سؤالات تحلیلی

در سوالات زیر لازم است مدلی مبتنی بر Multi-Armed Bandit ارائه کنید. برای هر مدل ارائه شده بازوها، پاداش و چگونگی پاسخ‌دهی به مسئله را به طور دقیق بیان کنید

۱) فرض کنید در ابتدای هر ماه مبلغی را به عنوان حقوق دریافت می‌کنید و می‌خواهید مقدار خاصی از آن را برای سرمایه‌گذاری استفاده کنید. برای این کار می‌توانید این پول را در بانک قرار دهید و در انتهای ماه اصل پول را با ۵ درصد سود آن دریافت کنید یا آنکه می‌توانید با این پول در بورس سرمایه گذاری کنید و در انتهای ماه سرمایه خود را از بورس بیرون بکشید. یک مدل Multi-Armed Bandit برای این مسئله طراحی کنید

۲) یکی از مسائلی که انسان در طول زندگی و به صورت روزانه با آن درگیر است، چگونگی اولویت بندی برای انجام کارهای متفاوت می‌باشد. فرض کنید شما یک دانشجو می‌باشید که در یک شرکت نیز مشغول به کار هستید. شما می‌خواهید برای تخصیص زمان بین چهار کار شامل وظایف خود در دانشگاه، انجام کارهای شرکت، روابط اجتماعی و گذراندن وقت با دوستان و آشنایان و پرداختن به امور شخصی خودتان اولویت‌بندی انجام دهید. یک مدل Multi-Armed Bandit برای حل این مسئله طراحی کنید(می‌توانید به جای ۴ کار گفته شده، بسته به شرایط و اولویت‌های خود ۴ کار دیگر را مشخص کرده و مسئله را برای آن حل کنید)

۳) فرض کنید که شما مدیر تبلیغات یک فروشگاه اینترنتی می‌باشید و از شما خواسته شده که با توجه به محصولات موجود در فروشگاه و از بین روش‌های مختلف انجام تبلیغات(شبکه‌های اجتماعی، تبلیغات تلویزیونی، تبلیغ در سطح شهر و ...)، استراتژی‌ای را پیدا کنید تا میزان فروش شرکت بیشینه شود. یک مدل Multi-Armed Bandit برای حل این مسئله طراحی کنید. با فرض اینکه این مدل طراحی شده حل شده و تابع پاداش هر کدام از بازوها را می‌دانیم، به نظر شما شرکت چگونه باید هزینه‌های تبلیغات را بین انتخاب‌های مختلف مدیریت کند؟

سؤال طراحی و پیاده‌سازی

یک شرکت آموزشی در حوزه مهندسی قصد دارد هر هفته از طریق اجرای سه برنامه تبلیغاتی هفتگی متفاوت متقاضیان کار را تشویق به شرکت در دوره آموزشی خود نماید. شرکت به صورت هفتگی ۱۰ نفر از متقاضیان را به صورت فشرده آموزش داده و به صنایع مختلف برای استخدام معرفی می‌کند. برنامه تبلیغاتی اول آگهی در سایت‌های کاریابی می‌باشد که برای هر بار، هزینه‌ای معادل ۱۰ میلیون تومان دارد. برنامه تبلیغاتی دوم از طریق تبلیغات محیطی می‌باشد که هزینه آن ۱۴ میلیون می‌باشد و برنامه تبلیغاتی سوم استفاده از شبکه‌های اجتماعی است که هزینه آن ۲ میلیون تومان می‌باشد. شرکت با اجرای این برنامه‌ها، می‌تواند متقاضیانی از بین افراد junior، mid-level و senior را جذب کند. هزینه آموزش در این دوره فشرده برای نیروی junior برابر با ۸ میلیون، نیروی mid-level برابر با ۶ میلیون و نیروی senior برابر با ۳ میلیون می‌باشد. شرکت در هر هفته تنها می‌تواند یکی از برنامه‌های تبلیغاتی را برای تنها یکی از گروه‌های هدف اجرا کند. در هر یک از این روش‌ها حداقل ۵۰ نفر به شرکت مراجعه می‌کنند. احتمال انتخاب ۱۰ نفرمناسب از میان مراجعین برای دوره آموزشی با اجرای هر کدام از برنامه‌ها در جدول یک نشان داده شده است. برنامه آموزش برای ۱۰ نفر اجرا شده

