی ـ سروش وفایی‌تبار

تمرین هفتم بخش اول — Markov Decision Process — مهلت ارسال: ۱۱ خرداد

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص‌شده است.
- 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 استفاده‌برای حل سوال مورد نظر را ذکرکنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

سوالات نظری (۱۰۰ نمره)

۱. (۵۰ نمره) همانطور که می‌دانید در الگوریتم value iteration پس از هر تکرار شما می‌توانید بهترین سیاست (policy) در آن لحظه را پیدا کنید.

حال اگر در value iteration بهترین سیاست در مرحله iام برابر با بهترین سیاست در مرحله 1+iام باشد آیا ممکن است که در ادامه سیاست ما تغییر کند؟

در صورت درستی عبارت بالا اثبات کنید در غیر این صورت مثال نقضی بیاورید.

۲. (۵۰ نمره) یک MDP با سه حالت و دو کنش (L, R) در نظر بگیرید که احتمال انتقالات به صورت زیر است:

Action L:

	State 1	State 2	State 3
In state 1:	0	1/4	3/4
In state 2:	3/4	0	1/4
In state 3:	1/4	3/4	0

Action R:

	State 1	State 2	State 3
In state 1:	0	3/4	1/4
In state 2:	1/4	0	3/4
In state 3:	3/4	1/4	0

امتیاز حالت ۲ برابر ۱ و بقیه‌ی حالت‌ها ۰ است. ($\gamma = ۰/۵$)

(آ) $V(s)$ تخمینی را به کمک روش Value Iteration برای دو مرحله محاسبه نمایید. (حالات اولیه را صفر در نظر بگیرید)

(ب) با در نظر گرفتن $V_۱ = V_۳ = ۰/۵$ و $V_۲ = ۰/۲۵$ بهترین کنش برای حالت اول را محاسبه کنید و توضیح دهید.