

به نام خدا



درس هوش مصنوعی و سیستم‌های خبره

---

## تمرین دوازدهم

---

مدرس درس:  
جناب آقای دکتر محمدی

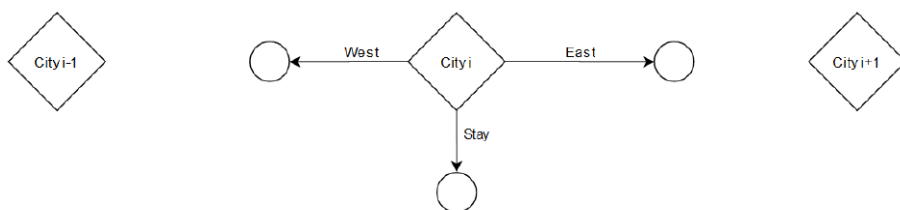
طراحان:  
سهیل حمزه بیگی  
حامد فیض آبادی

مهلت ارسال: ۱۴۰۱/۱۰/۱۴

## سوال ۱ - الگوریتم Q-Learning

جاده ای را تصور کنید که در طول آن  $N$  شهر وجود دارد که از ۱ تا  $N$  شماره گذاری شده اند. شما یک تاجر در شهر ۱ هستید. در هر روز می‌توانید به شهر همسایه سفر (شرق یا غرب) سفر کنید و یا در شهر کنونی خود بمانید و به تجارت بپردازید. اگر تصمیم بگیرید که از شهر  $i$  بروید با احتمال  $P_i$  با موفقیت به مقصد می‌رسید؛ اما، با احتمال  $1 - P_i$  هوا طوفانی خواهد بود و قادر به سفر کردن نخواهید بود. بدیهی است که در این صورت وقت آن روز هدر رفته است و مجبور هستید در شهر کنونی خود بمانید. اگر در شهر بمانید و تجارت انجام دهید پاداش  $r_i > 0$  دریافت خواهید کرد. اگر در حال سفر و تغییر شهر باشید یا هوا طوفانی شود و روزتان هدر رود هیچ پاداشی دریافت نمی‌کنید. ( $r_i = 0$ ) :

الف) گراف صورت مسئله را تکمیل کنید. Action ها را با فلش پررنگ و Transition ها را با فلش نقطه چین مشخص باشد. بر روی گراف احتمال ها و پاداش ها را نیز مشخص کنید.



ب) اگر سیاست ما همیشه انتخاب Stay باشد و برای همه  $i$  ها داشته باشیم  $P_i = 1, r_i = 1$  و مقدار پارامتر  $\gamma = 0.5$  باشد آنگاه مقدار  $V^{stay}(1)$  را به دست آورید.

پ) اگر برای همه  $i$  ها داشته باشیم  $P_i = 1, r_i = 1$  و پارامتر  $\gamma = 0.5$  باشد آنگاه مقدار آنگاه مقدار  $V^*(1)$  را به دست آورید.

ت) اگر بدانیم برای همه  $i$  ها  $P_i > 0, r_i > 0$  است و مقدار پارامتر  $\gamma = 1$  باشد آنگاه سیاست بهینه را تعریف کنید.

ث) فرض کنید موارد زیر را تجربه کرده ایم.

1 - ( $s = 1, a = stay, r = 4$ )

2 - ( $s = 1, a = east, r = 0$ )

3 - ( $s = 2, a = stay, r = 6$ )

4 - ( $s = 2, a = west, r = 0$ )

$$5 - (s = 1, a = \text{stay}, r = 4)$$

اگر  $LearningRate = 0.5$  باشد و مقدار پارامتر  $\gamma = 0.5$  باشد آنگاه جدول زیر را تکمیل کنید:

$(s, a, r, s')$	$Q(1, \text{Stay})$	$Q(1, \text{East})$	$Q(2, \text{West})$	$Q(2, \text{Stay})$
initial	0	0	0	0
(1, Stay, 4, 1)				
(1, East, 0, 2)				
(2, Stay, 6, 2)				
(2, West, 0, 1)				
(1, Stay, 4, 1)				

## ۱ سوال ۲- پیاده سازی Q-learning

در این سوال کد آموزش یک agent را که بر اساس آزمایش و خطا با محیط اطراف آشنا میشود را تکمیل می‌کنید. ابتدا فایل اولیه زیپ را از این [لینک](#) دانلود کنید. شما باید فایل qlearningAgents.py را تغییر بدهید و موارد خواسته شده را پیاده سازی کنید.

