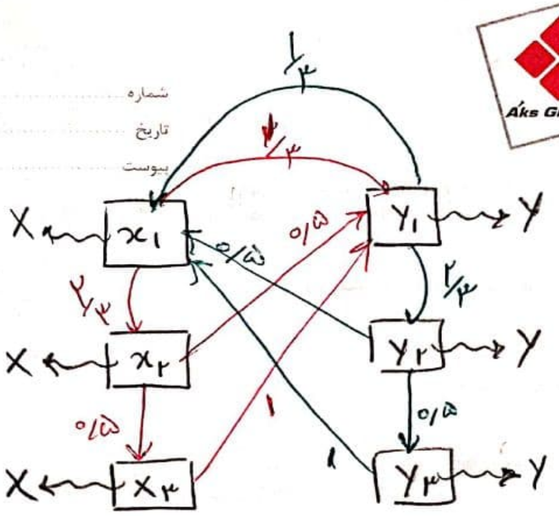
پاسخ تمرین پنجم، بخش اول — مدل مارکوف و HMM

سوالات نظری

۱. • الف



۶ تا state داریم

stateهای x_i فقط X را به عنوان emission دارند و y_i ها فقط Y

* x_1 نشان دهنده جایگاه اول است که x می‌آید و به همین ترتیب ...

Y_3 فقط به X_1 می‌تواند برود زیرا بعد از ۳ بار Y آمدن باید بعدش حتماً X بیاید (و برعکس هم همینطور است)

X_2 یعنی بعد از ۲ بار X آمدن با احتمال ۰٫۵ می‌توان X بیاید (۳ بار X بیاید) یا با احتمال ۰٫۵ Y بیاید (Y_1)

XXY
XXXY
XY
→ بعد از اولین X با احتمال $\frac{2}{3}$ X می‌آید و با احتمال $\frac{1}{3}$ Y

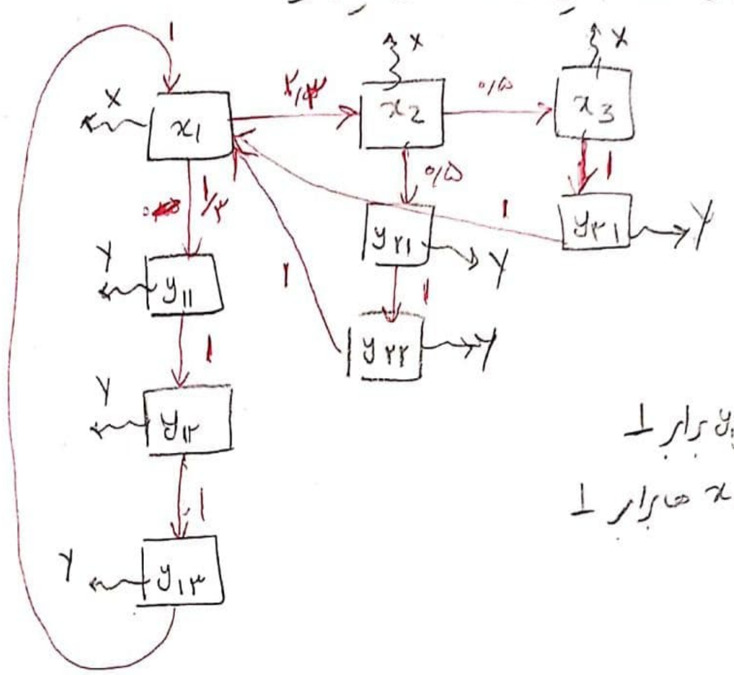
⇐ X_1 با احتمال $\frac{1}{3}$ به Y_1 و $\frac{2}{3}$ به X_2 می‌رود

