دانشگاه صنعتی شریف
دانشکده‌ی مهندسی کامپیوتر

هوش مصنوعی

بهار ۱۴۰۵

استاد: ماری‌اوریاد
گردآورندگان: ریحانه ابراهیم زاده، پارسا پورنعمت، فاطیما تیمارچی، باران حسینی، آریانا زال نژاد، امید شفاعت

تمرین سوم — **شبکه‌های بیزین و مدل‌های مخفی مارکوف** — مهلت ارسال: ۸ خرداد

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص‌شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمارین تا سقف ۴ روز و در مجموع ۱۰ روز، وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال‌شده پذیرفته نخواهند بود. همچنین، به ازای هر ساعت تأخیر غیر مجاز ۰.۵ درصد از نمره تمرین به صورت ساعتی کسر خواهد شد.
- هم‌کاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.
- پاسخنامه سوالات را در سامانه کوئرا و به فرمت AI_HW3_<STUDENT_ID> بارگزاری کنید.

سوالات نظری (۱۰۰ نمره)

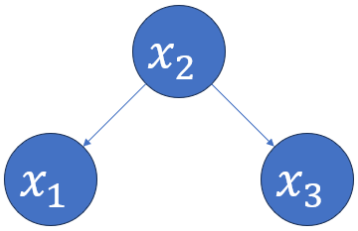
۱. (۲۰ نمره) درستی و نادرستی عبارات زیر را با ذکر دلیل مشخص کنید.

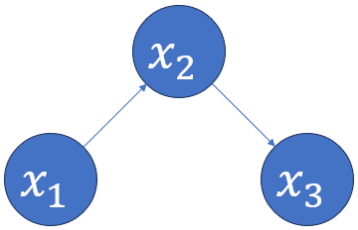
(آ) در یک شبکه‌ی بیزی، همواره هر گره به شرط دیدن والدین خود، از تمام گره‌های غیرفرزند خود مستقل شرطی است.

(ب) در ساختار common effect به صورت $X \to C \leftarrow Y$، اگر ما گره A را که یکی از اجداد گره C است مشاهده کنیم، مسیر بین X و Y فعال می‌شود.

(ج) الگوریتم d-separation شرط لازم و کافی‌ای برای استقلال است، به این معنی که اگر دو متغیر X و Y طبق این‌الگوریتم از هم جدا نشوند (یعنی تمام مسیرهای میان آن‌ها غیرفعال نباشد)، آنگاه این دو متغیر قطعا مستقل نمی‌باشند.

(د) هر توزیع توأم $P(X_1, X_2, X_3)$ که با استفاده از توزیع‌های احتمال شرطی (CPD) در شبکه بیزی سمت چپ قابل بیان باشد، با اعمال تغییرات مناسب روی CPD ها در شبکه بیزی سمت راست نیز قابل بیان است.





(ه) در یک شبکه‌ی بیز با دامنه‌ی مقادیر باینری، پیچیدگی زمانی الگوریتم Inference by Enumeration برای محاسبه‌ی $P(X|\mathbf{e})$ همواره $O(2^n)$ است که در آن n تعداد کل متغیرهای شبکه می‌باشد.

(و) در الگوریتم Variable Elimination، اگر دو فاکتور هیچ متغیر مشترکی نداشته باشند، join کردن آن‌ها بی‌فایده است.

(ز) حذف یک متغیر در Variable Elimination همیشه باعث کوچک‌تر شدن فاکتورها می‌شود.

(ح) در هنگام انجام Inference در یک شبکهٔ بیزی می‌توان همهٔ متغیرهایی را که از اجداد متغیرهای Query و Evidence نیستند، در نظر نگرفت.

(ط) اندازهٔ جداول احتمالات شرطی در یک شبکهٔ بیزی با N متغیر، با دامنه‌هایی حداکثر D عضوی، به شرطی که هر رأس حداکثر K پدر داشته باشد، دارای کران بالای $N \times D^{K+1}$ است.

(ی) Inference by Enumeration و Variable Elimination هر دو همواره مقادیر احتمال پسین یکسانی را تولید می‌کنند، اما حذف متغیر معمولاً سریع‌تر است، زیرا متغیرها را در اولین فرصت marginalize می‌کند، نه پس از محاسبهٔ کامل توزیع مشترک.

حل.

(آ) **درست.** این گزاره مستقیماً تعریف Local Markov Assumption (یا Local Markov Property) را در شبکه‌های بیزی بیان می‌کند. مطابق این ویژگی بنیادی و قوانین d-separation، با مشاهده‌ی پدرهای یک گره، تمام مسیرهای فعال میان آن گره و هر یک از گره‌هایی که از فرزندانش نیستند بسته می‌شود. در نتیجه، توزیع احتمال آن گره فقط به پدرهایش وابسته است و استقلال شرطی آن از تمام گره‌های دیگر که از فرزندانش نیستند تضمین می‌شود.

