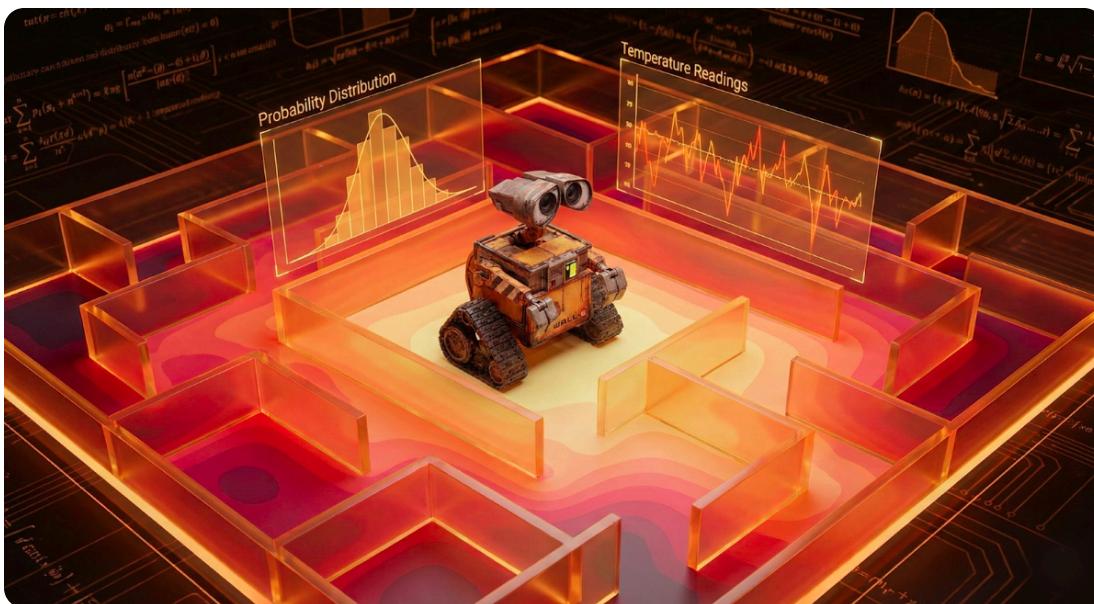




تمرین کامپیوتری پنجم - مجموع متغیرهای تصادفی، توزیع شرطی، توزیع بتا، کوواریانس و توابع دو متغیر تصادفی

طراحان: محمد عرفان دانایی، آرتین توسلی



پرسش ۱ - مجموع متغیرهای تصادفی و تابع دو متغیر تصادفی (۲۵ نمره)

یک شرکت می‌خواهد توزیع احتمال زیان مالی خود در پایان هفته‌ی آینده را برآورد کند. فرض کنید این شرکت با بت کالاهای مرجعی دچار ضرر می‌شود. تعداد کالاهای مرجعی نوع a و b در هر روز، از یک توزیع نمایی پیروی کرده و مستقل از هم هستند. شرکت برای هر کالای مرجعی، ۱ دلار ضرر می‌کند. متغیر X_i را تعداد کالاهای مرجعی نوع a و Y_i را تعداد کالاهای مرجعی نوع b در روز i -ام در نظر بگیرید.

$$X_i, Y_i \sim Exp(\lambda = 0.01)$$

۰. برای توزیع احتمالات بالا نمودار هیستوگرام و چگالی احتمال (PDF) داده‌ها را رسم کنید. (۵ نمره)

۱. میزان زیان شرکت ناشی از مرجع شدن محصول a پس از یک هفته را بدست بیاورید. فرض کنید این مقدار برابر با D است. (۱ نمره)

۲. با تولید ۱۰۰,۰۰۰ نمونه از زیان هر روز که توسط محصول a ایجاد شده است، ۱۰۰,۰۰۰ نمونه از D به دست آورید. سپس PDF و هیستوگرام آن را رسم کنید. (۲ نمره)

۳. میانگین و واریانس توزیع حاصل را محاسبه کنید. (۱ نمره)

می‌دانیم مجموع زیان شرکت پس از n روز، حاصل جمع n متغیر تصادفی نمایی مستقل است. این مجموع، از توزیع گاما (Gamma Distribution) پیروی می‌کند. است. میتوانید از این [لينک](#) بیشتر درباره این توزیع مطالعه بکنید.