	X_1	X_2	X_3	Y_1	Y_2	Y_3	X	Y
X_1	0	2/3	0	1/3	0	0	1	0
X_2	0	0	1/2	1/2	0	0	1	0
X_3	0	0	0	1	0	0	1	0
Y_1	1/3	0	0	0	2/3	0	0	1
Y_2	1/2	0	0	0	0	1/2	0	1
Y_3	1	0	0	0	0	0	0	1

transition — emission

• ب

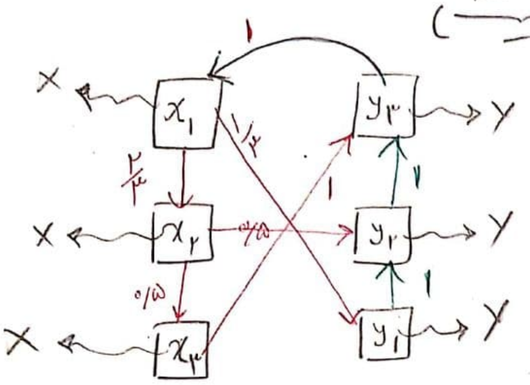
ب) مانند قسمت قبل است با این تفاوت که مینیمال نیست



احتمال Y برای y_{ij} برابر 1

و احتمال X برای x_i ها برابر 1

ب) حالت مینیمال (خواسته مسئله نیست)



	x_1	x_2	x_3	y_1	y_2	y_3	X	Y
x_1	0	2/3	0	1/3	0	0	1	0
x_2	0	0	1/2	0	1/2	0	1	0
x_3	0	0	0	0	0	1	1	0
y_1	0	0	0	0	1	0	0	1
y_2	0	0	0	0	0	1	0	1
y_3	1	0	0	0	0	0	0	1

۲. • الف

داده‌های مسئله: $e_1 = 1\ , e_2 = 1$ خواسته مسئله $p(x_2 = (2,1) \mid e_{1:2})$

$$p(x_2 = (2,1) \mid e_{1:2}) = \frac{p(x_2 = (2,1), e_2 \mid e_1)}{p(e_2 \mid e_1)} \propto p(x_2 = (2,1), e_2 \mid e_1)$$

$$= p(e_2 \mid x_2, \cancel{e_1})\, p(x_2 \mid e_1) = p(e_2 \mid x_2) \sum_{x_1} p(x_2 \mid x_1, \cancel{e_1})\, p(x_1 \mid e_1)$$

(*) $p(x_1 \mid e_1) \rightarrow$ می‌دانیم $e_1 = 1$ در نتیجه $p(x_1 \mid e_1)$ فقط برای خانه‌هایی که در سطر اول باشند مقدار دارد و برای بقیه خانه‌ها صفر است.

$$\underset{i \in \{1, \dots, 10\}}{p(x_1 = (1, i) \mid e_1 = 1)} = \frac{1}{10}\ , \quad \underset{i \neq 1}{p(x_1 = (i, j) \mid e_1)} = 0$$

(*) $\underset{i \in \{1, \dots, 10\}}{p(e_2 \mid x_2 = (2, i))} = 1\ , \quad \underset{i \neq 2}{p(e_2 \mid x_2 = (i, j))} = 0$

برای نرمال‌کردن باید احتمال‌های خانه‌های جدول به شرط $e_{1:2}$ را بدانیم. خانه‌هایی که در سطر دوم نیستند که ۰ است احتمالشان به خاطر $p(e_2 \mid x_2)$

برای خانه‌های سطر دوم به شکل زیر است

$$x_2 = (2,1) \rightarrow 1 \times \sum_{x_1 = (1, i)} \frac{1}{10} \times p(x_2 = (2,1) \mid x_1 = (1, i)) = \frac{1}{10} \left[\frac{9}{200} + \underbrace{\left(\frac{1}{200} + \frac{1}{4}\right)}_{p(x_2 \mid x_1)} \right]$$

با همه‌ی خانه‌های سطر اول بجز (۱،۱) به احتمال $\frac{1/2}{100} = \frac{1}{200}$

$x_1 = (1,1)$

$\frac{1}{200} \rightarrow$ پرواز شد

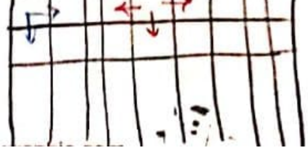
$\frac{1/2}{2} \rightarrow$ به یکی از دو خانه مجاورش برود

$$x_2 = (2,2) \text{ تا } x_2 = (2,9) \rightarrow \frac{1}{10} \left[\frac{9}{200} + \left(\frac{1}{200} + \frac{1}{4} \right) \right]$$