(ب) **نادرست.** در یک common effect به فرم $X \to C \leftarrow Y$ طبق قوانین d-separation، مسیر بین X و Y در ابتدا، وقتی هیچ evidenceای مشاهده نشده باشد، بسته است. برای فعال شدن این مسیر، باید خود C، یا دست‌کم یکی از فرزندان آن مشاهده شود. مشاهده‌ی یک جد برای C اطلاعاتی درباره‌ی C فراهم نمی‌کند به شکلی که میان X و Y هم‌بستگی ایجاد شود. بنابراین، با مشاهده‌ی یک جد، مسیر همچنان بسته می‌ماند.

(ج) **نادرست.** الگوریتم d-separation استقلال شرطی را فقط بر پایه‌ی ساختار گرافی شبکه بررسی می‌کند، یعنی یک شرط *کافی* برای استقلال به دست می‌دهد. اگر دو گره d-separated باشند، استقلال آن‌ها تضمین می‌شود. با این حال، این شرط *لازم* نیست، زیرا ممکن است دو متغیر از نظر ساختاری d-separated نباشند، اما همچنان به دلیل پارامترهای عددی خاص در CPT هایشان کاملاً مستقل باشند، چون این پارامترها اثر یکدیگر را خنثی می‌کنند. به این حالت نقض فرض faithfulness هم گفته می‌شود که یعنی استقلال‌های موجود در توزیع دقیقاً با استقلال‌های ساختاری گراف منطبق نیستند.

(د) **درست.** توزیع مشترک شبکه سمت چپ به صورت زیر تجزیه می‌شود:

$$P(X_1, X_2, X_3) = P(X_2) \cdot P(X_1 \mid X_2) \cdot P(X_3 \mid X_2)$$

با استفاده از قضیه‌ی بیز، می‌توانیم دو جمله‌ی اول را بازنویسی کنیم:

$$P(X_2) \cdot P(X_1 \mid X_2) = P(X_1, X_2) = P(X_1) \cdot P(X_2 \mid X_1)$$

با جایگذاری این رابطه در تجزیه‌ی اولیه، به دست می‌آوریم:

$$P(X_1, X_2, X_3) = P(X_1) \cdot P(X_2 \mid X_1) \cdot P(X_3 \mid X_2)$$

که دقیقا CPD های شبکه سمت راست است.

(ه) **نادرست.** پیچیدگی زمانی Inference by Enumeration همیشه $O(2^n)$ نیست. عامل نمایی به تعداد متغیرهای پنهان، یعنی k، وابسته است، نه به تعداد کل متغیرها، یعنی n. برای محاسبه‌ی مجموع، الگوریتم تمام 2^k انتساب ممکن به متغیرهای دودوییِ پنهان را پیمایش می‌کند. برای هر انتساب، ارزیابی احتمال مشترک نیازمند ضرب کردن احتمال‌های شرطیِ مربوط به تمام n گره است، که $O(n)$ عمل نیاز دارد. بنابراین، پیچیدگی زمانی دقیق برابر با $O(n \cdot 2^k)$ است. در نتیجه، مشاهده‌ی متغیرهای evidence بیشتر باعث کاهش k می‌شود، و این موضوع زمان اجرا را به‌صورت نمایی کاهش می‌دهد.

(و) **درست.** در الگوریتم Variable Elimination، هدف از join کردن (ضرب کردن) عامل‌ها این است که تمام وابستگی‌های مربوط به یک متغیر مشخص در کنار هم جمع شوند تا آن متغیر بتواند حذف شود. اگر دو عامل هیچ متغیر مشترکی در دامنه‌ی خود نداشته باشند، حاصل‌ضرب نقطه‌به‌نقطه‌ی آن‌ها یک ضرب دکارتی ایجاد می‌کند و یک عامل جدید بسیار بزرگ‌تر می‌سازد که اندازه‌اش برابر با حاصل‌ضرب اندازه‌های دو عامل اولیه است. این کار بدون آن‌که الگوریتم را به حذف یک متغیر نزدیک‌تر کند، به شدت هزینه‌ی حافظه و محاسبه را افزایش می‌دهد. بنابراین، عامل‌ها فقط زمانی باید join شوند که متغیر مشترکِ مربوط به حذفِ جاری را داشته باشند.

(ز) **نادرست.** فرآیند حذف یک متغیر از دو گام متمایز تشکیل می‌شود: نخست، تمام عامل‌هایی که شامل متغیر هدف هستند با هم join می‌شوند (در هم ضرب می‌شوند)، و سپس آن متغیر از عامل حاصل جمع بسته می‌شود. گرچه گام دوم (حاشیه‌گیری) بُعد عامل را یک واحد کاهش می‌دهد، اما گام اول (ضرب) یک عامل میانی می‌سازد که دامنه‌ی آن برابر با اجتماع تمام متغیرهای موجود در عامل‌های ضرب‌شده است. بسته به میزان اتصال در گراف و ترتیب حذف، این عامل میانی ممکن است به صورت نمایی از هر یک از عامل‌های اولیه بزرگ‌تر باشد.