۴. پارامترهای توزیع گاما را برای D به دست آورید و PDF آن را رسم کنید. (۲ نمره)

۵. میانگین و واریانس را به صورت تئوری محاسبه کرده و با بخش ۳ مقایسه کنید. (۱ نمره)

حال فرض کنید شرکت می‌خواهد بداند چند درصد از زیان شرکت پس از ۷ روز مربوط به محصول a است.

۶. رابطه ریاضی این نسبت را بنویسید. با تولید نمونه‌های تصادفی از زیان‌های شرکت، ۱۰۰,۰۰۰ نمونه از این نسبت را رسم کنید. همچنین میانگین و واریانس داده‌ها را محاسبه کنید. (۲ نمره) این توزیع جدید توزیع بتا (Beta Distribution) نامیده می‌شود.

میتوانید از این [لینک](#) بیشتر درباره این توزیع مطالعه بکنید.

۷. پارامترهای این توزیع را بدست بیاورید، سپس میانگین و واریانس آن را به صورت تئوری تخمین بزنید و با بخش ۶ مقایسه کنید. (۵ نمره)
۸. نمودار PDF توزیع بتا حاصل را رسم کرده و آن را با هیستوگرام بخش ۶ مقایسه کنید. (۲ نمره)
۹. حال فرض کنید شرکت می‌خواهد همین نسبت را برای یک روز (به جای یک هفته) محاسبه کند. این نسبت را مجدداً شبیه‌سازی کنید. (۲ نمره)
۱۰. توزیع حاصل شبیه به چه توزیعی است؟ این موضوع را با محاسبات تئوری بررسی و تایید کنید. (۲ نمره)

پرسش ۲ - کوواریانس و همبستگی (۲۵ نمره)

در این سوال از مجموعه داده `airquality` که در بسته‌های آماده R قرار دارد می‌خواهیم استفاده کنیم. این مجموعه داده مربوط به اطلاعات آب و هوای نیویورک در سال 1970 است. مرکز ما روی ستون‌های *Day* (روز)، *Month* (ماه) و *Temp* (دما) است. برای نمایش بهتر گذر زمان، یک متغیر جدید به نام *Day Number* تعریف می‌کنیم که نشان می‌دهد چند روز از شروع سال گذشته است.

۱. می‌خواهیم در این بخش کمی بیشتر با این مجموعه داده آشنا بشیم. نمودار دما (*Temp*) را بر حسب شماره روز (*Day Number*) رسم کنید. هدف این است که با این کار روند کلی تغییرات دما در طول سال را ببینید. (۷.۵ نمره)
۲. تابعی به نام `Calculate_Correlation` بنویسید که دو بردار *X* و *Y* را به عنوان ورودی بگیرد و همبستگی آنها را محاسبه کند. دقت کنید که باید فرمول ریاضی همبستگی را خودتان کدنویسی کنید و نباید از توابع آماده‌ی R برای این کار استفاده کنید. (۷.۵ نمره)

در داده‌های سری زمانی، معمولاً رابطه بین مقدار یک متغیر در زمان فعلی و مقادیر گذشته‌ی آن تحلیل می‌شود که به آن «خودهمبستگی» می‌گویند. این مفهوم نشان می‌دهد یک متغیر چگونه با خودش در فاصله‌های زمانی مختلف (که به آن *Lag* می‌گویند) ارتباط دارد. در این تمرین، خودهمبستگی دما را با «یک روز تاخیر» بررسی می‌کنیم.