با به یکی از سه خانه مجاور برود

$$x_2 = (2,10) = \frac{1}{10} \left[\frac{9}{200} + \left(\frac{1}{200} + \frac{1}{4} \right) \right]$$

$$\Big[\underbrace{0, 0 \cdots 0}_{\text{سطر اول}}, \underbrace{\frac{180}{4000}, \frac{1250}{4000}, \cdots, \frac{180}{4000}}_{\text{سطر دوم}}, \underbrace{0, \cdots 0}_{\text{بقیه}} \Big] \xrightarrow{\text{نرمال‌سازی}} p(x_2 = (2,1) \mid e_{1:2}) = \frac{180}{180 \times 2 + 1250 \times 8}$$



• ب

$P(X_2=(4,4) \mid e_{1:2})$ $\boxed{e_1=1 \quad e_2=4}$

مانند بخش قبلی حل می‌کنیم

$P(e_2 \mid X_2=(4,i)) = 1 \quad i \in \{1,\dots,10\}$

$P(e_2 \mid X_2=(i,j)) = 0 \quad i \neq 4$

⟸ فقط سطر چهارم احتمال غیر صفر دارد. در دانستن wall-E در گام اول در سطر اول بوده

$\underbrace{0,\dots,0}_{\text{سطر ۱ تا ۳}}, \underbrace{\tfrac{1}{200},\dots,\tfrac{1}{200}}_{\text{سطر ۴}}, \underbrace{0,\dots,0}_{\text{بقیه}}$

$X_2=(4,1) \rightarrow \frac{1}{10} \times \left(\frac{10}{200}\right) = \frac{10}{2000}$

$\vdots$

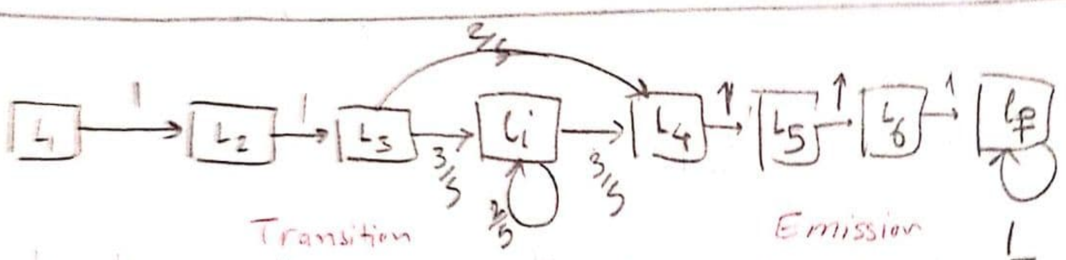
$(4,10) \rightarrow \frac{1}{10} \times \left(\frac{10}{200}\right) = \frac{10}{2000}$

فقط با پرواز از هر خانه سطر اول می‌تواند به سطر چهارم برود

در نهایت: $\boxed{P(X_2=(4,4) \mid e_{1:2}) = \frac{1}{10}}$

سوالات عملی

۱. • الف



	L_1	L_2	L_3	L_i	L_4	L_5	L_6	L_f	A	C	G	T
L_1	0	1	0	0	0	0	0	0	4/5	0	0	1/5
L_2	0	0	1	0	0	0	0	0	0	4/5	1/5	0
L_3	0	0	0	2/5	3/5	0	0	0	4/5	1/5	0	0
L_i	0	0	0	2/5	3/5	0	0	0	1/5	1/5	2/5	1/5
L_4	0	0	0	0	0	1	0	0	1	0	0	0
L_5	0	0	0	0	0	0	1	0	0	0	1/5	4/5
L_6	0	0	0	0	0	0	0	1	0	4/5	1/5	0
L_f	0	0	0	0	0	0	0	1	1/4	1/4	1/4	1/4

