هوش مصنوعی - تکلیف چهارم

موعد تحویل ۲۸ خرداد ۱۴۰۰

پیش از حل سوالات به موارد زیر دقت کنید:

- تکلیف شامل سه سوال تئوری و یک سوال عملی می‌باشد.
- پاسخ قسمت تئوری را به صورت یک فایل PDF آماده کنید و به همراه فایل‌های مربوطه به سوالات عملی به صورت فشرده شده با نام {HW2_{Student Number در سامانه آپلود کنید.
- در تحویل تکالیف به زمان مجاز تعیین شده دقت نمایید. موعد تکالیف قابل تمدید نمی‌باشند. اما تا یک هفته پس از موعد اعلام شده با تاخیر تحویل گرفته می‌شوند.
- در صورتی که مجموع تاخیر کل تکالیف شما کمتر از ۲۴ ساعت باشد نمره‌ای از شما کسر نمی‌گردد. در غیر این صورت به ازای هر روز تاخیر ده درصد از نمره تکلیف شما کسر می‌گردد.
- پاسخ تکالیف را حتما در سامانه آپلود کنید و از ارسال تکالیف به ایمیل یا تلگرام اکیدا خودداری نمایید.
- در صورت وجود شباهت غیر قابل اغماض نمره‌ای به سوال تعلق نمی‌گیرد.
- در صورت وجود هرگونه ابهام می‌توانید در گروه تلگرام یا گروه اسکایپ سوالات خود را مطرح کنید.
- از طریق ایمیل‌های زیر می‌توانید با ta درس در ارتباط باشید.
  - mroghani+ai@ec.iut.ac.ir
  - sahandzoufan79@gmail.com

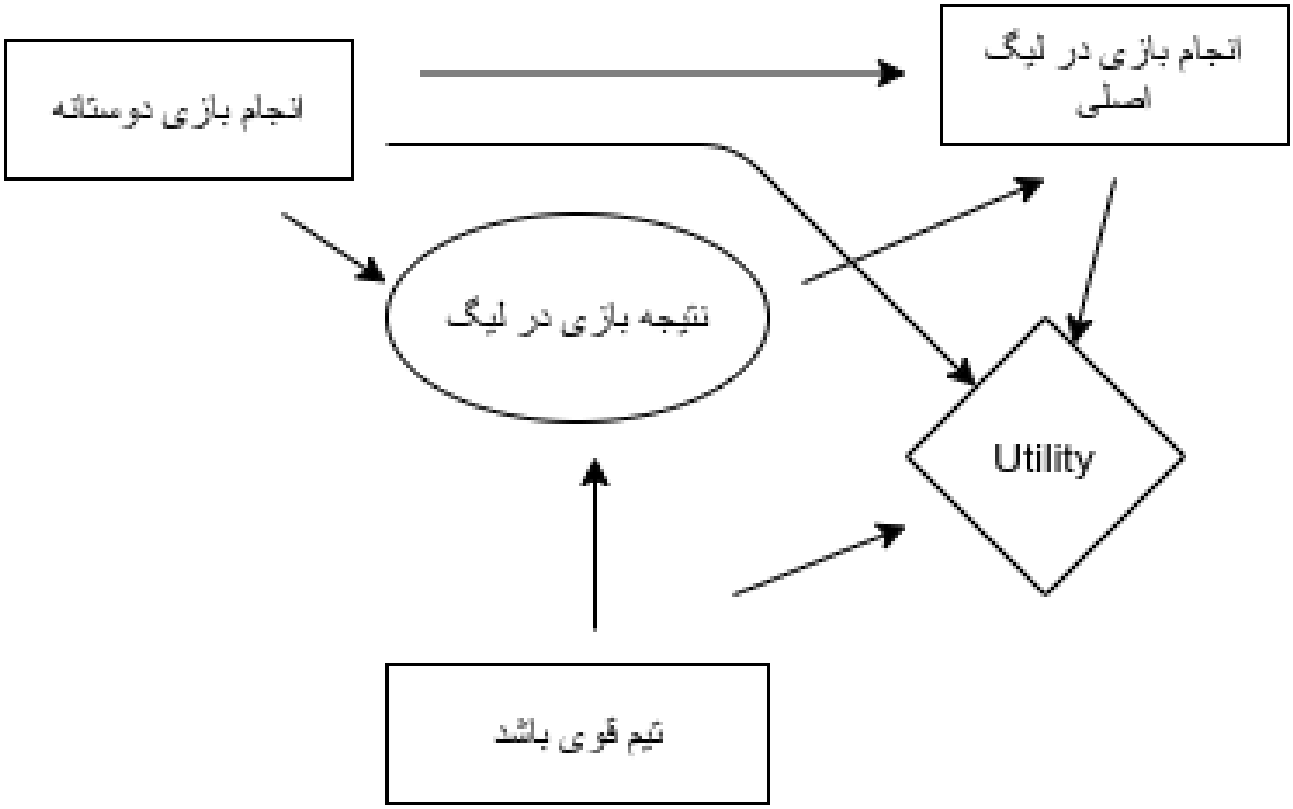
سوال ۱.

یک لیگ حذفی فوتبال فرضی را در نظر بگیرید با قوانینی که در ادامه گفته میشود.

