دانشگاه صنعتی امیرکبیر

دانشکده مهندسی کامپیوتر

«مبانی و کاربردهای هوش مصنوعی ترم بهار ۱۴۰۵»

تمرین پنجم

Bayes' Net

در انجام تمرین‌ها به نکات زیر توجه فرمائید:

۱- مطابق قوانین دانشگاه هر نوع کپی‌برداری و اشتراک کار دانشجویان غیرمجاز است و پاسخ به تمرین‌ها باید به صورت انفرادی انجام شود.

۲- پاسخ خود را باید در قالب یک فایل PDF به صورت تایپ‌شده و یا دستنویس (مرتب و خوانا) در سامانه کورسز آپلود نمائید؛ توجه کنید که مشکل ناخوانا بودن پاسخ‌ها در هنگام تصحیح، بر عهده دانشجو خواهد بود.

۳- در صورت هر گونه سوال و یا ابهام از طریق ایمیل زیر می‌توانید در ارتباط باشید:

h.masroor@aut.ac.ir

۴- توجه نمائید پاسخ تمرین‌ها تنها در صورت آپلود در سامانه کورسز پذیرفته خواهد شد و ارسال پاسخ از طریق ایمیل و یا روش‌های دیگر بررسی نخواهد شد.

۵- فایل پاسخ‌نامه خود را با فرمت StudentID_HW05.pdf تا ساعت ۲۳:۵۹ روز ۱۴۰۵/۰۴/۰۱ فقط در بخش مربوطه در سایت درس آپلود نمائید.

سوال اول

متغیرهای A, B, C, D دارای توزیع‌های زیر هستند:

A	P(A)
+a	0.8
-a	0.2

A	B	P(B|A)
+a	+b	0.8
+a	-b	0.2
-a	+b	0.3
-a	-b	0.7

B	C	P(C|B)
+b	+c	0.9
+b	-c	0.1
-b	+c	0.4
-b	-c	0.6

C	D	P(D|C)
+c	+d	0.15
+c	-d	0.85
-c	+d	0.5
-c	-d	0.5

الف) آیا بدون هیچ فرضی از استقلال متغیرها، می‌توان $P(-a,-b,+c)$ را محاسبه کرد؟ اگر بله، مقدار آن را محاسبه کنید. اگر خیر، کمترین فرض استقلالی لازم را بیان کنید.

ب) فرض کنید C نسبت به A با دانستن B مستقل است و D هم نسبت به تمام متغیرهای دیگر با دانستن C مستقل است. احتمال $P(+a,-b,+c,+d)$ را محاسبه کنید.

ج) بدون فرض هیچ نوع استقلالی، نشان دهید کدام تجزیه‌های زیر برابر با $P(R,S,T)$ هستند و دلیل خود را بنویسید. (می‌تواند بیش از یک گزینه باشد)

۱) $P(R \mid S,T)P(S \mid T)P(T)$

۲) $P(T,S \mid R)P(R)$

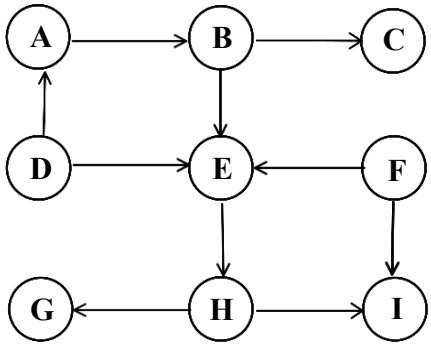
۳) $P(T \mid R,S)P(R,S)$

۴) $P(R \mid S,T)P(S \mid R,T)P(T \mid R,S)$

د) توضیح دهید در Variable Elimination چرا ترتیب حذف متغیرها بر پاسخ نهایی تأثیر نمی‌گذارد ولی می‌تواند بر زمان اجرا و حافظه موردنیاز الگوریتم اثر بگذارد.

سوال دوم

با توجه به شبکه بیزین زیر، تعیین کنید درستی یا نادرستی هرکدام از موارد زیر گارانتی می‌شود یا نمی‌توان مشخص کرد. اگر تضمین نمی‌شود، یک active path ارائه کنید.