۳. نمودار دمای روز *n* را در برابر دمای روز $1 - n$ (روز قبل) رسم کنید. (۴ نمره)
۴. با استفاده از تابع `Calculate_Correlation` که در مرحله قبل نوشته‌ید، همبستگی بین این دو سری دما را محاسبه کنید. به این کار محاسبه خودهمبستگی با $Lag = 1$ گفته می‌شود. (۱ نمره)
۵. حالا این ایده را گسترش دهید. همبستگی را برای تاخیرهای ۱ تا 100 روز محاسبه کنید و نمودار همبستگی دما را نسبت به روزهای گذشته رسم نمایید. (۵ نمره)

پرسش ۳ - توزیع شرطی (۵۰ نمره)

در این بخش می‌خواهیم موقعیت یک ربات (WALL-E) را در یک هزارتوی شبکه‌ای ۱۰ در ۱۵ با استفاده از سنسورهایش پیدا کنیم. ربات از اینکه در کدام موقعیت است هیچ اطلاع اولیه‌ای ندارد و تنها اطلاعاتی که دریافت می‌کند، ناشی از سنسورهایش است. ربات WALL-E دارای ۴ سنسور در ۴ جهت خود است که وجود یا عدم وجود دیوار را در هر جهت به او اطلاع می‌دهند. هر یک از این ۴ سنسور ۱۰٪ خطأ دارند. توزیع احتمال نتیجه هر سنسور به شرح زیر است:

$$P(\text{sensor} = 1 \mid \text{wall} = 1) = 0.9$$

$$P(\text{sensor} = 1 \mid \text{wall} = 0) = 0.1$$

$$P(\text{sensor} = 0 \mid \text{wall} = 1) = 0.1$$

$$P(\text{sensor} = 0 \mid \text{wall} = 0) = 0.9$$

WALL-E خیلی پیر است و پاهایش دیگر مثل سابق نیستند، بنابراین حتی اگر تصمیم بگیرد به سمت بالا برود، ممکن است به سمت جنوب حرکت کند!

با فرض هر جهت مورد نظر (Intended Direction)، احتمال ۸۰٪ برای حرکت صحیح و ۲۰٪ برای حرکت در جهت اشتباه (به صورت یکنواخت بین سه جهت دیگر) وجود دارد:

$$P(\text{move} = \text{UP} | \text{intendedDir} = \text{UP}) = 0.8$$

$$P(\text{move} = \text{DOWN} | \text{intendedDir} = \text{UP}) = \frac{0.2}{3}$$

$$P(\text{move} = \text{RIGHT} | \text{intendedDir} = \text{UP}) = \frac{0.2}{3}$$

$$P(\text{move} = \text{LEFT} | \text{intendedDir} = \text{UP}) = \frac{0.2}{3}$$

WALL-E ممکن است پیر باشد اما حافظه قوی‌ای دارد، او تاریخچه باورهایش (احتمال حضور در هر سلول در هر گام، که در ادامه توضیح داده می‌شود) را ذخیره می‌کند.

WALL-E می‌خواهد مکان خود را با دقت بالا با استفاده از روش‌های احتمالاتی و احتمال شرطی پیدا کند تا از هزارتو فرار کند. برای حل این مسئله، آن را با مجموعه‌ای از متغیرهای تصادفی مدل‌سازی می‌کنیم. فرض کنید X_t نشان‌دهنده موقعیت ربات پس از t حرکت باشد. در ابتدا، احتمال حضور ربات در تمام بلوک‌ها برابر است، بدین معنا که احتمال اولیه برای هر بلوک (x,y) برابر است با:

$$P(X_0 = (x, y)) = \frac{1}{\text{NUM OF FREE CELLS}}$$

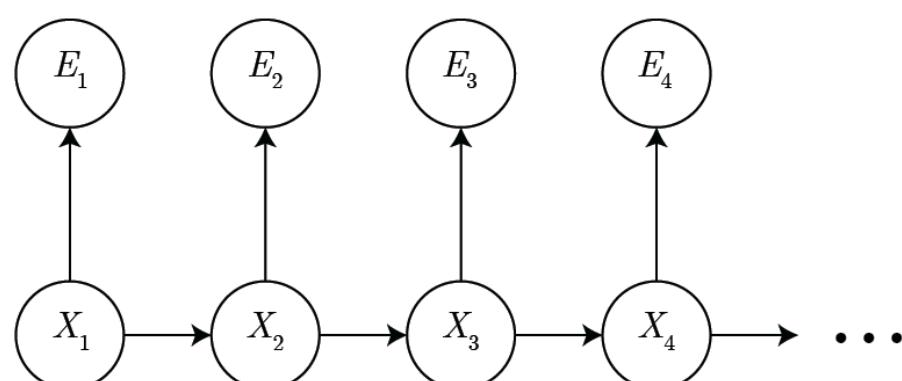
همچنین، فرض کنید داده‌های خوانده شده از سنسورها در زمان t را با E_t Evidence (شواهد یا پس از t حرکت).