اول اینکه هر تیم میتواند درخواست بازی دوستانه با تیم دیگری را بدهد. همچین میتواند انتخاب کند که دوست دارد با چه تیم هایی در لیگ بازی کند. اما باید به این نکته توجه داشت که با انجام بازی های دوستانه با تیم دیگر ممکن است استراتژی های آن تیم برای تیم دیگر معلوم شود و از این جهت علاوه بر آماده سازی تیم برای بازی های اصلی بازی دوستانه این عیب را نیز دارد. حال خود را به عنوان تیمی در این لیگ در نظر بگیرید بنابر آمار گذشته میدانیم که تیم شما ۸۰ درصد از بازی های خود را با تیم های دیگر شکست میخورد و تنها در ۲۰ درصد موارد بازی را میبرد. که تیم هایی که از آن ها میبازیم را $Strong$ مینامیم و بقیه تیم ها را $Weak$

هدف ما این است که با استفاده شبکه تصمیم زیر و Utility و احتمالات داده شده تصمیم بگیریم که آیا باید بازی دوستانه انجام دهیم یا خیر. و آیا اصلا باید با تیمی در لیگ بازی کنیم و یا خیر و اگر بله در چه حالت هایی میتوانیم بازی کنیم؟ همچینین دو احتمال زیر را نیز در دسترس داریم.

$$P(Win = True|Team = Strong) = 0.15$$
$$P(Win = True|Team = Weak) = 0.9$$



جدول میزان رضایت ها از هر تصمیم نیز در زیر آمده است (میزان رضایت را میتوان ترکیبی از میزان پول پرداختی و میزان خوشنودی دریافتی از تصمیم حساب کرد).

Utility	تیم قوی باشد	انجام بازی در لیگ اصلی	انجام بازی دوستانه
80	T	T	T
-100	F	T	T
30	T	F	T
-30	F	F	T
100	T	T	F
-80	F	T	F
0	T	F	F
0	F	F	F

حال به سوالات زیر پاسخ دهید.

آ) اگر از الگوریتم حذف متغیر استفاده کنیم فاکتور های اولیه ما چه توابعی هستند؟

ب) کدام Decision variable ابتدا حذف میشود و چرا؟

ج) بهترین تصمیم برای بازی دوستانه و بازی واقعی در لیگ چیست؟

د) Útility مورد انتظار از تصمیم گیری سوال قبلی چند است؟

سوال ۲. دو مسله یافتن مجموعه های همبند در گراف و کوتاه ترین مسیر در گراف ساده (V, E) را فرض کنید که V مجموعه راس ها و E مجموعه یال ها میباشد و v_G راس هدف میباشد. روش حل این مسله را به وسیله مدل سازی به MDP توضیح دهید(مقادیر reward هر نود از گراف و همچنین discount factor مشخص شود.)

سوال ۳. همگرا بودن الگوریتم value iteration به مقدار بهینه را اثبات کنید.

سوال ۴. همانطور که می‌دانید برای تعلیم الگوریتم‌های یادگیری تقویتی نیاز به یک محیط می‌باشد. OpneAI gym یک ابزار بسیار قوی شامل انواع و اقسام environmentهای مختلف برای یادگیری تقویتی است. environmentهای gym غالبا به عنوان یک محیط استاندارد برای آموزش و مقایسه عملکرد الگوریتم‌های یادگیری تقویتی محسوب می‌شود.
در این سوال شما باید الگوریتم value iteration را برای یافتن policy بهینه در محیط NChain-v0 پیاده‌سازی کنید.

در این محیط دو حرکت می‌توان انجام داد:

- حرکت به جلو که باعث می‌شود به استیت بعدی بریم.
- حرکت به عقب که باعث می‌شود به استیت اول برگردیم.

پاداش استیت اول یک پاداش کوچک است و پاداش استیت آخر یک پاداش بزرگ. بقیه استیت‌ها نیز پاداش ندارند.
در هر حرکت یک احتمال وجود دارد که agent لیز بخورد و به جای حرکتی که انتخاب کرده است حرکت دیگر اجرا شود. مثلا اگر احتمال لیز خوردن ۲۰ درصد باشد و ایجنت تصمیم بگیرد جلو برود به احتمال ۸۰ درصد جلو می‌رود و به احتمال ۲۰ درصد لیز می‌خورد و عقب می‌رود.

آ) الگوریتم value iteration را پیاده‌سازی کنید و $V^*(s)$ را چاپ کنید. (۲۰ نمره) (مقادیر محیط:
(n=5, small=2 large=10, slip=0.2, gamma=0.9

ب) تاثیر γ را روی یادگیری بیان کنید. (۱۰ نمره) (مقادیر محیط:
(n=5&10, small=2, large=10&100, slip=0.2&0.05, gamma=0.9&0.99&0.9999

ج) تاثیر تعداد استیت‌ها را روی یادگیری بیان کنید. (۱۰ نمره) (مقادیر محیط:
(n=5&10&1000, small=2 large=10&100,slip=0.2, gamma=0.9&0.9999

د) تاثیر پاداش بزرگ و کوچک را روی یادگیری بیان کنید. (۵ نمره) (مقادیر محیط:
(n=5&20t, small=2&4, large=10&100,slip=0.2, gamma=0.9

ه) آیا policy بهینه همیشه حرکت رو به جلو است؟ توضیح دهید. (۵ نمره)