(ح) **درست.** متغیرهایی که نه متغیر پرس‌وجو هستند و نه متغیر evidence و نیز هیچ متغیر پرس‌وجو یا شاهدی در میان فرزندانشان ندارند، barren nodes نامیده می‌شوند. چون مجموع احتمال‌ها روی تمام حالت‌های ممکن یک گره، با شرط پدرهایش، دقیقاً برابر با ۱ است ($\sum_x P(x \mid \text{parents}) = ۱$)، حاشیه‌گیری روی یک barren node فقط توزیع مشترک را در ضریب ۱ ضرب می‌کند. در نتیجه، این گره‌ها هیچ اثر ریاضی‌ای بر احتمال‌های پسین نهاییِ پرس‌وجو ندارند. برای افزایش کارایی، می‌توان و باید آن‌ها را پیش از اجرای استنتاج به طور کامل از شبکه هرس کرد.

(ط) **درست.** برای به‌دست‌آوردن کران بالای اندازه‌ی CPT، یک گره از شبکه را در نظر بگیرید. اگر آن گره حداکثر K پدر داشته باشد و هر متغیر بتواند حداکثر D مقدار متمایز بگیرد، آن‌گاه حداکثر D^K ترکیب ممکن برای انتساب پدرها وجود دارد. برای هر یک از این انتساب‌ها، CPT باید یک توزیع احتمال روی D حالت ممکن خود گره مشخص کند، که به D ورودی نیاز دارد (یا $D - ۱$ ورودی مستقل، اما D کران بالای ذخیره‌سازی است). پس بیشینه‌ی اندازه‌ی یک CPT برابر است با $D \times D^K = D^{K+۱}$. از آن‌جا که N گره در شبکه وجود دارد، اندازه‌ی کل تمام جدول‌ها با $N \times D^{K+۱}$ از بالا کران می‌خورد.

(ی) **درست.** هر دو الگوریتم Inference by Enumeration و Variable Elimination از الگوریتم‌های استنتاج دقیق هستند. چون هر دو دقیقاً همان جمع‌ها و ضرب‌های ریاضیِ توزیع مشترک را ارزیابی می‌کنند، تضمین می‌شود که احتمال‌های پسین یکسانی تولید کنند. تفاوت اصلی در کارایی است: Variable Elimination به‌طور گسترده از خاصیت توزیعی ضرب نسبت به جمع استفاده می‌کند ($\sum a \cdot b \cdot c = a \cdot \sum b \cdot c$) تا جمع‌ها را تا جای ممکن به درون ببرد. این کار از تولید کامل توزیع مشترک نمایی جلوگیری می‌کند و در عمل باعث می‌شود محاسبه بسیار سریع‌تر انجام شود.

۲. (۲۰ نمره) یک موزه دارای سیستم امنیتی هوشمند است. در هر شب، وضعیت واقعی موزه (متغیر مخفی) یکی از دو حالت زیر است:

$$X_t \in \{N, T\}$$

N: وضعیت عادی (Normal) - T: وجود دزد در موزه (Thief)

مشاهدات سیستم امنیتی در هر شب (خروجی قابل مشاهده) یکی از مقادیر زیر است:

$$E_t \in \{\text{silent}, \text{motion}, \text{alarm}\}$$

silent : بدون هشدار - motion : سنسور حرکتی - alarm : آژیر اصلی

پارامترهای مدل

$$P(X_۱ = T) = ۰/۳, \quad P(X_۱ = N) = ۰/۷$$
$$P(X_t = T \mid X_{t-۱} = T) = ۰/۸, \quad P(X_t = N \mid X_{t-۱} = T) = ۰/۲$$
$$P(X_t = T \mid X_{t-۱} = N) = ۰/۱۵, \quad P(X_t = N \mid X_{t-۱} = N) = ۰/۸۵$$

Observation E_t	$P(E_t \mid X_t = T)$	$P(E_t \mid X_t = N)$
silent	۰.۱	۰.۷
motion	۰.۳	۰.۲
alarm	۰.۶	۰.۱

فرض کنید در دو شب متوالی داریم:

$$E_۱ = \text{motion}, \quad E_۲ = \text{alarm}$$

(آ) تمام حالات ممکن $(X_۱, X_۲)$ را نوشته و برای هر کدام مقدار زیر را محاسبه کنید:

$$P(X_۱, E_۱, X_۲, E_۲)$$

(ب) احتمال زیر را محاسبه کنید:

$$P(X_۲ = T \mid E_۱ = \text{motion}, E_۲ = \text{alarm})$$

حل.

برای سهولت، از نمادگذاری زیر استفاده می‌کنیم:

$$E_۱ = \text{motion}, \quad E_۲ = \text{alarm}$$

(آ) حالات ممکن $(X_۱, X_۲)$:

$$(T,T),\ (T,N),\ (N,T),\ (N,N)$$

فرمول:

$$P(X_۱, E_۱, X_۲, E_۲) = P(X_۱)\, P(E_۱ \mid X_۱)\, P(X_۲ \mid X_۱)\, P(E_۲ \mid X_۲)$$

۱) $(X_۱ = T, X_۲ = T)$

$$P(X_۱ = T) = ۰/۳$$
$$P(E_۱ = motion \mid X_۱ = T) = ۰/۳$$
$$P(X_۲ = T \mid X_۱ = T) = ۰/۸$$
$$P(E_۲ = alarm \mid X_۲ = T) = ۰/۶$$
$$P(T, motion, T, alarm) = ۰/۳ \times ۰/۳ \times ۰/۸ \times ۰/۶ = ۰/۰۴۳۲$$