الف) $A \perp\!\!\!\perp B$

ب) $A \perp\!\!\!\perp F$

پ) $A \perp\!\!\!\perp F \mid E$

ت) $A \perp\!\!\!\perp H \mid E, I$

ث) $F \perp\!\!\!\perp E \mid B, D$

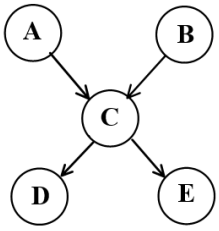
ج) $D \perp\!\!\!\perp G \mid A$

چ) $I \perp\!\!\!\perp B \mid E$

ح) $E \perp\!\!\!\perp G$

سوال سوم

شبکه بیزی زیر را در نظر بگیرید.



A	P(A)
+a	0.6
-a	0.4

B	P(B)
+b	0.7
-b	0.3

A	B	C	P(C|A, B)
+a	+b	+c	0.9
+a	+b	-c	0.1
+a	-b	+c	0.4
+a	-b	-c	0.6
-a	+b	+c	0.7
-a	+b	-c	0.3
-a	-b	+c	0.2
-a	-b	-c	0.8

C	D	P(D|C)
+c	+d	0.8
+c	-d	0.2
-c	+d	0.3
-c	-d	0.7

C	E	P(E|C)
+c	+e	0.9
+c	-e	0.1
-c	+e	0.4
-c	-e	0.6

الف) با استفاده از Enumeration مقدار $P(B = +b \mid E = +e)$ را محاسبه کنید. تمام مراحل محاسبه صورت و مخرج را بنویسید.

ب) برای محاسبه $P(B \mid E = +e)$ متغیرهای پنهان عبارت‌اند از A, C, D. کدام‌یک از ترتیب‌های حذف زیر بیشترین اندازه فاکتور واسط (در مرحله join و قبل از اعمال sum-out) را ایجاد می‌کند؟ اندازه بزرگ‌ترین فاکتور ایجادشده در هر ترتیب را مشخص کنید.

۱) $A \to C \to D$ ۲) $C \to D \to A$

۳) $D \to C \to A$ ۴) $A \to D \to C$

ج) یک ترتیب حذف را انتخاب کنید و با استفاده از روش Variable Elimination مقدار $P(B = +b \mid E = +e)$ را محاسبه کنید..

سوال چهارم

فرض کنید بوی گوگرد را با S نشان می‌دهیم. این بو می‌تواند به دو علت رخ دهد:

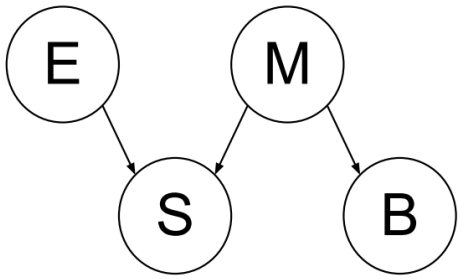
- وجود تخم‌مرغ گندیده (E)
- وقوع آخرالزمان (M)

همچنین، رخداد M باعث می‌شود اقیانوس‌ها به جوش بیایند. (B) ساختار شبکه بیزی مطابق شکل سؤال است و توزیع‌های زیر داده شده‌اند. (همه متغیرها دودویی هستند و با نمادهای $+/-$ نشان داده می‌شوند).

الف) با استفاده از تجزیه joint مطابق قانون زنجیره‌ای (chain rule) و ساختار شبکه بیزی، مقدار زیر را محاسبه کنید (پاسخ می‌تواند عدد نهایی یا یک عبارت ضربی/جمعی از اعداد جدول‌ها باشد): $P(-e, -s, -m, -b)$

$P(E)$	
$+e$	0.4
$-e$	0.6

$P(M)$	
$+m$	0.1
$-m$	0.9



$P(S\|E,M)$			
$+e$	$+m$	$+s$	1.0
$+e$	$+m$	$-s$	0.0
$+e$	$-m$	$+s$	0.8
$+e$	$-m$	$-s$	0.2
$-e$	$+m$	$+s$	0.3
$-e$	$+m$	$-s$	0.7
$-e$	$-m$	$+s$	0.1
$-e$	$-m$	$-s$	0.9

$P(B\|M)$		
$+m$	$+b$	1.0
$+m$	$-b$	0.0
$-m$	$+b$	0.1
$-m$	$-b$	0.9

ب) احتمال اینکه اقیانوس‌ها بجوشند چقدر است؟ یعنی: $P(+b)$

(راهنما: می‌توانید با marginalization روی M این مقدار را به‌دست آورید).

