به نام او، برای او

تمرین دوم درس یادگیری ماشین

پاییز ۱۳۹۸

سوال اول:

یک شرکت آرایشی بهداشتی قصد دارد محصول جدیدی به بازار معرفی کند. برای این‌کار نسخه‌های متفاوتی از محصول جدید را و با اندکی تغییر در ساختار تهیه‌ی آن‌ها تولید کرده‌است. این شرکت می‌خواهد از کیفیت محصولی که به بازار عرضه می‌کند اطمینان حاصل کند و بنابراین پیش از ارائه‌ی محصول تصمیم دارد روی گروهی از افراد نسخه‌های مختلف محصول را امتحان کند و نظرات آن‌ها را بپرسد. در مجموع ۶ نسخه‌ی متفاوت از محصول ایجاد شده است و در اختیار افراد قرار گرفته‌است. این شرکت از افراد می‌خواهد بین ۰ تا ۱۰۰ به هر محصول امتیاز دهند. اما با توجه به مواد اولیه‌ی نسخه‌های مختلف این محصول، قرار دادن محصول برای ارزیابی توسط افراد، هزینه‌ای نیز در بر خواهد داشت.

این شرکت می‌خواهد با توجه به هزینه‌ای که صرف این‌کار می‌کند و نظراتی که دریافت می‌کند معیاری برای ارزشمندی هر محصول محاسبه کند. معیار ارزشمندی را به صورت زیر تعریف می‌کنیم:

$$\text{Value} = a \times \text{Point} - \text{Cost}$$

در این مثال ضریب a را برابر ۲,۵ در نظر بگیرید.

در فایل Dataset.csv برای هر ۶ محصول امتیازات کاربران و هزینه‌ای که برای در دسترس افراد قرار دادن می‌کنیم وجود دارد.

توجه کنید که این مسئله یک مسئله‌ی یادگیری تقویتی است و هدف ما این است که با تعداد تجربه‌ی کمتر محصول بهتر را پیدا کنیم.

با توجه به توضیحات گفته‌شده به سوالات زیر پاسخ دهید:

۱) با استفاده از روش reinforcement comparison برای دو مقدار آلفا = ۰,۱ و بتا = ۰,۹ مسئله را حل کنید.

۲) نمودار پشیمانی در حسب تعداد تجربه را رسم کنید. به نظر شما مقادیر آلفا و بتا چه رابطه‌ای باید با هم داشته باشند؟

۳) آیا راهی وجود دارد که یک سیاست حریصانه به جواب همگرا شود؟

سوال دوم:

در یکی از روزهای پاییز، در یکی از واگن‌های مترو خط ۴، قتلی اتفاق افتاده و علی‌رغم تلاش شبانه‌روزی کارآگاهان، راز قتل حل نشده. به همین دلیل کارآگاهان به کمک‌های بین‌المللی برای گشودن گره این جنایت هولناک نیاز پیدا کرده‌اند. پس از نامه‌نگاری‌های فراوان، آقای هرکول پوآرو، خانم مارپل و شرلوک هلمز به تهران آمده‌اند تا راز جنایت را فاش کنند.

در بازدید اول از صحنه‌ی قتل و شواهد بدست آمده، هر یک از این ۳ کارآگاه ۱۰۰۰ سرنخ بدست آوردند که برای بررسی آنها باید به ۱۰۰۰ نقطه در شهر تهران سفر کنند. هر یک از کارآگاهان ترجیح می‌دهند که تحقیقات خود را به صورت جداگانه انجام دهند. برای این حجم از سفرها، کارآگاه شمسی و مادام به عنوان میزبانان مهمانان خارجی، به این سه نفر استفاده از سرویس‌های تاکسی اینترنتی موجود در تهران را پیشنهاد کرده است.

سه سرویس تاکسی اینترنتی در تهران فعال هستند، تَپْتاکسی، تاکسینَپ و تاکسیم. با هماهنگی جامعه کارآگاهان مقیم مرکز با این شرکت‌های تاکسی اینترنتی، قرار شده از سه مهمان آنها، هیچ هزینه‌ای برای سفرهایشان دریافت نشود. هر یک از این سه کارآگاه از روش win stay, lose shift در تصمیم‌گیری خود برای انتخاب یک سرویس تاکسی اینترنتی برای انجام سفر استفاده می‌کنند. به این صورت که در ابتدا، یک سرویس را به صورت تصادفی انتخاب می‌کنند. اگر در سرویس انتخاب شده، سفیری برای مسافران پیدا شد، سفر انجام می‌شود ولی اگر سفیری پیدا نشد، علاوه بر تلف شدن زمان ارزشمند مسافر، دو حالت ممکن است رخ دهد: ۱) ممکن است دوباره در همان سرویس تاکسی اینترنتی درخواست سفر بدهد یا ۲) به صورت تصادفی، یکی از دو سرویس دیگر تاکسی اینترنتی را انتخاب کند و درخواست سفر خود را در سرویس جدید انتخاب شده ارسال کند. و به همین ترتیب سفرهای بعدی نیز انجام می‌شود.

