(مهدی رحمانی)
۹۷۳۱۷۰۱

(تمرین تحویلی سری ٤)

سوال ۱)

الف) توزیع حاشیه‌ای در حقیقت یک زیر جدول از جدول توأم می‌باشد که تعدادی از متغیرهای تصادفی آن حذف شده اند.

برای حذف یک متغیر می‌بایست عمل summation را روی تمام مقادیر آن متغیر انجام دهیم و به عبارتی باید تمام مقادیر آن متغیر را با هم جمع کنیم. به این ترتیب تعدادی از سطرهای جدول توأم با هم یکی شده و تعداد سطرها کم می‌شود. و به جدول حاصل از آن جدول توزیع حاشیه‌ای گویند.

فرمول آن نیز به صورت روبرو می‌باشد:

$$P(X_1 = x_1) = \sum_{x_2} p(X_1 = x_1, X_2 = x_2)$$

	A	B	C	P
×	+	−	−	0.2
×	+	−	+	0.05
✓	+	+	−	0.1
✓	+	+	+	0.2
×	−	−	−	0.25
×	−	−	+	0.05
✓	−	+	−	0.1
✓	−	+	+	0.05

برای مثال اگر A و C را حذف کنیم خواهیم داشت:

$$P(B) = \sum_{a,c} p(B, a, c)$$

$\Rightarrow$ P(B)

B	P
+	0.45
−	0.55

که B+ ها با علامت ✓ و B- ها با علامت × در کنار جدول مشخص شده اند

اگر B و C را هم حذف کنیم باز خواهیم داشت:

$$P(A) = \sum_{b,c} P(A, b, c)$$

$\Rightarrow$ P(A)

A	P
+	0.55
−	0.45

ب) فرمول احتمال شرطی به صورت روبرو است:

$$P(x|y) = \frac{P(x,y)}{P(y)}$$

$$* \; P(A+|B+) = \frac{P(A+, B+)}{P(B+)} = \frac{0.1 + 0.2}{0.1 + 0.2 + 0.1 + 0.05} = \frac{0.3}{0.45} = \frac{2}{3}$$

$$* \; P(B-|C+) = \frac{P(B-, C+)}{P(C+)} = \frac{0.05 + 0.05}{0.05 + 0.2 + 0.05 + 0.05} = \frac{0.1}{0.35} = \frac{2}{7}$$

$$* \; P(A+|B-, C+) = \frac{P(A+, B-, C+)}{P(B-, C+)} = \frac{0.05}{0.05 + 0.05} = \frac{1}{2}$$

ج) با توجه به اینکه $P(A|C=+)$ را می‌خواهیم، ابتدا سطرهایی را که متغیر تصادفی C مقدار + دارند را انتخاب می‌کنیم بنابراین خواهیم داشت: (همچنین متغیر B را نیز حذف می‌کنیم)

$P(C+, A)$

C	A	P
+	+	B- ۰/۰۵ + B+ ۰/۲ = ۰/۲۵
+	-	B- ۰/۰۵ + B+ ۰/۰۵ = ۰/۱

مرحله‌ی اول ←
(select the joint probabilities matching the evidence)

سپس لازم است تا مقادیر را normalize کنیم و جمع مقادیر احتمال با ۱ برابر شود:

$P(A|C=+)$

A	P
+	$\frac{۰/۲۵}{۰/۳۵} = \frac{۵}{۷}$
-	$\frac{۰/۱}{۰/۳۵} = \frac{۲}{۷}$

مرحله‌ی دوم ←
(Normalize the selection)

د) ابتدا برای $P(A|B+)$ داریم:

$$P(A|B+) \begin{cases} \text{Query}: Q = A \\ \text{Evidence}: B \\ \text{Hidden}: C \end{cases}$$

مرحله‌ی اول: ابتدا سطرهایی که با evidence ما یعنی B=+ همخوانی ندارند را حذف می‌کنیم

مرحله‌ی دوم: در این گام باید Hidden variable ها را حذف کنیم.

مرحله‌ی سوم: عملیات نرمال‌سازی را انجام دهیم.

