**سید علی موسوی رضا فرخ**

**سوال ۱**

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom\_map\_1:

برای مقدار discount\_factor برابر 1 مسئله در یک لوپ میوفتد، زیرا مقدار دلتا کاهش پیدا نمیکند و همواره بزرگتر از تتا است، پس شرط break اتفاق نمیوفتد.

برای بقیه مقادیر discount\_factor نتیجه یکسان است، زیرا محیط به گونه ای است که، مقدار discount\_factor تاثیر خاصی نمیگذارد. (به دلیل سادگی محیط)

**سوال ۲**

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom\_map\_2:

برای مقدار discount\_factor برابر 1 مسئله در یک لوپ میوفتد، زیرا مقدار دلتا کاهش پیدا نمیکند و همواره بزرگتر از تتا است، پس شرط break اتفاق نمیوفتد.

برای مقادیر 0.5 و 0.1 در یک خانه ثابت می‌ماند، زیرا در این مقادیر ارزش پیشنهادی انقدر کم است که بی حرکت ماندن را بهتر از حرکت به سمت هدف میبیند، ولی در مقدار 0.9 چون هنوز ارزش حرکت رو به آینده بیشتر است، به سمت هدف حرکت میکند.

**سوال ۳**

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom\_map\_3:



با فایل بالا، زمانی که True است به معنای غیر قطعی بودن حرکات است.(زمین سر است.) و با بودن این شرایط، تابع ارزش حالت داری حرکات رندوم بیشتری است و رفته رفته V[state] ما به حالت stable تری میرسد. و در نقاط hole ارزش تابع منفی است. ولی در False حرکت به سمت reward به صورت قطعی انجام میشود.

**سوال ۴**

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom\_map\_4:



با فایل بالا، زمانی که True است به معنای غیر قطعی بودن حرکات است.(زمین سر است.) و با بودن این شرایط، تابع ارزش حالت داری حرکات رندوم بیشتری است و رفته رفته V[state] ما به حالت stable تری میرسد. و در نقاط hole ارزش تابع منفی است. ولی در False حرکت به سمت reward به صورت قطعی انجام میشود.

**سوال ۵ و ۶**

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom\_map\_5:

با توجه به داده ها ارزش تابع حالت، ممکن است تغییر کند اما از نظر نسبی تغییر نمیکند و یک مسیر معین به سمت هدف تعیین میشود.

سوال ۷