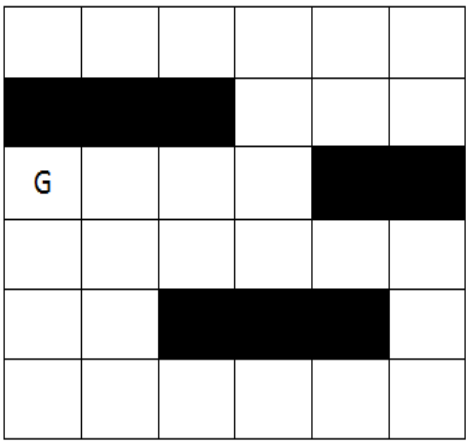
به نام آنکه جان را فکرت آموخت

تکلیف سری سوم – یادگیری ماشین　　دانشگاه خوارزمی　　پاییز ۱۳۹۶

موعد تحویل: ۱۳۹۶/۱۰/۱۵

مقایسه الگوریتم‌های یادگیری Q و SARSA

۱- محیطی گسسته را به ابعاد ۶×۶، مطابق شکل زیر در نظر بگیرید. هر گاه عامل به خانه‌ی هدف (G) برسد، پاداش ۱۰ و هرگاه به مانع برخورد کند، تنبیه ۱- دریافت می‌کند. عامل در هر خانه قادر است به یکی از چهار جهت اصلی تغییر مکان دهد. هدف آن است که عامل به خانه‌ی هدف (G) برسد، به عبارت دیگر، عامل باید وظیفه رسیدن به خانه‌ی هدف را در محیط زیر بیاموزد. در آغاز، عامل به طور تصادفی در یکی از خانه‌های محیط قرار می‌گیرد.



به سوالات زیر پاسخ دهید:

i. یک episode یا trial در این محیط چگونه تعریف می‌شود؟

ii. جدول Q را چگونه تعریف می‌کنید؟

iii. عامل را با الگوریتم یادگیری Q را با انتخاب عمل به روش ε-Greedy آموزش دهید:

iv. عامل را با الگوریتم SARSA را با انتخاب عمل به روش ε-Greedy آموزش دهید:

گفتنی است که الگوریتم SARSA به صورت زیر است:

```
Initialize Q(s, a) arbitrarily
Repeat (for each episode):
   Initialize s
   Choose a from s using policy derived from Q (e.g., ε-greedy)
   Repeat (for each step of episode):
      Take action a, observe r, s'
      Choose a' from s' using policy derived from Q (e.g., ε-greedy)
      Q(s, a) ← Q(s, a) + α[r + γQ(s', a') − Q(s, a)]
      s ← s'; a ← a';
   until s is terminal
```

v. مقادیر پارامترهای یادگیری نظیر: نرخ یادگیری، نرخ تخفیف را چگونه مقداردهی می‌کنید؟ چرا؟

vi. عملکرد عامل را بر حسب دو معیار زیر ارزیابی کنید، نمودار تغییرات هر معیار را در طول یادگیری در هزار episode، رسم کنید.

- Discounted Cumulative Reward per Episode
- Steps needed to reach goal per Episode

vii. نتایج بدست آمده از اجرای هر دو الگوریتم را با هم مقایسه کنید. به نظر شما کدام الگوریتم برای یادگیری در این محیط مناسب‌تر است؟ چرا؟