متغیرهای X_t و E_t بردار هستند.

- موقعیت X_t WALL-E را با استفاده از مختصات x, y نشان می‌دهد، مثلاً [3,4].
- وضعیت سنسورهای WALL-E E_t را نشان می‌دهد، مثلاً [0,0,0,1] (تنها یک سنسور روشن است در حالی که بقیه خاموش هستند).

در مدل ما، X_t تنها به X_{t-1} وابسته است و هر E_t نیز تنها به X_t وابسته است (مستقل از سایر متغیرهای تصادفی).

می‌توانیم مجموعه‌ای از متغیرهای تصادفی را در نظر بگیریم:



هر فلش نشان‌دهنده وابستگی متغیر انتهایی فلش به متغیر ابتدایی فلش است.

ما به دنبال تخمین X_t بر اساس شواهد E_t تا E_1 برای هر بلوک از نقشه هستیم. بنابراین برای هر گام باید مقدار احتمال زیر را بدست آوریم:

$$P(X_t | E_1, E_2, \dots, E_t)$$

این احتمال را می‌توان به صورت زیر نوشت:

$$P(X_t | E_{1:t}) \propto P(X_t, E_{1:t}) = \sum_{X_{t-1}} P(X_{t-1}, X_t, E_{1:t})$$

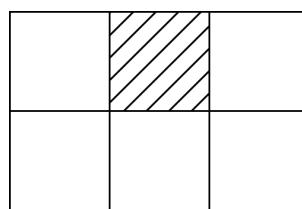
$$= \sum_{X_{t-1}} P(X_{t-1}, E_{1:t-1}) P(X_t | X_{t-1}) P(E_t | X_t) \propto P(E_t | X_t) \sum_{X_{t-1}} P(X_t | X_{t-1}) P(X_{t-1} | E_{1:t-1})$$

بنابراین ما به ۳ احتمال نیاز خواهیم داشت:

1. $P(E_t | X_t)$
 2. $P(X_t | X_{t-1})$
 3. $P(X_{t-1} | E_{1:t-1})$
-

۰. محاسبه تئوری احتمال حضور (۵ نمره)

- فرض کنید یک زمین ۲ در ۳ به شکل زیر داریم.



اگر ربات جهت پایین را انتخاب کند و پس از حرکت مشاهده زیر را انجام دهد:

$$E_2 = [\text{UP} = 0, \text{RIGHT} = 0, \text{DOWN} = 1, \text{LEFT} = 1]$$

آنگاه $P(X_2 | E_{1:2})$ را به دست آورید. (فرض کنید $P(X_1 | E_1)$ به طور یکنواخت برای همه خانه‌های خالی برابر ۰.۲ است). آیا نتیجه مطابق انتظار است؟

۱. پیدا کردن موقعیت بعدی WALL-E (۱۰ نمره)

- تابعی بنویسید که موقعیت فعلی WALL-E و جهت انتخاب شده توسط آن را بگیرد و موقعیت را پس از حرکت برگرداند.

۲. محاسبه $P(E_t | X_t)$ (۱۰ نمره)

- ۲.۱. به صورت تئوری احتمال فوق را با استفاده از $E_t = [0, 1, 0, 1]$ و $X_t = [3, 4]$ برای نقشه داده شده در کد محاسبه کنید.
- ۲.۲. تابعی بنویسید که آخرین شواهد (E_t) و موقعیت WALL-E (X_t) را بگیرد و احتمال $P(E_t | X_t)$ را برگرداند.
- ۲.۳. پاسخ قسمت ۲.۱ خود را با استفاده از تابع قسمت ۲.۲ بررسی کنید.

۳. محاسبه $P(X_t | X_{t-1})$ (۱۰ نمره)

- تابعی بنویسید که X_{t-1} ، X_t و جهت انتخاب شده توسط WALL-E را بگیرد و احتمالی را که WALL-E را با توجه به جهت ورودی از X_{t-1} به X_t برود، خروجی دهد.