- ب

$$p(y_1, y_2, \dots, y_k) = \sum_{x_1} p(y_1, \dots, y_k, x_1) = \sum_{x_1} p(x_1) p(y_1 \cdots y_k | x_1) =$$

$$\sum_{x_1} p(x_1)\, p(y_1 | x_1, \cancel{y_2 \cdots y_k})\, \underbrace{p(y_2 \cdots y_k | x_1)} = \sum_{x_1} p(x_1) p(y_1|x_1) \times \sum_{x_2} p(y_2 \cdots y_k | x_1, x_2) \times p(x_2|x_1)$$

$$= \sum_{x_1} p(x_1) p(y_1|x_1) \times \sum_{x_2} p(x_2|x_1) p(y_2|x_2) \sum_{x_3} \cdots \cdots \sum_{x_k} p(x_k|x_{k-1}) p(y_k|x_k)$$

همانطور که می‌بینید می‌توان به شکل بازگشتی عبارت بالا را حساب کرد (که در آن تابع مشاهده است)

* همچنین طبق صورت سوال می‌دانیم احتمال $p(x_j = i)$ برای هر ۸ حالت برابر با $\frac{1}{8}$ است

کد این بخش در فایل likelihood.py موجود است.

- پ

ابتدا با استفاده از کد خوب است چندین بار آزمایش کنید

اگر رشته طول k داشته باشد و همچنین طولانی باشد

(در آنجا می‌توانند با کد رشته کنند)

مثلا شکل رشته + ...TTTT باشد

به احتمال $\frac{1}{8}$ می‌دانیم با l_e شروع می‌شود و می‌دانیم $p(l_i|l_e)=0$ ⇐ $i\neq e$

$$p(y_1 \cdots y_k) = \frac{1}{8} \times \left(\frac{1}{4}\right)^k$$

و همچنین اگر با هر state دیگری شروع شود در این رشته می‌بینیم

اگر l_1, l_2, l_3, l_4, l_6 باشد ⇐ $p(y_1|x_1)=0$ ⇐ $p(y_1 \cdots y_k)=0$

اگر با l_5 شروع شود $\sum p(x_1)p(y_1|x_1)\sum p(x_2|x_1)\underbrace{p(y_2|x_2)}\cdots$

$x_2=l_4$ $\quad p(y_2=T|x_2=l_4)=0$

اگر با l_1 شروع شود برای اینکه $p(y_1 \cdots y_k)$ صفر نشود تنها حالت را داریم که همیشه در l_1 بماند

$$\sum \underbrace{p(x_1)}_{\frac{1}{8}}\,\underbrace{p(y_1|x_1)}_{\frac{1}{5}} \sum \underbrace{p(x_2|x_1)}_{p(l_1|l_1)=\frac{2}{5}} \times \frac{1}{5} \cdots = \frac{1}{8} \times \left(\frac{2}{5}\right)^n \times \left(\frac{1}{5}\right)^n \to 0$$

$$\boxed{\text{Score} = \frac{\left(\frac{1}{4}\right)^k \times \frac{1}{8} + 0}{\left(\frac{1}{4}\right)^k} = \frac{1}{8}}$$

- ت

۲. دقیقاً مانند مسئله فیلترینگ عمل می‌کنیم. در هر خانه از جدول viterbi

داریم: خانه $(A, t+1)$

state = A   seq_len = t+1

$$\max_{x_1 \dots x_t} p(x_1, \dots, x_t, X_{t+1} \mid e_{1:t+1}) =$$

$$\alpha\, p(e_{t+1} \mid X_{t+1}=A) \max_{x_t} p(X_{t+1}=A \mid x_t) \max_{x_1 \dots x_{t-1}} p(x_1, \dots, x_t \mid e_{1:t})$$

(می‌دانیم برای حل این مسئله از dynamic programming کمک می‌گیریم)

در آخرین ستون چون که ستون k ام همان seq_len است

بیشینه احتمال state ها به شرط observation ها را داریم که به state

آن خانه ختم شده باشند

حال اگر ماکزیمم احتمال‌هایی که در ستون آخر داریم را بگیریم

احتمال optimal path با توجه به پارامترهای HMM آموزش شده ما را

به شرط observation ها می‌دهد

و معیار ما همان احتمال optimal path به شرط emissions را داریم

کد این بخش در فایل viterbi.py موجود است.