مرحله‌ی اول →

A	B	C	P
+	+	+	۰/۲
+	+	-	۰/۱
-	+	-	۰/۱
-	+	+	۰/۰۵

مرحله‌ی دوم →

A	B	P
+	+	۰/۱ + ۰/۲ = ۰/۳
-	+	۰/۱ + ۰/۰۵ = ۰/۱۵

مرحله‌ی سوم →

$P(A|B+)$

A	P
+	$\frac{۰/۳}{۰/۴۵} = \frac{۲}{۳}$
-	$\frac{۰/۱۵}{۰/۴۵} = \frac{۱}{۳}$

حال برای $P(A|B+, C-)$ داریم:

$$P(A|B+, C-) \begin{cases} \text{Query}: A \\ \text{Evidence}: \{B, C\} \\ \text{Hidden}: \{\} = \varnothing \end{cases}$$

مراحل مانند بالاست و فقط در مرحله‌ی اول، سطرهایی که B=+, C=- باید نگه داشته شوند، بقیه حذف شوند:

مرحله‌ی اول →

A	B	C	P
+	+	-	۰/۱
-	+	-	۰/۱

مرحله‌ی دوم →
از آنجایی که متغیر hidden نداریم پس چیزی را حذف نمی‌کنیم

مرحله‌ی سوم →

$P(A|B+, C-)$

A	P
+	$\frac{۰/۱}{۰/۲} = \frac{۱}{۲}$
-	$\frac{۰/۱}{۰/۲} = \frac{۱}{۲}$

۵)

رخداد (outcome) :

در اینجا یک assignment کامل به متغیرها انجام می‌شود و P (احتمال) مربوط به آن بدست می‌آید

رویداد (event) :

مجموعه‌ای از outcome ها برای یک متغیر تصادفی خاص می‌باشد

حال برای مثال، احتمال یک رخداد داریم :

$$P(A+, B+, C-) = 0/1$$

همانطور که مشاهده می‌شود، احتمال یک رخداد دقیقاً معادل احتمال یک سطر خاص از جدول توزیع توأم می‌باشد

حال برای مثال، احتمال یک رویداد داریم :

$$P(A+) = 0/2 + 0/05 + 0/1 + 0/2 = 0/55$$

سؤال دوم)

طبق تعریف استقلال باید به ازای هر x و y در جدول، $P(x,y)=P(x)P(y)$ برقرار باشد :

$$\forall x,y \quad , \quad P(x,y) = P(x)P(y)$$

یا به عبارتی داشته باشیم : $\forall x,y \quad P(x|y) = P(x)$

اگر در یک مورد هم این تعریف برقرار نباشد و مثال نقض بیابیم، x و y مستقل از هم نیستند.

برای مثال اگر $P(x+, y-)$ را چک کنیم :

X	Y	P
+	-	۰/۲
+	+	۰/۵
-	-	۰/۱
-	+	۰/۲

$$P(x+, y-) \overset{?}{=} P(x+)\,P(y-)$$

$$\begin{cases} P(x+) = 0/2 + 0/5 = 0/7 \\ P(y-) = 0/2 + 0/1 = 0/3 \\ P(x+, y-) = 0/2 \end{cases}$$

$$\Rightarrow 0/2 \neq 0/7 \times 0/3 \Longrightarrow P(x+, y-) \neq P(x+)P(y-)$$

چون مثال نقض یافتیم، بنابراین x و y مستقل نیستند.

(سوال سوم)

(قسمت الف)

فایده و حسن یافتن استقلال بین دو متغیر این می‌باشد که مدل ما ساده‌تر و راحت‌تر می‌شود
راحت‌تر به این معنی که برای مثال فرض کنید اگر ۲ تا متغیر X و Y داشته باشیم که دامنه هر یک
هر یک ۵ عضو داشته باشد، در این حالت جدول توزیع توأم آن $5\times5=25$ خانه دارد.
اما اگر فرض استقلال را داشته باشیم فقط کافی می‌باشد که جدول توزیع حاشیه‌ای ۲ متغیر
X و Y را نگه داریم و در واقع $5+5=10$ خانه نگهداری می‌شود و احتمال هر رخدادی از روی این
۲ جدول به دست می‌آید.
بنابراین، حجم اطلاعات ذخیره‌شده نیز به این ترتیب کم می‌شود و پیچیدگی محاسبات نیز کاهش
می‌یابد.
از آنجایی که استقلال غیرشرطی در دنیای واقعی خیلی نادر و کم است از استقلال شرطی استفاده می‌کنیم