۲) $(X_۱ = T, X_۲ = N)$

$$P(X_۲ = N \mid X_۱ = T) = ۰/۲$$
$$P(E_۲ = alarm \mid X_۲ = N) = ۰/۱$$
$$P(T, motion, N, alarm) = ۰/۳ \times ۰/۳ \times ۰/۲ \times ۰/۱ = ۰/۰۰۱۸$$

۳) $(X_۱ = N, X_۲ = T)$

$$P(X_۱ = N) = ۰/۷$$
$$P(E_۱ = motion \mid X_۱ = N) = ۰/۲$$
$$P(X_۲ = T \mid X_۱ = N) = ۰/۱۵$$
$$P(E_۲ = alarm \mid X_۲ = T) = ۰/۶$$
$$P(N, motion, T, alarm) = ۰/۷ \times ۰/۲ \times ۰/۱۵ \times ۰/۶ = ۰/۰۱۲۶$$

۴) $(X_۱ = N, X_۲ = N)$

$$P(X_۲ = N \mid X_۱ = N) = ۰/۸۵$$
$$P(E_۲ = alarm \mid X_۲ = N) = ۰/۱$$
$$P(N, motion, N, alarm) = ۰/۷ \times ۰/۲ \times ۰/۸۵ \times ۰/۱ = ۰/۰۱۱۹$$

(ب) ابتدا احتمال کل مشاهدات را حساب می‌کنیم:

$$P(E_۱, E_۲) = \sum_{X_۱, X_۲} P(X_۱, E_۱, X_۲, E_۲)$$
$$P(E_۱, E_۲) = ۰/۰۴۳۲ + ۰/۰۰۱۸ + ۰/۰۱۲۶ + ۰/۰۱۱۹ = ۰/۰۶۹۵$$

حالا فقط مسیرهایی را در نظر می‌گیریم که در آن‌ها $X_۲ = T$ است:

$$P(E_۱, E_۲, X_۲ = T) = P(T, motion, T, alarm) + P(N, motion, T, alarm)$$
$$= ۰/۰۴۳۲ + ۰/۰۱۲۶ = ۰/۰۵۵۸$$

در نتیجه:

$$P(X_۲ = T \mid E_۱, E_۲) = \frac{P(E_۱, E_۲, X_۲ = T)}{P(E_۱, E_۲)} = \frac{۰/۰۵۵۸}{۰/۰۶۹۵} \approx ۰/۸۰۳$$

بنابراین:

$$P(X_۲ = T \mid E_۱ = motion, E_۲ = alarm) \approx ۰/۸۰۳$$

۳. (۱۵ نمره) یک تیم فوتبال در هر مسابقه می‌تواند یکی از سه نتیجه را کسب کند: W = برد، D = مساوی، L = باخت

تحلیلگر تیم متوجه شده است که نتیجه مسابقه بعد فقط به نتیجه مسابقه فعلی بستگی دارد. ماتریس احتمال انتقال P به صورت زیر است:

$$P = \begin{bmatrix} ۰/۵ & ۰/۳ & ۰/۲ \\ ۰/۴ & ۰/۴ & ۰/۲ \\ ۰/۲ & ۰/۳ & ۰/۵ \end{bmatrix}$$

ترتیب ستون‌ها از چپ به راست W, D, L است. یعنی:

- بعد از برد (W) ← با احتمال ۰/۵ دوباره برد، ۰/۳ مساوی، ۰/۲ باخت
- بعد از مساوی (D) ← با احتمال ۰/۴ برد، ۰/۴ مساوی، ۰/۲ باخت

- بعد از باخت (L) ← با احتمال ۰/۲ برد، ۰/۳ مساوی، ۰/۵ باخت

(آ) نمودار حالت (State Diagram) این زنجیره مارکوف را رسم کنید.

(ب) اگر توزیع اولیه در مسابقه اول به صورت زیر باشد:

$$P(X_۱ = W) = ۰/۴, \quad P(X_۱ = D) = ۰/۴, \quad P(X_۱ = L) = ۰/۲$$

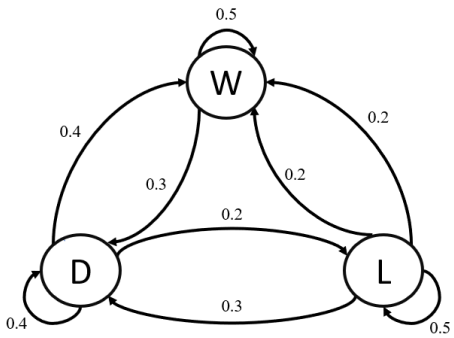
آنگاه محاسبه کنید:

$$P(X_۱ = L,\, X_۲ = D,\, X_۳ = W) = ?$$

(ج) با فرض توزیع اولیه بالا، احتمال اینکه تیم در مسابقه سوم ببرد را محاسبه کنید.

حل.