تذکر: کدهای مشابه زیادی برای این سوال در اینترنت موجود است که طبیعتاً سوزاندن فرصت یادگیری برای شما است، بنابراین سعی کنید با مطالبی که یاد گرفته‌اید این سوال را حل کنید (هر گونه کپی، پیگرد قانونی دارد!).

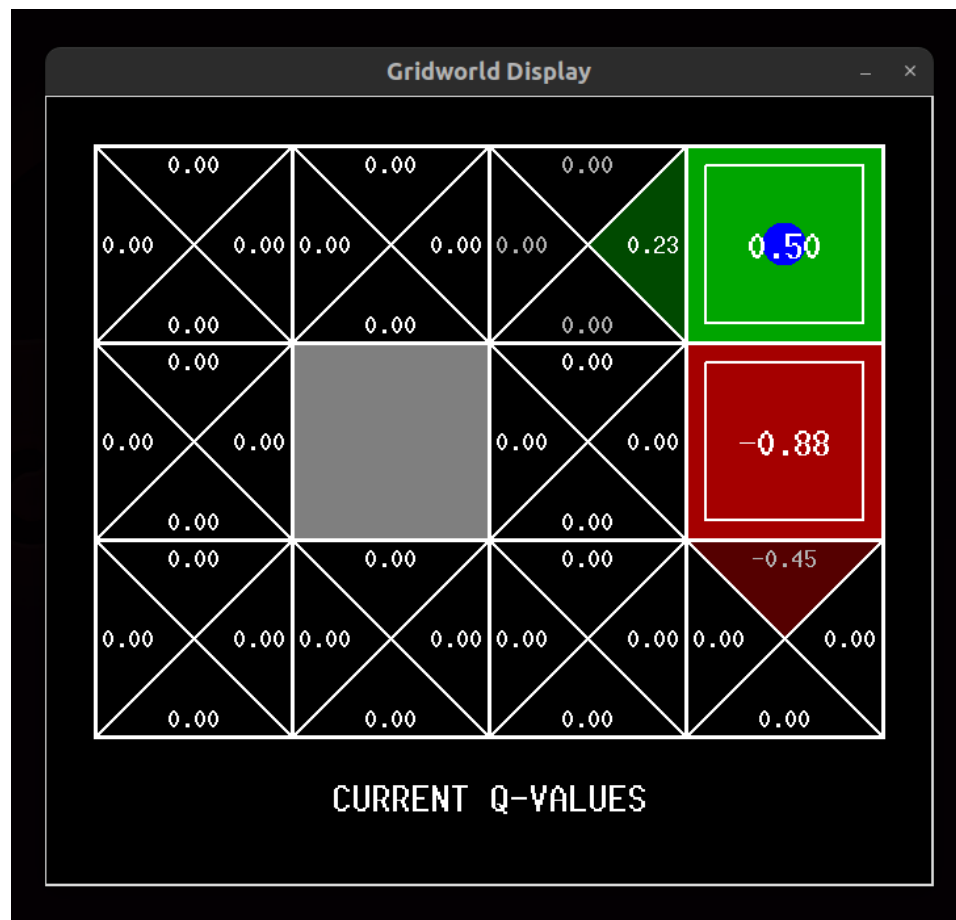
### ۱.۱ مراحل پیاده سازی

- برای اجرای سوال شما نیاز به یک محیط پایتون 3.6 دارید که پیشنهاد میشود با conda و دستور `conda create -name ai-iust python=3.6` این کار را انجام دهید. در صورت مشکل در راه اندازی میتوانید طبق این [لینک](#) عمل کنید یا از TA مربوطه کمک بگیرید.
- متدهای `getQValue` و `computeValueFromQValues` و `computeActionFromQValues` و `update` باید پیاده سازی بشوند.
- به متد `init` این خط را اضافه کنید: `self.QValues = util.Counter()`  
کلاس `counter` یک نوع دیکشنری `extend` شده است که بعضی از متدها را به صورت آماده دارد و کار پیاده سازی را ساده‌تر میکند (به ساختار کلاس مراجعه کنید و این متدها را مشاهده کنید). کلاس `counter` به این شکل است که اگر بخواهید به یک کلید از دیکشنری دسترسی پیدا کنید و وجود نداشته باشد به جای اینکه خطای `KeyError` بدهد، مقدار صفر را به شما برمیگرداند بنابراین یک ساختمان داده مناسب برای ذخیره سازی مقادیر `QValue` ها است.
- به عنوان اولین متد `getQValue` را پیاده سازی کند و هر جایی از کد که مقدار `QValue` را نیاز داشتید از این تابع استفاده کنید، این کار برای بخش‌های بعد لازم میباشد و باعث میشود کد شما کلی‌تر باشد (این مورد برای مقدار `state_value` ها نیز رعایت شود و با تابع `computeValueFromQValues` به مقدار آن‌ها دسترسی پیدا کنید).
- تابع `update` را با توجه به فرمول اسلاید ۲۵ از جلسه ۲۳ کامل کنید. پارامترهای  $\alpha$  و  $\gamma$  جزئی از attribute های کلاس هستند و میتوانید از آن‌ها استفاده کنید.
- تابع `getAction` برای پیاده سازی مفهوم `Exploration` و `Exploitation` است که در اسلاید ۳۴ در جلسه ۲۳ در مورد آن صحبت شده است. این تابع را باید جوری کامل کنید که به احتمال  $\epsilon$  رندوم و احتمال  $1 - \epsilon$  بر اساس `policy` عمل کند. میتوانید از `flipCoin` در `utils` استفاده کنید.
- برای پیاده سازی دو تابع باقی مانده میتوانید `action` های مجاز را با استفاده از `self.getLegalActions(state)` بدست آورید.  
در تابع `computeValueFromQValues` باید مقدار `state_value` را از `QValue` ها طبق