(قسمت ب)

در واقع در شبکه‌ی بیزین به جای محاسبه‌ی کل جدول توزیع توأم که حافظه نگهداری آن نمایی
می‌باشد، برای هر متغیر تصادفی فقط جدول احتمال شرطی یا همان CPT نگهداری می‌شود.
برای مثال اگر n متغیر boolean داشته باشیم، حافظه مورد نیاز برای جدول توأم، 2^n می‌باشد اما اگر
هر متغیر تصادفی در گراف شبکه بیزین حداکثر k والد داشته باشد، حافظه مورد نیاز برای نگهداری
CPT ها از مرتبه‌ی $O(n\times2^{k+1})$ می‌باشد که خیلی کمتر از 2^n می‌باشد.
بنابراین به کمک شبکه بیزین، فضای کمتری مصرف می‌شود. اگر k مستقل از n باشد
هم فضا و هم محاسبات شبکه، خطی نسبت به تعداد متغیرها می‌باشد.
همچنین حتی اگر فضای کافی داشته باشیم، به دست آوردن جدول احتمال توأم کار دشواری می‌باشد
در حالی که به دست آوردن CPT ها به دلیل وجود رابطه‌ی علت و معلول بین گره‌ها، راحت‌تر می‌باشد.
همچنین در شبکه‌ی بیزین پاسخ دادن به query ها سریع‌تر می‌باشد. در این شبکه به جای نگهداری
توزیع‌های توأم پیچیده، فقط وابستگی‌های محلی در یک گراف DAG نگهداری می‌شود و ارتباطات
global با قانون زنجیره‌ای به دست می‌آیند. که مدل زیر هم بیانگر آن است که محاسبات پیچیده کم‌تر شده.

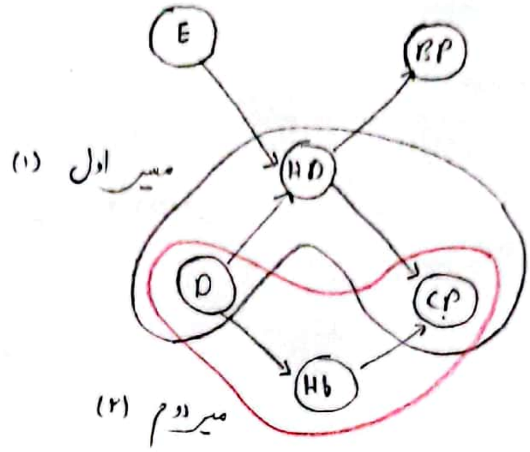
$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i \mid Parent(X_i))$$

(سؤال ۴)

(الف) دو مسیر بین D و CP داریم:

مسیر اول: نوع آن causal chains می باشد
اگر در این مسیر HD دیده نشود active است

مسیر دوم: نوع آن causal chains است
اگر در این مسیر Hb دیده نشود active است



زمانی که بین ۲ متغیر تصادفی بیش از یک مسیر باشد در صورتی که حداقل یک مسیر active وجود داشته باشد، در مورد استقلال آن ۲ متغیر ضمانتی نیست و تنها زمانی می توان گفت ۲ متغیر مستقلند که تمام مسیرهایی که بین آن دو وجود دارد، inactive باشد.

پس در واقع می توان گفت، برای آنکه داشتن درد در قفسه سینه و رژیم غذایی مستقل از هم باشند باید بیماری قلبی (HD) و همچنین Hb همزمان هر دو تا دیده شوند.

برای آنکه HD مشاهده شود یا خودش باید مستقیم مشاهده شود یا هر کدام از E و BP به دلیل زنجیری بودن مشاهده شوند.