(آ) نمودار حالت:



(ب) با استفاده از قانون زنجیره‌ای و خاصیت مارکوف داریم:

$$P = P(X_۱ = L) \times P(X_۲ = D \mid X_۱ = L) \times P(X_۳ = W \mid X_۲ = D)$$

از ماتریس P:

$$P(X_۱ = L) = ۰/۲$$
$$P(X_۲ = D \mid X_۱ = L) = P_{L,D} = ۰/۳$$
$$P(X_۳ = W \mid X_۲ = D) = P_{D,W} = ۰/۴$$

بنابراین:

$$P = ۰/۲ \times ۰/۳ \times ۰/۴ = ۰/۰۲۴$$

$$\boxed{۰/۰۲۴}$$

(ج) ابتدا $P(X_۲)$ را محاسبه می‌کنیم:

$$\begin{aligned}
P(X_۲ = W) &= P(X_۱ = W) \times P(X_۲ = W \mid X_۱ = W) \\
&\quad + P(X_۱ = D) \times P(X_۲ = W \mid X_۱ = D) \\
&\quad + P(X_۱ = L) \times P(X_۲ = W \mid X_۱ = L) \\
&= ۰/۴ \times ۰/۵ + ۰/۴ \times ۰/۴ + ۰/۲ \times ۰/۲ \\
&= ۰/۲۰ + ۰/۱۶ + ۰/۰۴ = ۰/۴۰
\end{aligned}$$

$$\begin{aligned}
P(X_۲ = D) &= P(X_۱ = W) \times P(X_۲ = D \mid X_۱ = W) \\
&\quad + P(X_۱ = D) \times P(X_۲ = D \mid X_۱ = D) \\
&\quad + P(X_۱ = L) \times P(X_۲ = D \mid X_۱ = L) \\
&= ۰/۴ \times ۰/۳ + ۰/۴ \times ۰/۴ + ۰/۲ \times ۰/۳ \\
&= ۰/۱۲ + ۰/۱۶ + ۰/۰۶ = ۰/۳۴
\end{aligned}$$

$$\begin{aligned}
P(X_۲ = L) &= P(X_۱ = W) \times P(X_۲ = L \mid X_۱ = W) \\
&\quad + P(X_۱ = D) \times P(X_۲ = L \mid X_۱ = D) \\
&\quad + P(X_۱ = L) \times P(X_۲ = L \mid X_۱ = L) \\
&= ۰/۴ \times ۰/۲ + ۰/۴ \times ۰/۲ + ۰/۲ \times ۰/۵ \\
&= ۰/۰۸ + ۰/۰۸ + ۰/۱۰ = ۰/۲۶
\end{aligned}$$

حال $P(X_۳ = W)$:

$$\begin{aligned}
P(X_۳ = W) &= P(X_۲ = W) \times P(X_۳ = W \mid X_۲ = W) \\
&\quad + P(X_۲ = D) \times P(X_۳ = W \mid X_۲ = D) \\
&\quad + P(X_۲ = L) \times P(X_۳ = W \mid X_۲ = L) \\
&= ۰/۴۰ \times ۰/۵ + ۰/۳۴ \times ۰/۴ + ۰/۲۶ \times ۰/۲ \\
&= ۰/۲۰ + ۰/۱۳۶ + ۰/۰۵۲ = ۰/۳۸۸
\end{aligned}$$

$$\boxed{۰/۳۸۸}$$

۴. (۲۰ نمره) شبکه بیزی زیر را در نظر بگیرید. درستی یا نادرستی هر یک از گزاره‌های استقلال شرطی زیر را با ذکر دلیل مشخص کنید.

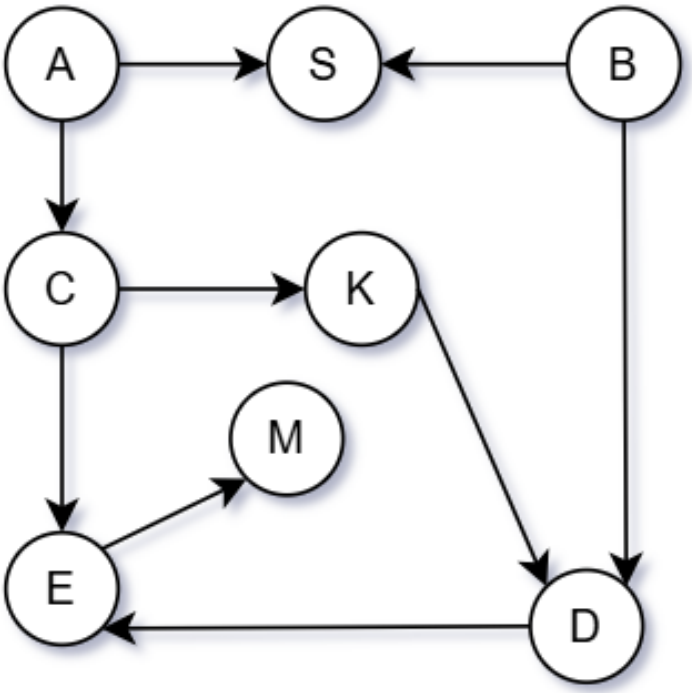
(آ) $A \perp B$

(ب) $A \perp B \mid M$

(ج) $A \perp B \mid \{M, K\}$

(د) فرض کنیم گره‌های S و M مشاهده شده باشند. آیا میتوان با اضافه کردن گره‌های دیگری به مجموعه شواهد، متغیرهای A و B را از یکدیگر کاملاً مستقل کرد؟

حل.