فرمول‌هایی که خوانده‌اید، محاسبه کنید و در تابع `computeActionFromQValues` باید بهترین action را بر اساس مقادیر QValue ها محاسبه کنید.

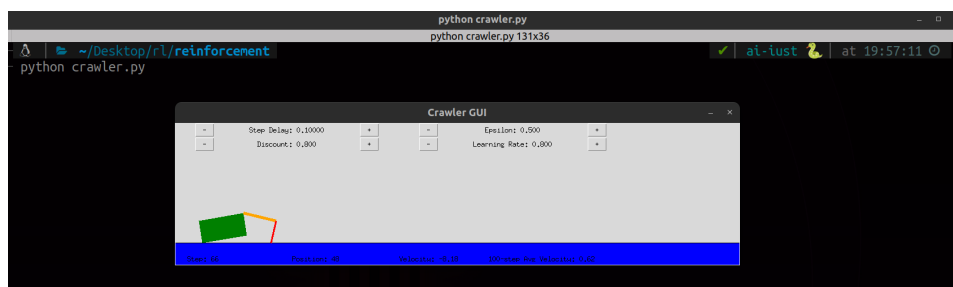
## ۲.۱ ارزیابی پیاده سازی

برای تست کد خود ابتدا دستور `python gridworld.py -a q -k 5 -m` اجرا کنید. با این دستور `gridworld` به شما نمایش داده میشود و با توجه به مقدار `k` که ۵ است، ۵ episode میتواند agent را با دکمه‌های جهت کیبرد کنترل کنید و نتایج آموزش را بر روی صفحه ببینید. به چه علت است که همواره agent مطابق جهتی که شما میدهید حرکت نمیکند؟ آپشن `-m` را از دستور بالا بردارید که از حالت `manual` دربیاید و لازم به کنترل شما نباشد. تعداد iteration ها یا `k` را از ۵ به ۲۰ تغییر بدهید و تصویر نهایی همراه با نتیجه‌ای که در لاگ ترمینال نمایش داده میشود را به همراه برداشت خود در گزارش ذکر کنید.



بدون هیچ تغییری، کد `crawler` را با دستور `python crawler.py` اجرا کنید. این کد یک

ربات است که دارای یک بازو است که از دو جا قابلیت خم شدن دارد و با کد Qlearning ای که شما نوشته‌اید بازوی ربات را حرکت می‌دهد و آموزش می‌بیند و هدف این است که به سمت راست حرکت کند (امتیاز مثبت در این جهت است و جهت خلاف آن امتیاز منفی دارد). پارامترهای موجود در تصویر را توضیح دهید و بگویید که تغییر هر کدام چه تاثیری بر روی ربات دارد.



اگر کد مربوط به ربات شما کار نکرد، این احتمال وجود دارد که کدی که در بخش qllearning نوشته‌اید خیلی کلی نیست و مربوط به یک مسئله خیلی خاص مانند gridworld است. سعی کنید مطابق با توضیحات پیاده سازی، کد خود را کلی تر و general تر بنویسید.

### قوانین:

۱. تمرین ها به صورت فردی انجام شوند و حل گروهی تمرین ها مجاز نیست.
۲. نمره شما بر اساس گزارش راه طی شده برای حل مسئله و پاسخ صحیح خواهد بود لذا از هرگونه اطناب در گزارش پرهیز و به موارد خواسته شده به صورت کامل پاسخ دهید.
۳. برای تحویل تمرین یک فایل zip شامل گزارش حل سوالات، با نام [HW12\_ID\_NAME] در سامانه LMS بارگذاری کنید.