تمرین سری سوم درس یادگیری تعاملی

پاییز ۱۴۰۰

بخش اول مدلسازی MDP

برای مسائل زیر مدلی برای حل ارائه دهید. لازم به ذکر است برای مدل نیاز دارید استیت ها، اکشن ها انتقال بین استیت ها به صورت احتمالی و پاداش را تعیین کنید. (توجه کنید که جواب یکسان برای مسائل وجود ندارد و همچنین به پاسخ های خلاقانه نمره امتیازی تعلق خواهد گرفت)

۱. در یک دهکده قصد داریم در ابتدای هر ماه تصمیم بگیریم که فروش خرچنگ مجاز است یا خیر. هر سری که تصمیم به فروش خرچنگ گرفته شود مقداری از خرچنگ ها کاسته شده و مقداری سود از فروش خرچنگ ها به دست ما خواهد رسید. لازم به ذکر است که اگر جمعیت خرچنگ ها از حدی کمتر شوند نیازمند هزینه زیادی برای جبران جمعیت آن ها هستیم وگرنه کلا صنعت فروش خرچنگ در این شهر ورشکسته میشود.

۲. در خرید سهام در آخر هر دوره باید تصمیم بگیریم که سهام خود را نگه داریم و یا آن را بفروشیم. فرض کنید خرید و یا فروش سهام در هر دوره به طور ۷ بار انجام میشود و برای سادگی reset صورت میگیرد. همچنین در انتهای هر تصمیم با احتمالی به سود فرد اضافه شده و با احتمالی درصد از ارزش سهام خود را از دست می‌دهد.

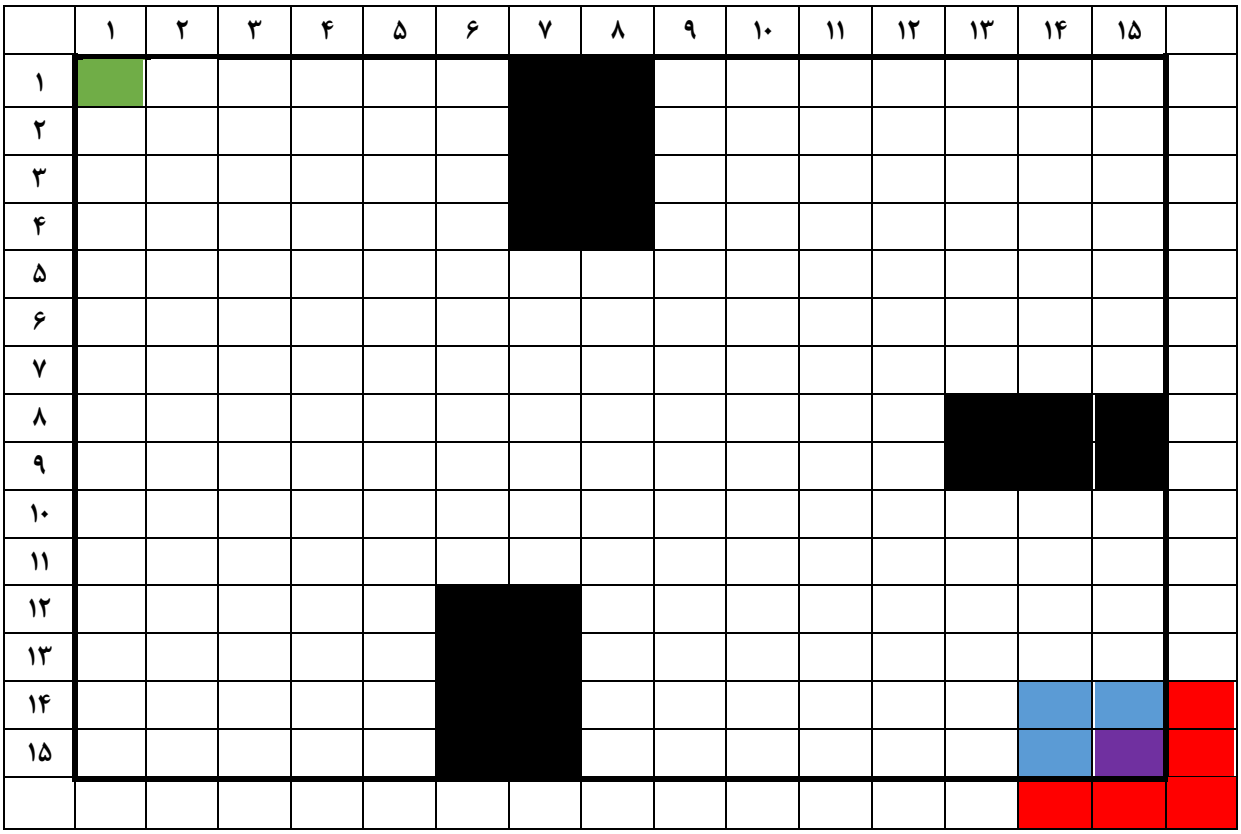
۳. در یک کارخانه قصد داریم با توجه به تقاضای سال گذشته میزان تولید سال آینده را تعیین کنیم. توجه کنید که هر مشتری که درخواست کالا داشته باشد و کالا تولید کافی نداشته باشد علاوه بر ضرر آن مشتری را کامل از دست میدهیم که برای ما ضرری اضافه تر نیز دارد. همچنین کالاهایی که تولید شدند اما خریداری نشد نیز ضرر دارند.