و در صورتی که تعداد متقاضی کمتر از ۱۰ نفر باشد برگزار نمی‌گردد. همچنین احتمال جذب هر فرد در صنعت پس از گذراندن دوره آموزشی برای هر گروه در جدول دو به شما داده شده است. اطلاعات این دو جدول تنها برای پیاده‌سازی محیط به شما داده شده است و از دید شرکت و عامل آن مخفی می‌باشد. به ازای جذب هر فرد در صنعت، شرکت آموزشی به ازای سطوح مختلف مبلغی دریافت می‌کند که در جدول شماره ۳ مشخص شده است. شرکت مایل است بهترین برنامه تبلیغاتی را که به صورت متوسط بیشترین سود را برای آن دارد، به کرّات اجرا کند.

جدول 1 :احتمال جذب ۱۰ نفر از هر گروه برای آموزش به ازای اجرای هر برنامه

	شبکه‌های اجتماعی	تبلیغات محیطی	آگاهی کاریابی
junior	0.8	0.3	0.5
mid-level	0.5	0.5	0.6
senior	0.2	0.7	0.8

جدول 2: احتمال استخدام در صنعت به ازای هر نفر در گروه های مختلف

junior	mid-level	senior
0.7	0.5	0.3

جدول 3: درآمد شرکت به ازای استخدام هر فرد متعلق به گروه ها مختلف در صنعت

Junior	mid-level	senior
27 میلیون	26 میلیون	25 میلیون

۱) یک مدل Multi-Armed Bandit برای این سوال طراحی کنید. بازوها و پاداش را به طور دقیق مشخص نمایید.

۲) براساس مدلی که طراحی کردید، مدل محیط را پیاده‌سازی نمایید.

۳) ۱۰ شرکت (عامل) در ۱۰ شهر مختلف می‌خواهند در محیط تعریف شده با استفاده از الگوریتم های یادگیری تعاملی برنامه تبلیغاتی بهینه را برای جامعه هدف، به منظور بیشینه کردن پاداش، بدست آورند. این شرکت ها برنامه دارند به مدت ۱۰ سال این تعامل با محیط را انجام دهند.(فرض کنید در طول این ۱۰ سال محیط ثابت می ماند) به ازای الگوریتم ها

یادگیرنده‌ی ε-Greedy و UCB ، با پارامتر های اپسیلون ثابت و برابر ۰.۱ و c برابر ۲ عملکرد شرکت‌ها را شبیه‌سازی کنید و نمودارهای مربوط به Regret و Reward را رسم کنید. هر نمودار باید شامل ۲ خم باشد که نتیجه اجرای ۲ الگوریتم متفاوت است.

نکات تمرین

- استفاده از LLM ها در این تمرین مشکلی ندارد. اما در صورت استفاده لطفاً منبع و prompt خود را ذکر نمایید تا تقلب محسوب نشود.
- مهلت ارسال این تمرین تا پایان روز چهارشنبه ۲۴ آبان ماه خواهد بود.
- انجام این تمرین به‌صورت یک‌نفره می‌باشد. اما بحث و گفت‌وگو در دیسکورد مانعی ندارد.
- لطفاً گزارش و کد تمرین را در فایل‌هایی که از طریق google Doc و google colab با شما به اشتراک گذاشته شده است، وارد نمایید.
- در صورت وجود سؤال و یا ابهام می‌توانید در channel مربوط به این تمرین با دانشجویان دیگر مطرح نمایید و یا برای ارتباط با دستیاران آموزشی از طریق ایجاد یک thread در همان channel دیسکورد، سؤال خود را مطرح نمایید.