برای مشاهده شدن Hb چون والد یا فرزند ندارد باید خودش مشاهده شود

$CP \perp\!\!\!\perp D \mid \{HD, Hb\}$

$CP \perp\!\!\!\perp D \mid \{E, Hb\}$

$CP \perp\!\!\!\perp D \mid \{BP, Hb\}$

$CP \perp\!\!\!\perp D \mid \{E, BP, Hb\}$

$CP \perp\!\!\!\perp D \mid \{E, BP, HD, Hb\}$

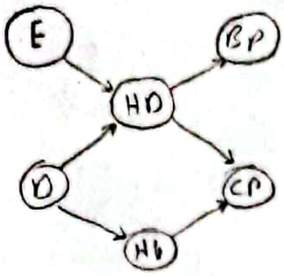
$CP \perp\!\!\!\perp D \mid \{E, HD, Hb\}$

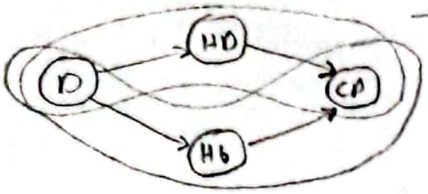
$CP \perp\!\!\!\perp D \mid \{BP, HD, Hb\}$

تمام حالات ← استقلال CP, D

Scanned with CamScanner

(ب) لازم به ذکر است که نودهایی که هاشور خورده به این معنی است که آن متغیر مشاهده شده است



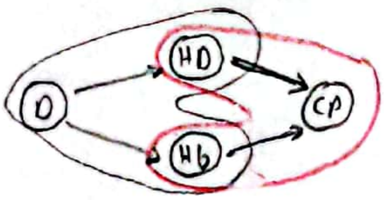


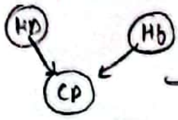
$(D \perp\!\!\!\perp CP) \leftarrow$ (Not gauranteed) مسیر از نوع → causal chains

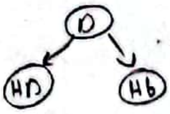
× در شکل روبرو، چون مشاهده‌ای صورت نگرفته هر دو مسیر active هستند و می‌دانیم اگر حداقل یک مسیر active باشد، استقلال دو متغیر ضمانت نمی‌شود.

$(D \perp\!\!\!\perp CP \mid HD) \leftarrow$ (Not gauranteed)

× در این صورت مسیر بالایی inactive است ولی همچنان پایینی active است بنابراین چون یک مسیر active هست ضمانتی برای استقلال D و CP نیست.




مسیر راست

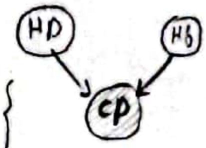

مسیر چپ

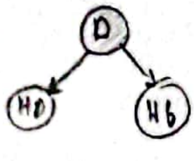
$(HD \perp\!\!\!\perp Hb) \leftarrow$ (Not gauronteed)

- مسیر راستی از نوع Common effect است و چون مشاهده‌ای نداریم، inactive می‌باشد.
- مسیر چپ از نوع Common cause است و چون مشاهده‌ای نیست active است.

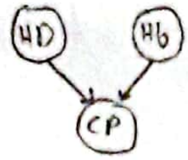
بنابراین چون یک مسیر active داریم ضمانتی برای استقلال نیست.

$(HD \perp\!\!\!\perp Hb \mid D, CP) \leftarrow$ (Not geuronteed)

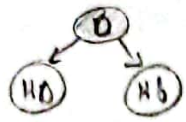

مسیر راست HD به Hb → Common effect, active


مسیر چپ HD به Hb → Common cause, inactive

چون مسیر راست active است و یک مسیر active داریم پس ضمانتی برای استقلال HD و Hb نیست.



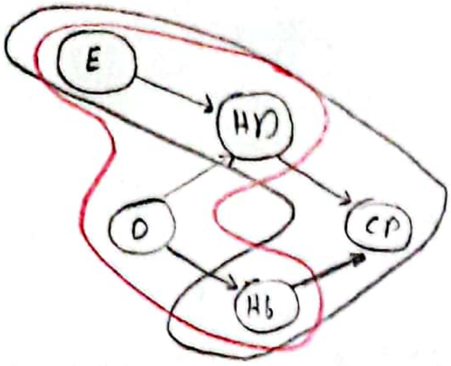
مسیر راست از Hb به HD → Common effect inactive



مسیر چپ از Hb به HD → Common cause inactive

$(HD \perp\!\!\!\perp Hb \mid D)$ ← استقلال داریم

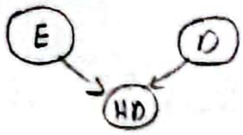
چون هر دو مسیر بین HD و Hb inactive می‌باشند، پس HD و Hb با شرط مشاهده D مستقل اند.