۴. در یک آتش نشانی برای هر آلارم آتش که از هر منطقه (منطقه های مسکونی، اداری، تجاری، زراعی و ...) می بایست تصمیم بگیریم از تعداد ماشین های آتش نشانی موجود چند ماشین ارسال شوند. با توجه به نوع آن منطقه و بلندی و کوتاهی آژیر تعداد ماشین نیازمند به ارسال را تعیین کنید. در نظر داشته باشید که ارزش انسان ها از ساختمان ها بیشتر است و در تعیین ارزش منطقه ها باید پیش فرض جمعیتی آن منطقه را بیان کنید. همچنین اگر آژیر زده شود و ماشینی در آتش نشانی نباشد ضرر زیادی خورده و برای مناطق پر جمعیت میبایست تعدادی بیشتر ماشین برای نجات انسان ها ارسال شوند.

بخش دوم پیاده سازی

مسئله مسیریابی برای رباتها از مسائل مهم و اساسی برای رباتهای امروزی میباشد. مواردی مانند عدم برخورد با موانع موجود و پیدا کردن بهترین مسیر در محیط داده شده از این قبیل مسائل میباشند. این مسائل را میتوان در ترکیب با مسائل MDP حل کرد. محیط زیر را طبق توضیحات داده شده در نظر گرفته، و پس از پیاده سازی موارد خواسته شده، به سوالات هر بخش پاسخ داده و تحلیل خود را ارائه کنید.

یک محیط grid طبق شکل۱ با ابعاد ۱۵ در ۱۵ را در نظر بگیرید. ربات ما در ابتدا در نقطه (۱۵،۱۵) قرار گرفته است. هدف ربات این است که به خانه (۱،۱) برود. ربات مورد نظر ما در هر استیت قادر به انجام ۹ عمل مختلف میباشد. ۸ عمل برای جابجا در جهت های ۸ گانه (حرکت های مورب مجاز است) و یک عمل برای باقی ماندن در نقطه فعلی. برای ربات دو سری مجموعه استیت داریم. مجموعه استیت‌های قابل دسترسی و مجموعه استیت‌های ممکن. همسایه‌ی ممکن همسایه‌ای است که خارج از محدوده محیط نباشد و مانع نباشد. همسایه ی در دسترس همسایه ای است که با یکی از اکشن های ممکنه بتوان به آن رسید. برای مثال در موقعیت ابتدایی ربات نقاط قرمز نشان داده شده نقاطی غیر قابل دسترسی هستند. و نقاط آبی نشان داده شده نقاط قابل دسترس می‌باشند. همچنین تمامی استیت های موجود در محیط که مانع نیستند نیز نقاط ممکن می‌باشند.



شکل ۱ – ربات در ابتدا در خانه بنفش – هدف رسیدن به خانه سبز – نقاط مشکلی مانع

در انجام هر اکشن ربات با احتمال p به جهت انتخابی میرود و در غیر این صورت به یکی از همسایگان "ممکن و در دسترس" لیز میخورد. دقت شود که احتمال لیز خوردن به همه خانه‌ها به صورت یکنواخت و یکسان است. برای انجام هر حرکت بخاطر وجود انرژی مصرف شده و زمان تلف شده پاداش منفی‌ای درنظر گرفته شده است. همچنین هنگام برخورد با مانع هزینه‌ی برخورد با مانع نیز برای ربات در نظر گرفته شده است. همچنین در هنگام رسیدن به هدف ربات پاداش دریافت میکند.