پ) اگر مشاهده کنیم که اقیانوس‌ها در حال جوشیدن هستند $(B = +b)$، احتمال اینکه آخرالزمان رخ داده باشد چقدر است؟ $P(+m \mid +b) = \ ?$

ج) اگر سه شاهد (evidence) زیر را هم‌زمان داشته باشیم:

- بوی گوگرد وجود دارد $(S = +s)$
- اقیانوس‌ها در حال جوشیدن هستند $(B = +b)$
- تخم‌مرغ گندیده وجود دارد $(E = +e)$

احتمال اینکه آخرالزمان رخ داده باشد چقدر است؟ $P(+\mathrm{m} \mid +s, +b, +e) = \ ?$

د) اگر بدانیم آخرالزمان رخ داده است $(M = +m)$، احتمال اینکه تخم‌مرغ گندیده وجود داشته باشد چقدر است؟ $P(+e \mid +m) = \ ?$

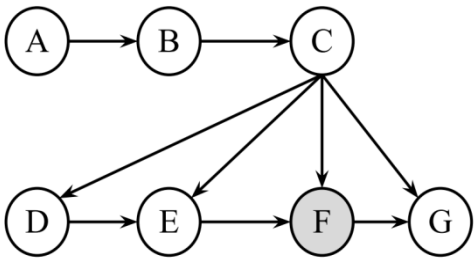
سوال پنجم

برای شبکه بیزیِ نشان‌داده‌شده در شکل (همه متغیرها دودویی هستند)، کوئری زیر داده شده است:

$$P(B, D \mid +f)$$

فرض کنید برای محاسبه این کوئری از حذف متغیر (Variable Elimination) استفاده می‌کنیم و ترتیب حذف به شکل روبه‌رو است: A, C, E, G

(یعنی ابتدا A را sum-out می‌کنیم، سپس C، سپس E، و در نهایت G).



الف) فرآیند تولید factor ها را کامل کنید.

پس از اعمال evidence یعنی $(F = +f)$، factor های اولیه به شکل زیر هستند:

$$P(A),\ P(B \mid A),\ P(C \mid B),\ P(D \mid C),$$
$$P(E, \mid CD),\ P(+, f \mid CE),\ P(G, \mid C + f)$$

با حذف A یک factor جدید به‌دست می‌آید: $f_1(B) = \sum_a P(a)\, P(B \mid a)$

و سپس factor های باقی‌مانده عبارت‌اند از:

$$P(C \mid B),\quad P(D \mid C),\quad P(E \mid C, D),\quad P(+f \mid C, E),\quad P(G \mid C, +f),\quad f_1(B)$$

اکنون باید برای سه حذف بعدی، factor های جدید را بنویسید:

۱. هنگام حذف C، factor جدید f_2 را بنویسید (به‌صورت حاصل‌ضرب factor های درگیر و سپس $\sum_c$).

۲. هنگام حذف E، factor جدید f_3 را بنویسید (به‌صورت $\sum_e$).

۳. هنگام حذف G، factor جدید f_4 را بنویسید (به‌صورت $\sum_g$).

(در هر مرحله همچنین بنویسید بعد از تولید factor جدید، چه factor هایی باقی می‌مانند).

ب) در یک جمله توضیح دهید بعد از پایان بخش (الف) وقتی فقط چند factor باقی مانده‌اند، چگونه می‌توان از آن‌ها مقدار $P(B, D \mid +f)$ را به‌دست آورد؟

پ) در بین factor های (f_1, f_2, f_3, f_4)، بزرگ‌ترین factor کدام است و اندازه آن چقدر است؟

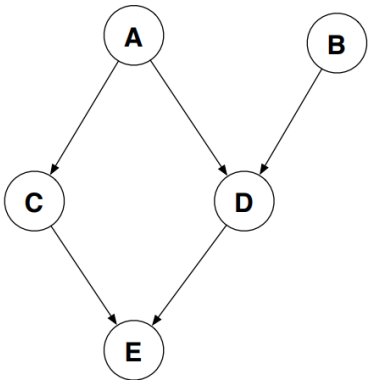
فرض کنید همه متغیرها دودویی‌اند و اندازه یک factor برابر تعداد سطرهای جدول آن (تعداد assignment های ممکن) است.

ت) یک ترتیب حذف جدید برای همین کوئری پیدا کنید (یعنی برای $P(B, D \mid +f)$) که در آن بزرگ‌ترین factor تولیدشده در طول اجرا حداقل شود.