$(E \perp\!\!\!\perp Hb)$ ← استقلال داریم

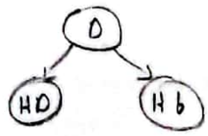
مسیرها بین E و Hb به صورت روبرو است

حال باید ۳ تایی‌های مختلف بررسی شوند



Common effect inactive

\+



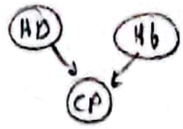
Common cause active

= مسیر چپ

این مسیر inactive است چون حداقل یک ۳ تایی inactive دارد.

causal chains active

\+

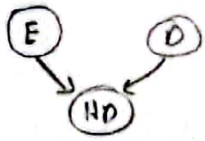


Common effect inactive

= مسیر راست " " " "

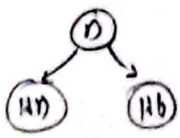
← چون همه مسیرها بین E و Hb، به صورت inactive اند پس E و Hb مستقلند.

$(E \perp\!\!\!\perp Hb \mid cp)$ ← not garonteed



Inactive

\+

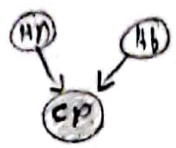


active

= مسیر چپ → این مسیر inactive است

active

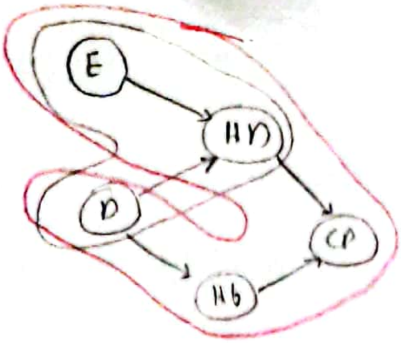
\+



active

= مسیر راست → این مسیر active است

چون حداقل یک مسیر active بین E و Hb هست، بنابراین ضمانتی برای استقلال نیست.



استقلال داریم ← $(E \perp\!\!\!\perp D)$

مسیر سه‌تایی: $E \to HD \leftarrow D$ inactive

مسیر فرعی: $E \to HD \to CP$ active + $HD \to CP \leftarrow Hb$ inactive + $D \to Hb \to CP$ active

پس این مسیر چون یک inactive است inactive است — مسیر مشترکم inactive

همه مسیرها بین E و D ← inactive ← E و D مستقل.

$E \to HD \leftarrow D$ → چون HD دیده شده پس مسیر مشترک در بالا ← active

not gaurranteed ← $(E \perp\!\!\!\perp D \mid HD)$

$E \to HD \to CP$ inactive + $HD \to CP \leftarrow Hb$ inactive + $D \to Hb \to CP$ inactive → inactive

چون حداقل یک مسیر active داریم ← استقلال ضمانت نمی‌شود

(ج)

طبق قوانین بیزنت داریم:

$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(x_i))$$

بنابراین داریم:

$$P(HD+) = P(HD+ \mid E, D)\, P(E)\, P(D)$$

$$\to P(HD+) = \sum_{e,d} P(HD+ \mid e, d)\, P(e)\, P(d)$$

$$= \underbrace{P(HD+ \mid e+, d+)\, P(e+)\, P(d+)}_{1} + \underbrace{P(HD+ \mid e+, d-)\, P(e+)\, P(d-)}_{2}$$
$$+ \underbrace{P(HD+ \mid e-, d+)\, P(e-)\, P(d+)}_{3} + \underbrace{P(HD+ \mid e-, d-)\, P(e-)\, P(d-)}_{4}$$

حال نیاز به ۳ CPT داریم:

P(E)

E	P(E)
+	۰/۷
-	۰/۳

P(D)

D	P(D)
+	۰/۲۵
-	۰/۷۵

P(HD+ | E,D)