حالت های زیر حالت‌های ممکن در محیط هستند:

- حالت پایه:
  - احتمال انجام اکشن و رفتن به استیت بعدی برابر ۰,۸
  - احتمال لیز خوردن ربات و رفتن به یک خانه "ممکن و در دسترس" یا ماندن در استیت فعلی برابر ۰,۲
  - هزینه برخورد با مانع ۱-
  - برای انجام هر حرکت به خاطر از دست وقت و انرژی پاداش منفی برابر با ۰,۰۱ -
  - پاداش رسیدن به خانه هدف برابر با ۱۰۰۰
- حالت حرکت بدون هزینه : همانند حالت پایه می باشد با این فرق که هزینه حرکت برابر با ۰ در نظر گرفته شود. و در صورت برخورد با مانع پاداش منفی ۰,۰۱ برای ربات در نظر گرفته شود.
- حالت حرکت با هزینه زیاد: همانند حالت پایه میباشد با این تفاوت که هزینه هر حرکت برابر با ۱- (برای همه اکشن‌ها به جز اکشن ماندن در خانه)، هزینه برخورد با مانع برابر با ۱۰- و پاداش رسیدن به خانه هدف را برابر با ۱۰۰ در نظر بگیرید.

با توجه به MDP تعریف شده، توابع مشخص شده در فایل نوتبوک پیوست شده به همراه تمرین را کامل نموده تا مراحل زیر را پیاده سازی کرده و به سوالات مربوطه پاسخ دهید. لازم به ذکر است که اگر فرمت ارائه شده در توابع را رعایت ننمایید نصف نمره از شما کاسته خواهد شد.

۱- شبه کدی برای روش مونت کارلو بنویسید. هدف از روش مونت کارلو به دست اوردن ارزش استیت‌های محیط می‌باشد.

۲- سیاست بهینه را برای حالت پایه با استفاده از روش policy iteration به دست آورید. مقدار discount factor برابر با ۰,۹ در نظر گرفته شود.

۳- سیاست بهینه را با روش policy iteration را برای حالت بدون اصطکاک به دست آورده و با نتایج مرحله دوم مقایسه کنید. در مقایسه طول مسیر طی شده توسط ربات را در نظر داشته باشید. مقدار discount factor برابر با ۰,۹ در نظر گرفته شود.

۴- حال حالت با اصطکاک زیاد را در نظر گرفته و سیاست بهینه را با استفاده از روش policy iteration به دست آورده و با دو حالت قبل مقایسه کنید. مقدار discount factor برابر با ۰,۹ در نظر گرفته شود.

۵- با توجه به مراحل ۲ و ۳ بهترین حالت برای ریوارد محیط را در نظر گرفته و نقش تفاوت مقدارهای مختلف برای discount factor را برای ۴ مقدار مختلف در مسئله بررسی کنید. تحلیل خود از

نتایج به دست امده و همچنین آینده نگری ربات با توجه به discount factor تعیین شده را بررسی کنید.

۶- الگوریتم value iteration را برای محیط داده شده اجرا کرده و نتایج به دست آمده را با بهترین نتیجه قسمت ۵ مقایسه کنید.

۷- (امتیازی) دلیل تفاوت بخش ۲و ۳و ۴ را بررسی کنید و راه حلی برای آن ارائه دهید.

نکات تکمیلی:

- سعی کنید از پاسخ های روشن در گزارش خود استفاده کنید و اگر پیش فرضی در حل سوال در ذهن خود دارید، حتما در گزارش خود آن را ذکر نمایید.
- حجم گزارش شما به هیچ وجه معیار نمره دهی نیست، پس لطفا در حد نیاز توضیح دهید.
- از نمودارهای واضح در گزارش خود استفاده کنید، نمودارهایتان حتما دارای لیبل واضح روی هر محور و توضیح مناسب باشد.
- لطفا در گزارش و کدهای خود از تمرین دیگران استفاده نکنید. مشورت و همفکری در مورد سوال ها اشکالی ندارد اما اگر شباهت بیش از اندازه در تمرین ها دیده شود منجر به صفر شدن نمره خواهد شد.
- تمام فایل ها را در قالب یک فایل zip در سایت درس بارگذاری کنید.
- حتما فرمت گزارش که در سایت درس قرار داده شده است را رعایت نمایید.
- در صورت وجود هر نوع سوال در رابطه با این سری تمرین میتونید از طریق ایمیل‌های اعلام شده با دستیاران آموزشی درارتباط باشید.

بنفشه کریمیان – banafshehkarimian@ut.ac.ir
امیرحسین مصباح – amir.mesbah@ut.ac.ir

شاد و سلامت باشید :)