(آ) درست
بین A و B سه مسیر وجود دارد که هر سه به دلیل وجود حداقل یک ساختار Common Effect مشاهده نشده، مسدود هستند.

i. مسیر $A \to S \leftarrow B$: مسیر در S مسدود شده است.

ii. مسیر $A \to C \to E \leftarrow D \leftarrow B$: مسیر در E مسدود شده است.

iii. مسیر $A \to C \to K \to D \leftarrow B$: مسیر در D مسدود شده است.

(ب) در مورد استقلال چیزی نمی‌توان گفت.
مشاهده گره M (نواده گره E) باعث فعال شدن گره E و در نتیجه مسیر دوم میشود و استقلال را از بین می‌برد.

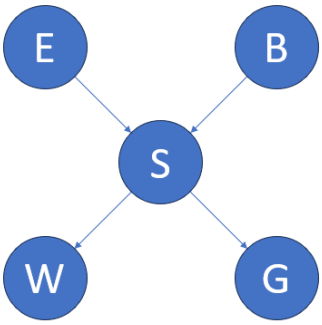
(ج) در مورد استقلال چیزی نمی‌توان گفت.
علیرغم اینکه K مسیر سوم را مسدود می‌کند اما تاثیری روی مسیر دوم ندارد و مسیر دوم توسط E فعال شده (چون M نواده آن است). بنابراین استقلالی نخواهیم داشت.

(د) خیر
برای اینکه دو گره مستقل باشند باید تمامی مسیرهای میان آن‌ها مسدود باشد، در اینجا با مشاهده گره S مسیر اول فعال می‌شود. حتی اگر بخواهیم مشکلی که M ایجاد کرده را با گره‌های D و C در مسیرهای دوم و سوم درست کنیم باز هم A و B نمی‌توانند لزوما مستقل باشند.

۵. (۱۵ نمره) در شبکه‌ی بیز با توزیع‌های داده‌شده‌ی زیر، مقادیر خواسته شده را محاسبه کنید.

(آ) $P(W \mid S, G)$

(ب) $P(E \mid G)$



S	B	E	$P(S \mid B, E)$
$+s$	$+b$	$+e$	۰.۹
$+s$	$+b$	$-e$	۰.۸
$+s$	$-b$	$+e$	۰.۲
$+s$	$-b$	$-e$	۰
$-s$	$+b$	$+e$	۰.۱
$-s$	$+b$	$-e$	۰.۲
$-s$	$-b$	$+e$	۰.۸
$-s$	$-b$	$-e$	۱

B	$P(B)$
$+b$	۰.۱
$-b$	۰.۹

E	$P(E)$
$+e$	۰.۱
$-e$	۰.۹

G	S	$P(G \mid S)$
$+g$	$+s$	۰.۵
$+g$	$-s$	۰.۴
$-g$	$+s$	۰.۵
$-g$	$-s$	۰.۶

W	S	$P(W \mid S)$
$+w$	$+s$	۰.۸
$+w$	$-s$	۰.۳
$-w$	$+s$	۰.۲
$-w$	$-s$	۰.۷

حل.

(آ) با دانستن S، W از G مستقل می‌شود. پس:

$$P(W|S,G) = P(W|S)$$

$$P(W = ۰ \mid S = ۰) = ۰/۷, \quad P(W = ۱ \mid S = ۰) = ۰/۳$$
$$P(W = ۰ \mid S = ۱) = ۰/۲, \quad P(W = ۱ \mid S = ۱) = ۰/۸$$

(ب)

$$P(E|G) = \frac{P(E,G)}{P(G)}$$

$$\begin{aligned}
P(E,G) &= \sum_{b,s} P(E,G,s,b) = \sum_{b,s} P(E,b)\, P(s \mid E,b)\, P(G \mid E,b,s) \\
&= \sum_{b,s} P(E)\, P(b)\, P(s \mid E,b)\, P(G \mid s) \\
&= P(E)\, P(+b)\, P(+s \mid E,+b)\, P(G \mid +s) + P(E)\, P(+b)\, P(-s \mid E,+b)\, P(G \mid -s) \\
&+ P(E)\, P(-b)\, P(+s \mid E,-b)\, P(G \mid +s) + P(E)\, P(-b)\, P(-s \mid E,-b)\, P(G \mid -s)
\end{aligned}$$