E	D	HD+
+	+	۰/۲۵
+	-	۰/۱۵
-	+	۰/۵۵
-	-	۰/۷۵

$$P(HD+) = ۰٫۲۵ \times ۰٫۷ \times ۰٫۲۵ + ۰٫۴۵ \times ۰٫۷ \times ۰٫۷۵ + ۰٫۵۵ \times ۰٫۳ \times ۰٫۲۴$$
$$+ ۰٫۷۵ \times ۰٫۳ \times ۰٫۷۵ = \boxed{۰٫۴۹}$$

(قسمت ۱) در واقع مقدار $P(HD+ \mid BP+)$ را می‌خواهیم:

طبق بیز $\rightarrow$ $P(HD+ \mid BP+) = \dfrac{P(BP+ \mid HD+)}{P(BP+)} \times P(HD+)$

* از قسمت ج مقدار $P(HD+)$ را بدست آوردیم
* مقدار $P(BP+ \mid HD+)$ از CPT به راحتی بدست می‌آید

باید مقدار $P(BP+)$ را حساب کنیم:

$$P(BP+) = \sum_{hd} P(BP+ \mid HD) \times P(hd) = ۰٫۸۵ \times ۰٫۴۹ + ۰٫۲ \times ۰٫۵۱ = ۰٫۵۱۸۵$$

$$\Longrightarrow P(HD+ \mid BP+) = \frac{۰٫۸۵}{۰٫۵۱۸۵} \times ۰٫۴۹ = \boxed{۰٫۸۰۳}$$

(قسمت ۵) در واقع مقدار $P(HD+ \mid BP+, E+, D+)$ را می‌خواهیم.

$$P(HD+ \mid BP+, E+, D+) \overset{\text{احتمال شرطی}}{=\!=} \frac{P(HD+, BP+, E+, D+)}{P(BP+, E+, D+)} \rightarrow \text{مستقل شرطی}$$

$$\overset{\text{بیز}}{=\!=} \frac{P(HD+ \mid E+, D+)\, \cancel{P(E+)}\, \cancel{P(D+)}\, P(BP+ \mid HD+)}{P(BP+ \mid HD)\, P(HD \mid E+, D+)\, \cancel{P(D+)}\, \cancel{P(E+)}}$$

از CPT می‌خوانیم $\longleftarrow$

$$= \frac{P(HD+ \mid E+, D+)\, P(BP+ \mid HD+)}{P(BP+ \mid HD)\, P(HD \mid E+, D+)} = \frac{۰٫۲۵ \times ۰٫۸۵}{\sum_{hd} P(BP+ \mid hd)\, P(hd \mid E+, D+)}$$

$$= \frac{۰٫۲۵ \times ۰٫۸۵}{P(BP+ \mid HD+)\, P(HD+ \mid E+, D+) + P(BP+ \mid HD-)\, P(HD- \mid E+, D+)}$$

$$= \frac{۰٫۲۵ \times ۰٫۸۵}{(۰٫۸۵ \times ۰٫۲۵) + (۰٫۲ \times ۰٫۷۵)} = \boxed{۰٫۵۸۶}$$

	BP+
HD+	۰٫۸۵
HD-	۰٫۲

Scanned with CamScanner

(قسمت و)

روش (rejection-sampling) ←

در این روش نمونه‌هایی که با مشاهده‌ی ما سازگار نیست خط می‌خورند و سپس با باقی‌مانده‌ی نمونه‌ها احتمال را حساب می‌کنیم.

BP+ و CP+ مشاهدات اند ——— $P(E+ \mid BP+, CP+)$

① E+, HD-, BP-, CP- ← ✗ Reject می‌شود و با مشاهدات ما سازگار نیست و خط می‌خورد.

