



دانشگاه صنعتی اصفهان
دانشکده مهندسی برق و کامپیوتر

درس هوش مصنوعی
تکلیف چهارم

تاریخ تحویل: ۱۴۰۲

۱. به سوالات زیر پاسخ دهید.

- سود (Utility) در مسائل تصمیم‌گیری چیست و چه چیزی را نشان می‌دهد؟ توضیح دهید.
- سود موردانتظار (Expected Utility) برای یک تصمیم چگونه محاسبه می‌شود؟
- منظور از Policy و Optimal Policy برای یک شبکه‌ی تصمیم‌گیری چیست؟
- توضیح دهید که چرا در MDP از ضریب Discount استفاده می‌کنیم. رفتار یک عامل (Agent) زمانی که Discount Factor برابر 0.9 است با زمانی که برابر 0.6 است چه تفاوتی می‌کند؟
- تفاوت اصلی میان asynchronous value iteration و synchronous value iteration چیست و کدام یک معمولاً بهتر عمل می‌کند؟

۲. فرض کنید در یک شبکه‌ی تصمیم، متغیر تصمیمی (Decision Variable) به اسم Run وجود دارد و دارای دو والد (Parent) به نام‌های Look و See می‌باشد که متغیرهای تصادفی (Random Variable) بولی هستند. فرض کنید در حال اجرای الگوریتم حذف متغیرها هستیم تا به Optimal Policy برسیم و همیه‌ی متغیرهای دیگر را حذف کرده ایم و به فاکتور زیر رسیده ایم.

<i>Look</i>	<i>See</i>	<i>Run</i>	<i>Value</i>
<i>true</i>	<i>true</i>	<i>yes</i>	23
<i>true</i>	<i>true</i>	<i>no</i>	8
<i>true</i>	<i>false</i>	<i>yes</i>	37
<i>true</i>	<i>false</i>	<i>no</i>	56
<i>false</i>	<i>true</i>	<i>yes</i>	28
<i>false</i>	<i>true</i>	<i>no</i>	12
<i>false</i>	<i>false</i>	<i>yes</i>	18
<i>false</i>	<i>false</i>	<i>no</i>	22

- فاکتوری که پس از حذف متغیر Run بدست می‌آید را بدست آورید. (به نوع متغیر Run دقت کنید).
 - Optimal Policy را برای Run بدست آورید.
 - با دانستن Look، رفتار عامل چه تغییری می‌کند؟
۳. فرض کنید یک دانش‌آموز می‌خواهد تصمیم بگیرد که آیا کتاب مربوط به بک درس را خریداری کند یا خیر. این مسئله تصمیم‌گیری را با یک Decision Node به اسم B و دو Chance Node به اسم‌های M و P و یک Utility Node به اسم U مدل می‌کنیم. در این مسئله M به معنای مسلط شدن به درس و P به معنای پاس شدن در امتحان درس است. فرصت کنید سود دانش‌آموز از نخریدن کتاب 0 و از خریدن کتاب \$100- و از پاس شدن امتحان \$2000 و از پاس نشدن امتحان 0 است. احتمالات شرطی این مسئله به شکل زیر هستند. فرض می‌کنیم امتحان این درس کتاب‌باز است پس متغیرهای P و B با داشتن M از یکدیگر مستقل نیستند.

$$\begin{aligned}
 P(p|b, m) &= 0.9 & P(m|b) &= 0.9 \\
 P(p|b, \neg m) &= 0.5 & P(m|\neg b) &= 0.7 \\
 P(p|\neg b, m) &= 0.8 \\
 P(p|\neg b, \neg m) &= 0.3
 \end{aligned}$$

- شبکه‌ی تصمیم این مسئله را رسم کنید.
- سود موردانتظار (Expected Utility) خریدن و نخریدن کتاب را محاسبه کنید.

- تصمیم بهینه کدام است؟

۴. مثال Grid World بررسی شده در کلاس را در نظر بگیرید. فرض کنید $R(s_{24})$ برابر 1 و $R(s_{34})$ برابر 1 است و R برای استیت های دیگر برابر 0.05- می باشد و عامل با احتمال 0.8 در جهت مورد نظر و با احتمال 0.1 به سمت چپ و با احتمال 0.1 به سمت راست حرکت میکند. فرض کنید Discount Factor برابر 0.9 است.

- الگوریتم Value Iteration را تا 4 مرحله انجام دهید و Optimal Policy در این مرحله را بدست آورید.
- الگوریتم policy Iteration را تا 4 مرحله انجام دهید و Optimal Policy در این مرحله را بدست آورید.
- فرض کنید عاملی در یک trial از مسیر $s_{11}-s_{21}-s_{31}-s_{32}-s_{33}-s_{34}$ (از راست به چپ) حرکت کرده است و V^π حاصل از خروجی Value Iteration قسمت قبل را بدست آورده است، با استفاده از Temporal Difference Learning و با نرخ یادگیری برابر با یک، همه ی مقادیر V^π برای state های داخل مسیر را یکبار آپدیت کنید.

نکات تکمیلی

۱. پاسخ ها باید خوانا باشند و در قالب فایل pdf ارسال شوند.

۲. فرمت نام گذاری تکلیف ارسالی باید به صورت زیر باشد :

HWX_LastName_StudentID

که X شماره تکلیف ، LastName نام خانوادگی شما و StudentID شماره دانشجویی شما است.

۳. انجام این تکلیف به صورت تکی نفره است. در صورت مشاهده تقلب، نمرات هم مبدا کپی و هم مقصد آن صفر لحاظ می شود.

۴. در صورتی که مجموع تاخیر در ارسال برای همه ی تکالیف کمتر از 8 روز باشد نمره ای کسر نمی شود و در غیر این صورت به ازای هر روز ده درصن نمره کسر میگردد.

۵. در صورت وجود هر گونه ابهام و یا سوال می توانید سوالات خود را در گروه تلگرام بپرسید. همچنین می توانید برای رفع ابهامات با دستیاران آموزشی از طریق تلگرام در تماس باشید.