$$
\begin{aligned}
&P(+e,+g)\\
&\quad= P(+e)\,P(+b)\,P(+s \mid +e,+b)\,P(+g \mid +s) + P(+e)\,P(+b)\,P(-s \mid +e,+b)\,P(+g \mid -s)\\
&\quad+ P(+e)\,P(-b)\,P(+s \mid +e,-b)\,P(+g \mid +s) + P(+e)\,P(-b)\,P(-s \mid +e,-b)\,P(+g \mid -s)\\
&\quad= ۰/۱ \times ۰/۱ \times (۰/۹ \times ۰/۵ + ۰/۱ \times ۰/۴)\\
&\quad+ ۰/۱ \times ۰/۹ \times (۰/۲ \times ۰/۵ + ۰/۸ \times ۰/۴)\\
&\quad= ۰/۰۴۲۷
\end{aligned}
$$

$$
\begin{aligned}
&P(+e,-g)\\
&\quad= P(+e)\,P(+b)\,P(+s \mid +e,+b)\,P(-g \mid +s) + P(+e)\,P(+b)\,P(-s \mid +e,+b)\,P(-g \mid -s)\\
&\quad+ P(+e)\,P(-b)\,P(+s \mid +e,-b)\,P(-g \mid +s) + P(+e)\,P(-b)\,P(-s \mid +e,-b)\,P(-g \mid -s)\\
&\quad= ۰/۱ \times ۰/۱ \times (۰/۹ \times ۰/۵ + ۰/۱ \times ۰/۶)\\
&\quad+ ۰/۱ \times ۰/۹ \times (۰/۲ \times ۰/۵ + ۰/۸ \times ۰/۶)\\
&\quad= ۰/۰۵۷۳
\end{aligned}
$$

$$
\begin{aligned}
&P(-e,-g)\\
&\quad= P(-e)\,P(+b)\,P(+s \mid -e,+b)\,P(-g \mid +s) + P(-e)\,P(+b)\,P(-s \mid -e,+b)\,P(-g \mid -s)\\
&\quad+ P(-e)\,P(-b)\,P(+s \mid -e,-b)\,P(-g \mid +s) + P(-e)\,P(-b)\,P(-s \mid -e,-b)\,P(-g \mid -s)\\
&\quad= ۰/۹ \times ۰/۱ \times (۰/۸ \times ۰/۵ + ۰/۲ \times ۰/۶)\\
&\quad+ ۰/۹ \times ۰/۹ \times (۰ \times ۰/۵ + ۱ \times ۰/۶)\\
&\quad= ۰/۵۳۲۸
\end{aligned}
$$

$$
\begin{aligned}
&P(-e,+g)\\
&\quad= P(-e)\,P(+b)\,P(+s \mid -e,+b)\,P(+g \mid +s) + P(-e)\,P(+b)\,P(-s \mid -e,+b)\,P(+g \mid -s)\\
&\quad+ P(-e)\,P(-b)\,P(+s \mid -e,-b)\,P(+g \mid +s) + P(-e)\,P(-b)\,P(-s \mid -e,-b)\,P(+g \mid -s)\\
&\quad= ۰/۹ \times ۰/۱ \times (۰/۸ \times ۰/۵ + ۰/۲ \times ۰/۴)\\
&\quad+ ۰/۹ \times ۰/۹ \times (۰ \times ۰/۵ + ۱ \times ۰/۴)\\
&\quad= ۰/۳۶۷۲
\end{aligned}
$$

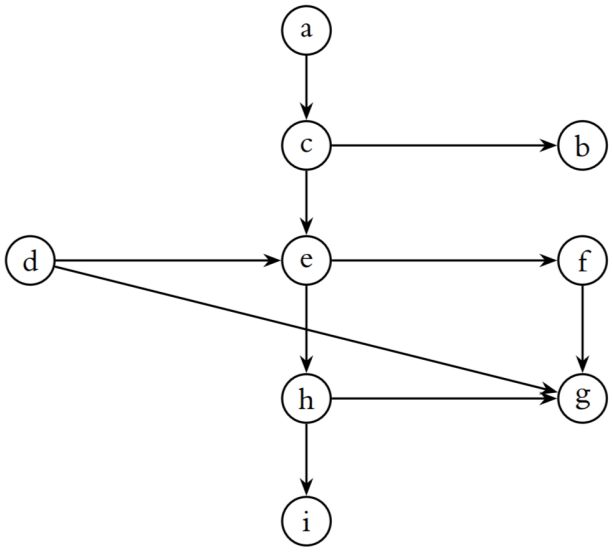
با توجه به مقایر به دست آمده می‌فهمیم که:

$$
\begin{aligned}
P(+g) &= P(+g,-e) + P(+g,+e) = ۰/۳۶۷۲ + ۰/۰۴۲۷ = ۰/۴۰۹۹\\
P(-g) &= P(-g,-e) + P(-g,+e) = ۰/۵۳۲۸ + ۰/۰۵۷۳ = ۰/۵۹۰۱
\end{aligned}
$$

پس:

$$
\begin{aligned}
P(E=۰ \mid G=۰) &= \frac{۰/۵۳۲۸}{۰/۵۹۰۱} \approx ۰/۹۰۲, \quad P(E=۱ \mid G=۰) = \frac{۰/۰۵۷۳}{۰/۵۹۰۱} \approx ۰/۰۹۸\\
P(E=۰ \mid G=۱) &= \frac{۰/۳۶۷۲}{۰/۴۰۹۹} \approx ۰/۸۹۶, \quad P(E=۱ \mid G=۱) = \frac{۰/۰۴۲۷}{۰/۴۰۹۹} \approx ۰/۱۰۴
\end{aligned}
$$

۶. (۱۰ نمره) گراف زیر را در نظر بگیرید:



(آ) فرض کنید متغیرهای a, b, e, h, i دو مقداری، متغیرهای c, f, g سه مقداری و متغیر d چهار مقداری است. با توجه به والدهای هر گره در گراف بالا، کمترین تعداد سطر موردنیاز برای جدول‌های احتمالات شرطی هر گره (CPT) را مشخص کنید.