◦ برای مثال، $f((3,4), (3,5), 'u')$ برابر است با احتمالی که WALL-E از خانه (3,5) به (3,4) برود در حالی که جهت بالا (Up) را انتخاب کرده است.

◦ اگر X_{t-1} دیوار باشد، باید ۰ برگردانیم.

◦ اگر X_{t-1} همسایه‌ی X_t نباشد یا خود X_t نباشد، واضح است که رفتن ربات از X_{t-1} به X_t غیرممکن است. بنابراین باید ۰ برگردانیم.

◦ اگر X_t دیوار باشد، باید ۰ برگردانیم.

۴. حلقه اصلی برنامه (۱۰ نمره)

- حلقه اصلی را با استفاده از فرمول مدل ارائه شده برای محاسبه $P(X_t | E_{1:t})$ بنویسید.
- نرمال‌سازی (Normalize) را فراموش نکنید! می‌توانید از تابع `normalize` برای نرمال‌سازی باور (Belief) استفاده کنید.

از شما تنها خواسته شده است بخشنده از حلقه اصلی (تابع `model`) را کامل کنید. تابع `model` یک لیست دوتایی شامل باورهای مدل و موقعیت‌های واقعی ربات را با گرفتن نقشه و جهت‌های انتخاب شده توسط ربات بر می‌گرداند:

```
return (list(beliefs = beliefs, real_pos = real_pos))
```

- لیستی برای ذخیره باور WALL-E در مورد مکانش در هر گام زمانی است. هر عنصر در این لیست یک ماتریس `beliefs` خواهد بود که توزیع احتمال حضور ربات در هر سلول شبکه (باور) را نشان می‌دهد. در ابتدا، ما یک توزیع یکنواخت به آن اضافه می‌کنیم.
- از `beliefs` برای شبیه‌سازی مدل و از `real_pos` برای شبیه‌سازی ردیابی واقعی WALL-E استفاده می‌شود. ما می‌توانیم عملکرد مدل خود را با مقایسه باورها و موقعیت‌های واقعی بسنجیم.

۵. ارزیابی نتایج (۵ نمره)

خروجی مدل خود را به توابع زیر بدهید:

- تابع `plot_belief_maxima` باور بیشینه درست‌نمایی (Maximum Likelihood) را در مقابل وضعیت واقعی رسم می‌کند.
 - تابع `plot_belief_accuracy` تعداد پیش‌بینی‌های صحیح در هر گروه ۵ تایی را به همراه دقت کلی چاپ می‌کند.
- ۱.۵. درصد خطای مدل را طبق جدول زیر تغییر دهید و نمودارها را برای ۳۰۰ حرکت مقایسه کنید.
- در هر مرحله، مقادیر `MOVE_ERROR` و `SENSOR_ERROR` را دوباره فراخوانی کنید و خروجی مدل را با توابع داده شده رسم کنید.

خطای حرکت	خطای سنسور
0.01	0.01
0.05	0.05
0.6	0.6
0.001	0.6
0.6	0.001

- نکته: از حلقه `for` استفاده نکنید و هر نمودار را در یک سلول جداگانه رسم کنید.
- ۲.۵. در نهایت، نتایج را تحلیل کنید و دلیل رفتار هر نمودار را بررسی نمایید. همچنین، اهمیت خطای سنسور و خطای حرکت را با توجه به نتایج مقایسه کنید.

نکات پایانی

- یک کپی از نوتبوک خود با نام EPS_CA5_SID در گوگل درایو خود تهیه کنید.
- در فایل خود بخش‌هایی که به وسیله TODO مشخص شده‌اند را با کدهای مناسب جایگزین کنید.
- تمام بخش‌های تئوری را داخل نوتبوک تان بیاورید.
- تمامی عکس‌ها و نمودارهای تولید شده باید تحلیل شده باشند و در گزارش کار موجود باشند.
- فایل دفترچه پاسخ خود را در جایگاه آپلود این تمرین در سامانه ایلن قرار دهید.

موفق باشید.