احتمال پیدا شدن راننده در سرویس‌های مختلف یکسان نیست و زمان تلف شده مسافر در صورت پیدا نشدن راننده نیز برای هر سرویس متفاوت است. نحوه‌ی تصمیم‌گیری کارآگاهان به صورت زیر است:

- آقای هرکول پوآرو در صورت پیدا شدن راننده، در درخواست سفر بعدی، با احتمال ۰٫۹ از همان سرویس استفاده می‌کند و در صورت پیدا نشدن راننده، با احتمال ۰٫۹ سرویس تاکسی اینترنتی خود را تغییر می‌دهد.
- خانم مارپل در صورت پیدا شدن راننده، در درخواست سفر بعدی، با احتمال ۰٫۹ از همان سرویس استفاده می‌کند و در صورت پیدا نشدن راننده، با احتمال ۰٫۲ سرویس تاکسی اینترنتی خود را تغییر می‌دهد.
- شرلوک هلمز در صورت پیدا شدن راننده، در درخواست سفر بعدی، با احتمال ۰٫۳ از همان سرویس استفاده می‌کند و در صورت پیدا نشدن راننده، با احتمال ۰٫۸ سرویس تاکسی اینترنتی خود را تغییر می‌دهد.

همچنین در مورد تاکسی‌های اینترنتی می‌دانیم:

- در صورت درخواست سفر از تَپْتاکسی، با احتمال ۰٫۱ راننده پیدا نمی‌شود و زمان تلف شده‌ی مسافر از توزیع $N(5,3)$ بدست می‌آید.
- در صورت درخواست سفر از تاکسینَپ، با احتمال ۰٫۳ راننده پیدا نمی‌شود و زمان تلف شده‌ی مسافر از توزیع $N(2,2)$ بدست می‌آید.
- در صورت درخواست سفر از تاکسیم، با احتمال ۰٫۷ راننده پیدا نمی‌شود و زمان تلف شده‌ی مسافر از توزیع $N(1,1)$ بدست می‌آید.

در حالی که کارآگاهان آماده می‌شوند تا تحقیقات خود را برای حل معما شروع کنند، با استفاده از شبیه‌سازی بررسی کنید که زمان کدام‌یک از کارآگاهان کمتر تلف خواهد شد؟

* در صورت منفی شدن زمان تلف شده‌ی مسافر برای هر سفر، قدر مطلق مقدار آن را در نظر بگیرید.

سوال سوم:

سارا برای رفتن به سر کار باید یکی از سه روش زیر را انتخاب کند:

- روش اول: با استفاده از مترو سر کار برود. در این صورت هزینه مسیر بر حسب هزار تومان از تابع توزیع $N(2, 0.0625)$ بدست می‌آید. اما با توجه به این که مسیری که سارا باید طی کند طولانی می‌شود، میزان تاخیر او بر حسب دقیقه از توزیع $N(0, 0.25)$ محاسبه می‌شود.
- روش دوم: با استفاده از اتوبوس سر کار برود. در این صورت هزینه مسیر بر حسب هزار تومان از تابع توزیع $N(3.5, 0.25)$ بدست می‌آید. در این صورت، میزان تاخیر او بر حسب دقیقه از توزیع $U(-3, 0.5)$ محاسبه می‌شود.
- روش سوم: با استفاده از تاکسی سر کار برود. در این صورت هزینه مسیر بر حسب هزار تومان از تابع توزیع $U(3.5, 4.5)$ بدست می‌آید. در این صورت هم میزان تاخیر او بر حسب دقیقه از توزیع $N(-2.5, 0.25)$ محاسبه می‌شود.

با توجه به این که سارا به ازای هر دقیقه دیر رسیدن 1500 تومان جریمه و به ازای هر دقیقه زود رسیدن 1000 تومان تشویق می‌شود، بهترین روش را با استفاده از الگوریتم UCB2 بدست آورید و نمودار میانگین پاداش هر روش را در طول یادگیری رسم کنید. مسئله را به ازای مقادیر مختلف پارامتر α حل کنید و در مورد نقش این پارامتر در پاسخ خود بحث کنید.

- تاخیر منفی به معنای زود رسیدن است.
- پارامتر دوم تابع توزیع نرمال واریانس است.

سوال چهارم (سوال امتیازی):

پاداش یک عامل یادگیر در یک مسئله 2-armed bandit یک ترکیب خطی از پاداش عمل و میزان اطمینان او به تصمیم اجرا شده است و عامل از این امر آگاه نیست. اطمینان عامل همان احتمال انتخاب عمل در مدل بولتزمن است. یادگیری این عامل را با حالتی که اثر اطمینان در پاداش وجود ندارد مقایسه کنید.