② E+, HD-, BP+, CP- ← ✗ 〃 〃 〃 〃

③ E+, HD+, BP+, CP- ← ✗ 〃 〃 〃 〃

④ E-, HD+, BP+, CP+ ← ✓ با مشاهدات سازگار است و پذیرفته شود

⑤ E+, HD-, BP-, CP+ ← ✗ پذیرفته نمی‌شود و خط می‌خورد

فقط نمونه‌ی ۴ ام قبول شد و در آن E- منفی است و احتمال گفته شده ۰ می‌شود

$$P(E+ \mid BP+, CP+) = \frac{0}{1} = 0$$

روش (likelihood weighting) ←

برای هر نمونه weight آن را می‌یابیم در ضمن evidence ها فیکس می‌شوند پس نمونه‌ها به شکل زیر است:

① E+, HD-, BP+, CP+ — $w_1 = 1 \times 1 \times 0.4 \times 0.2 = 0.08$

② E+, HD+, BP+, CP+ — $w_2 = 1 \times 1 \times 0.8 \times 0.9 = 0.72$

③ E-, HD+, BP+, CP+ — $w_3 = 1 \times 1 \times 0.8 \times 0.9 = 0.72$

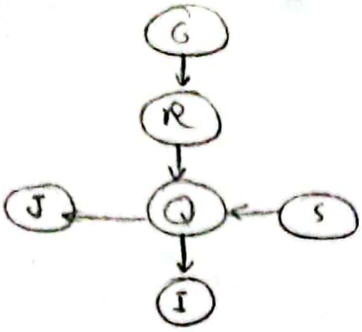
④ E+, HD-, BP+, CP+ — $w_4 = 1 \times 1 \times 0.4 \times 0.2 = 0.08$

⑤ E-, HD+, BP+, CP+ — $w_5 = 1 \times 1 \times 0.8 \times 0.9 = 0.72$

$$\rightarrow \hat{P}(E+ \mid BP+, CP+) = \frac{2 \times 0.08 + 0.72}{2 \times 0.08 + 3 \times 0.72} = \frac{0.88}{2.32} = \boxed{0.379}$$

روش likelihood موثرتر می‌باشد چرا که نمونه‌هایی که پیدا کنیم در روش اول چون احتمال رخداد نمونه‌ها با توجه به مشاهدات پایین بود، اکثراً reject شدند. چون BP+ و CP+ به صورت همزمان بعید است رخ دهد. اکثر نمونه‌ها ناسازگار با مشاهدات می‌شوند. در روش دوم که evidence ها را فیکس کردیم بهتر عمل کرد و با در نظر گرفتن وزن احتمال خواسته شده حساب شد.

(سؤال ۵)



می‌توان گراف را به صورت روبرو کشید :

R, Q, S : hidden
G : Query
i, j : evidence

هم جنبه دادیم

$$P(G)\ P(R|G)\ P(Q|R,S)\ P(S)\ P(j|Q)\ P(i|Q)$$

$$P(G|i,j) \underset{\beta}{\propto} P(G,i,j) = \sum_{r,q,s} P(G)P(r|G)P(q|r,s)P(s)P(j|q)P(i|q)$$

باید یکی یکی متغیرهای hidden را حذف کنیم .

$$= \sum_{r,q} P(G)\,P(r|G)\,P(j|q)\,P(i|q) \sum_{s} P(s)\,P(q|r,s)$$

متغیر s را با کمک summation روی آن حذف کنیم و باید به جای $\sum_s$ عبارت زیر را قرار دهیم:

$$f_1(r,q) = \sum_s P(s)\,P(q|r,s)$$

بنابراین خواهیم داشت :

$$P(G,i,j) = \sum_{r,q} P(G)\,P(r|G)\,P(j|q)\,P(i|q)\,f_1(r,q)$$

$$= \sum_q P(G)\,P(j|q)\,P(i|q) \sum_r P(r|G)\,f_1(r,q)$$

$f_2(G,q) = \sum_r P(r|G) f_1(r,q)$

$$\sum_q P(G)\,P(j|q)\,P(i|q)\,f_2(G,q)$$

$$= P(G) \sum_q P(j|q)\,P(i|q)\,f_2(G,q)$$

$f_3(j,i,G) = \sum_q P(j|q)P(i|q)f_2(G,q)$

$$P(G)\,f_3(j,i,G)$$

حال که حذف hidden ها شدن باید نرمالایز کنیم :

$$P(G|i,j) = \frac{P(G,i,j)}{P(i,j)} = \frac{P(G)\,f_3(i,j,G)}{\sum_g P(g,i,j)} = \frac{P(G)\,f_3(i,j,G)}{\sum_g P(g)\,f_3(i,j,g)}$$