(ب) با استفاده از Variable Elimination، مقدار زیر را حساب کنید:

$$P(e, f, c, g, i)$$

ابتدا ترتیب دلخواه خود را نوشته و سپس محاسبات خود را کامل بنویسید. جواب نهایی و مراحل شما باید به صورت روابط ریاضیاتی مختلف روی توزیع‌های احتمالاتی مختلف موجود در شبکه باشند.

حل.

(آ) ابتدا، دامنه‌های هر متغیر را فهرست می‌کنیم:

- $|a| = ۲, |b| = ۲, |e| = ۲, |h| = ۲, |i| = ۲$
- $|c| = ۳, |f| = ۳, |g| = ۳$
- $|d| = ۴$

تعداد سطرها در جدول کامل احتمال شرطی (CPT) برای یک گره X با توجه به والدهایش $Pa(X)$ برابر است با $|X| \times \prod_{Y \in Pa(X)} |Y|$. (توجه: اگر فقط سطرهای پارامترهای مستقل را در نظر بگیریم، این مقدار کمتر خواهد بود، اما اندازهٔ کل جدول همهٔ حالات احتمالات شرطی را نشان می‌دهد.)

بر اساس گراف، والدها و اندازه‌های CPT به‌صورت زیر هستند:

- گره a: $Pa(a) = \emptyset \Rightarrow |a| = ۲$
- گره d: $Pa(d) = \emptyset \Rightarrow |d| = ۴$
- گره c: $Pa(c) = \{a\} \Rightarrow |c| \times |a| = ۳ \times ۲ = ۶$
- گره b: $Pa(b) = \{c\} \Rightarrow |b| \times |c| = ۲ \times ۳ = ۶$
- گره e: $Pa(e) = \{c, d\} \Rightarrow |e| \times |c| \times |d| = ۲ \times ۳ \times ۴ = ۲۴$
- گره f: $Pa(f) = \{e\} \Rightarrow |f| \times |e| = ۳ \times ۲ = ۶$
- گره h: $Pa(h) = \{e\} \Rightarrow |h| \times |e| = ۲ \times ۲ = ۴$
- گره g: $Pa(g) = \{d, f, h\} \Rightarrow |g| \times |d| \times |f| \times |h| = ۳ \times ۴ \times ۳ \times ۲ = ۷۲$
- گره i: $Pa(i) = \{h\} \Rightarrow |i| \times |h| = ۲ \times ۲ = ۴$

(ب) می‌خواهیم احتمال حاشیه‌ای $P(e,f,c,g,i)$ را محاسبه کنیم. توزیع مشترک کامل شبکهٔ بیزی برابر است با:

$$P(a,b,c,d,e,f,g,h,i) = P(a)P(c|a)P(b|c)P(d)P(e|c,d)P(f|e)P(h|e)P(g|d,f,h)P(i|h)$$

باید متغیرهای $a,b,d,$ و h را حاشیه‌گیری کنیم:

$$P(e,f,c,g,i) = \sum_{a,b,d,h} P(a,b,c,d,e,f,g,h,i)$$

ترتیب حذف را به‌صورت b,a,h,d انتخاب می‌کنیم.

گام ۱: حذف b

تنها عاملی که شامل b است، $P(b|c)$ می‌باشد.

$$m_۱(c) = \sum_b P(b|c) = ۱$$

گام ۲: حذف a

عوامل شامل a عبارت‌اند از $P(a)$ و $P(c|a)$.

$$m_۲(c) = \sum_a P(a)P(c|a) = P(c)$$

گام ۳: حذف h

عوامل شامل h عبارت‌اند از $P(h|e), P(g|d,f,h),$ و $P(i|h)$.

$$m_۳(e,g,d,f,i) = \sum_h P(h|e)P(g|d,f,h)P(i|h)$$

گام ۴: حذف d

عوامل شامل d عبارت‌اند از $P(d), P(e|c,d),$ و $m_۳(e,g,d,f,i)$.

$$m_۴(c,e,g,f,i) = \sum_d P(d)P(e|c,d)m_۳(e,g,d,f,i)$$

نتیجهٔ نهایی:

عوامل باقی‌مانده را در هم ضرب می‌کنیم:

$$P(e,f,c,g,i) = m_۱(c)\cdot m_۲(c)\cdot P(f|e)\cdot m_۴(c,e,g,f,i)$$

$$P(e,f,c,g,i) = P(c)P(f|e)\sum_d \left(P(d)P(e|c,d)\sum_h P(h|e)P(g|d,f,h)P(i|h) \right)$$