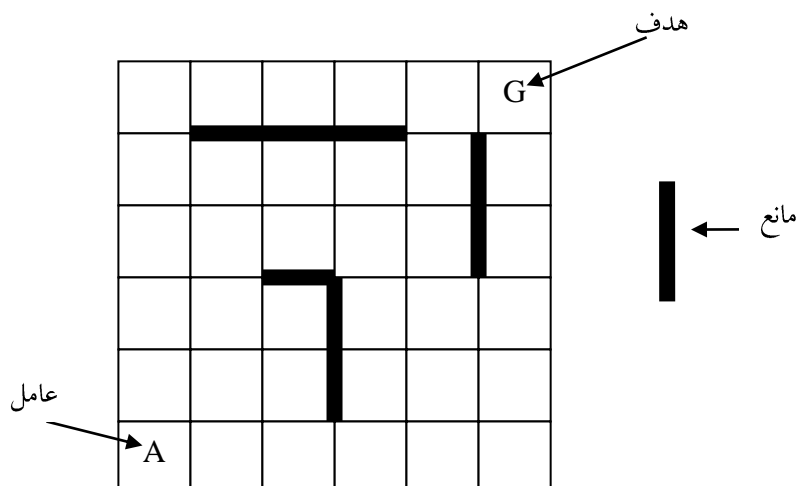


یادگیری ماشین - نیمسال اول ۱۳۹۳-۹۴
تکلیف شماره ۲ - تحویل چهارشنبه ۱۳۹۳/۹/۱۹

آشنائی با یادگیری تقویتی

هدف از این تکلیف آشنائی با روشهای یادگیری تقویتی و پیاده سازی یادگیری Q می باشد.

- برنامه ای به زبان دلخواه خود بنویسید که الگوریتم های یادگیری $SARSA(\lambda)$ و $Q(\lambda)$ را در یک دنیای شبکه 6×6 پیاده سازی نماید. عامل باید خود را از هر خانه ای که در آن قرار گرفته است بدون برخورد به مانع به خانه هدف برساند.



برنامه را بصورتی بنویسید که حداقل پس از یادگیری بتوان شبکه و حرکت عامل را مشاهده نمود و خانه شروع را انتخاب نمود. پس از مرحله یادگیری برنامه باید بتواند از هر خانه شروعی خود را بدون برخورد با مانع و با کمترین فاصله و حداقل حرکات به هدف برساند. در صورت برخورد به موانع و یا دیواره های شبکه، عامل در خانه فعلی خود باقی خواهد ماند ولی حرکت انجام شده برای او در نظر گرفته خواهد شد.

پاداشها برای همه حرکات صفر بوده به جز هنگام رفتن به خانه هدف که پاداش ۱۰ خواهد گرفت. حرکات نیز بصورت احتمالی همانند مثال کتاب انجام خواهد شد (با احتمال ۰.۸ کار خواسته شده انجام شود و با احتمال ۰.۱ به یکی از دو سمت عمود به حرکت خواسته شده خواهد رفت). مقدار γ را نیز ۰.۹ در نظر بگیرید.

پس از انجام عمل فوق سیاست بهینه را فرا گرفتید آنرا ذخیره نمایید و با استفاده از روش برنامه نویسی پویای ولفی و روش $TD(\lambda)$ سودمندی هر حالت را محاسبه کنید و بر روی آن نمایش دهید. روند پیشرفت محاسبه سودمندی حالات $(1, 1)$ ، $(3, 3)$ ، و $(5, 5)$ توسط دو روش را در دو نمودار رسم نمایید.

برنامه را به گونه ای بنویسید که براحتی بتوان پارامترها را تغییر داد و برنامه را آزمود. فراموش نکنید که برنامه ها را بصورت یک روال محیط و یک روال عامل طراحی نمائید. برنامه به همراه یک گزارش درباره نحوه پیاده سازی آن بایستی تحویل داده شود. سعی کنید برنامه را به صورت شیء گرا پیاده سازی نمائید و اشیاء استفاده شده خود را در گزارش ذکر نمائید.

لطفاً نکات زیر را در نظر بگیرید:

- به تکالیف مشابه نمره ای تعلق نخواهد گرفت.
- تکالیف فقط تا دو روز بعد از تاریخ موعود قابل تحویل هستند و به ازاء هر روز تأخیر ۱۰٪ از نمره تکالیف کاسته خواهد شد.
- فقط به تکالیفی که در سامانه تحویل داده شوند نمره داده خواهد شد.

موفق باشید

پالهنګ