راهنما: در ترتیب بهینه، بزرگ‌ترین factor فقط ۲ متغیر دارد (پس اندازه جدول آن $2^2 = 4$ است).

ترتیب حذف و factor تولیدشده در هر مرحله را در جدول مربوطه بنویسید.

سوال ششم



یک شبکه بیزی با ۵ متغیر مطابق شکل زیر در نظر بگیرید:

فرض کنید هر متغیر X دو مقدار دارد x_1 و x_2.

(مثلا A مقادیر a_1 و a_2 را دارد).

راهنمای نوشتن factor ها:

در پاسخ‌ها، هر factor را با نوشتن متغیرهای داخل آن مشخص کنید. اگر شواهد (evidence) داشته باشیم، به‌جای نام متغیر، مقدار مشاهده‌شده را بنویسید. مثال: $f_3(E,\ C,\ d_1)$ یا $f_1(A,\ E)$

الف) نشان دهید چگونه با الگوریتم Variable Elimination و با ترتیب حذف زیر:

$$A,\ E,\ C,\ D$$

می‌توان $P(B)$ را محاسبه کرد.

- دنباله factor های جدید ساخته‌شده را بنویسید .
- مشخص کنید در هر گام، کدام factor ها در هم ضرب می‌شوند و سپس روی کدام متغیر جمع زده می‌شود (sum-out) تا factor جدید ساخته شود .
- به هر factor جدید یک نام یکتا بدهید (مثل $f_5(A,\ E)$).

ب) در فرایند بخش (الف)، بزرگ‌ترین factor ی که ساخته می‌شود کدام است و اندازه آن چقدر است؟ (بر حسب تعداد assignment ها)

نکته :اندازه را قبل از اجرای sum-out گزارش کنید.

پ) اگر به‌جای ترتیب قبلی، ترتیب حذف این باشد: $D,\ C,\ A,\ E$

بزرگ‌ترین factor در طول اجرای VE چقدر بزرگ می‌شود؟ (راهنما: لازم نیست کل فرایند را مرحله به مرحله بنویسید؛ فقط کافی است بفهمید در بدترین مرحله، factor روی چند متغیر خواهد بود.)

ج) اگر هدف این باشد که به‌جای $P(B)$، مقدار زیر را محاسبه کنیم: $P(B \mid C = c_1)$

فرایند VE که در بخش (الف) نوشتید دقیقاً چگونه تغییر می‌کند؟ توضیح دهید. برای اینکه پاسخ مشخص باشد، حتماً بگویید: evidence یعنی $C = c_1$ روی کدام factor ها اعمال می‌شود (یعنی کدام factor ها باید به مقدار c_1 انتخاب شوند) و آیا در ترتیب حذف، متغیری هست که دیگر لازم نباشد حذف شود یا ترتیب حذف باید تغییر کند.

د) (**امتیازی**) فرض کنید می‌خواهید با likelihood weighting مقدار زیر را تقریب بزنید:

$$P(E' = e_1 \mid C = c_1 \; A = a_2)$$

فرمول وزن (weight) مربوط به نمونه زیر را بنویسید: $(a_2,\ b_1,\ c_1,\ d_2,\ e_1)$

فرمول را بر حسب CPT های شبکه بنویسید .

ه) فرض کنید یک شبکه بیزی بسیار بزرگ داریم که ارتباط بین صدها بیماری و صدها نشانه (symptom) را مدل می‌کند:

- گره‌های بیماری در بالای شبکه هستند و والد ندارند .
- تعداد زیادی گره میانی وجود دارد (بین بیماری‌ها و نشانه‌ها) که وضعیت اندام‌های داخلی مثل کلیه و ریه را نشان می‌دهند .
- گره‌های نشانه‌ها در پایین شبکه هستند .
- یک نشانه می‌تواند در نهایت به‌وسیله چند بیماری مختلف ایجاد شود .

می‌خواهیم از sampling (به‌خصوص likelihood weighting) برای تخمین احتمال‌های شرطی استفاده کنیم. نقاط قوت و ضعف این الگوریتم را برای کوئری‌های زیر بیان کنید:

(۱) $P\left(\text{symptom}_i \mid \text{disease}_j\right)$

(۲) $P\left(\text{disease}_j \mid \text{symptom}_i\right)$

(۳) چه زمانی (اگر اصلاً زمانی باشد)، rejection sampling را به‌جای likelihood weighting استفاده می‌کنیم؟