به‌نام خدا

TPM and Bayesian Network – تمرین۳

هوش مصنوعی

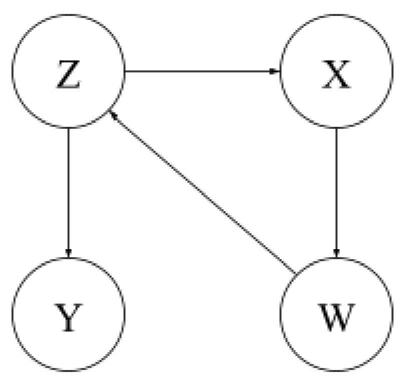
مهراد میلانلو

۹۹۱۰۵۷۷۵

سوال ۱.

شبکه‌های بیزی در واقع نوعی گراف جهت‌دار بدون دور (DAG) هستند. اکنون فرض کنید می‌خواهیم یک شبکه‌ی بیزی با n راس با بیش‌ترین تعداد یال بسازیم. میان این n راس $\binom{n}{2}$ مکان برای قرار دادن یال وجود دارد. اکنون فرض کنید $\binom{n}{2} + 1$ یال در شبکه قرار بدهیم. طبق اصل لانه کبوتری، دو راس وجود دارد که بین آن‌ها دو یال قرار می‌گیرد. بنابراین در گراف دور داریم و حداکثر $\frac{n\times(n-1)}{2}$ یال می‌توانیم داشته باشیم. اکنون فرض کنید راس‌ها را به ترتیب از ۱ تا n نام‌گذاری کرده‌ایم. برای راس nام می‌توانیم حداکثر n-1 یال به راس‌های دیگر متصل کنیم. برای راس دوم حداکثر n-2 راس برای وصل کردن موجود است. (زیرا اگر به راس‌های قبلی یال رسم کنیم دور داریم) به همین ترتیب در نهایت حداکثر $\frac{n\times(n-1)}{2}$ یال می‌توان در شبکه‌ای با n راس داشت.

سوال ۲.



در شبکه‌ی بالا واضح است که W و Y به شرط دانستن Z مستقل‌اند(Causal Chain). همین‌طور Y و X به شرط دانستن Z مستقل‌اند(Common Cause). همچنین مشخص است که میان بقیه‌ی رئوس در حالت کلی هیچ استقلالی برقرار نیست.

سوال ۳.

می‌دانیم:

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i | x_1 ... x_{i-1})$$

$$P(x_i|x_1 \dots x_{i-1}) = P(x_i|parents(X_i))$$

$$\Rightarrow P(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i))$$

$$\Rightarrow P(A, B, C, D, E, F, G) = P(A) \times P(B|A) \times P(C|A) \times P(D) \times P(E|D, A) \times P(F|D) \times P(G|B, C, E, F)$$

سوال ۴.

ابتدا توزیع توام رئوس را بر اساس initial factorها می‌نویسیم:

$$P(A, B, C, D, E, F, G) = P(A) \times P(B|A) \times P(C|A, B) \times P(D|A, C) \times P(E|A, D) \times P(F|A, E) \times P(G|B, C)$$

اکنون الگوریتم variable elimination را اجرا می‌کنیم. دقت کنید از آن‌جا که رئوس B, G, E در Query موردنظر وجود دارند، نیازی به حذفشان نیست.

آ)

طبق ترتیب حذف ذکر شده داریم:

Initial factors:

$P(A),\ P(B|A),\ P(C|A, B),\ P(D|A, C),\ P(e|A, D),\ P(F|A, e),\ P(g|B, C)$

حذف A:

$$f_1(B, C, D, e, F) = \sum_a P(a) \times P(B|a) \times P(C|B, a) \times P(D|a, C) \times P(e|D, a) \times P(F|e, a)$$

New factors:

$f_1(B, C, D, e, F),\ P(g|B, C)$

حذف C:

$$f_2(B, D, e, F, g) = \sum_c f_1(B, c, D, e, F) \times P(g|B, c)$$

New factors:

$f_2(B, D, e, F, g)$

حذف D:

$$f_3(B, e, F, g) = \sum_d f_2(B, d, e, F, g)$$

New factors:

$f_3(B, e, F, g)$

حذف F:

$$f_4(B, e, g) = \sum_f f_3(B, e, f, g)$$

New factors:

$f_4(B, e, g)$

اکنون با normalize کردن $f_4(B, e, g)$، به پاسخ مورد نظر می‌رسیم.

ب)

طبق ترتیب ذکر شده برای حذف داریم:

Initial factors:

$P(A),\ P(B|A),\ P(C|A, B),\ P(D|A, C),\ P(e|A, D),\ P(F|A, e),\ P(g|B, C)$

حذف F:

$$f_1(A, e) = \sum_f P(f|A, e)$$

New factors:

$P(A),\ P(B|A),\ P(C|A, B),\ P(D|A, C),\ P(e|A, D),\ P(g|B, C),\ f_1(A, e)$

حذف D:

$$f_2(A, C, e) = \sum_d P(d|A, C) \times P(e|d, A)$$

New factors:

$P(A),\ P(B|A),\ P(C|A, B),\ P(g|B, C),\ f_1(A, e),\ f_2(A, C, e)$

حذف C:

$$f_3(A, e, B, g) = \sum_c P(c|B, A) \times P(g|B, c) \times f_2(A, c, e)$$

New factors:

$P(A),\ P(B|A),\ f_1(A, e),\ f_3(A, B, e, g)$

حذف A:

$$f_4(B, g, e) = f_1(A, e) \times f_3(a, c, B, g) \times \sum_a P(a) \times P(B|a)$$

New factors:

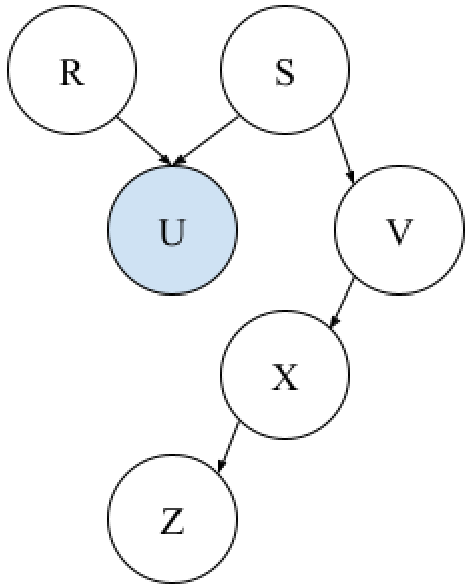
$f_4(B, e, g)$

ج)

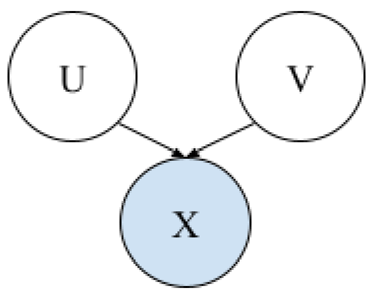
در عملیات بخش (آ)، فاکتورهای جدید ساخته‌شده حداکثر ۴ متغیر نهان دارند که همان f_1 است. در قسمت (ب) فاکتورها حداکثر ۲ متغیر نهان دارند؛ یعنی f_3 , f_2. بنابراین برای محاسبه‌ی این توابع در بخش (الف) نیاز به $2^4 = 16$ سطر داریم اما در بخش (ب)، $2^2 = 4$ سطر از جدول توزیع توام کافی‌است. بنابراین ترتیب قسمت (ب) بهتر است.

سوال ۵.

آ) نادرست. (مسیر فعال: R, U, S, V, X, Z)

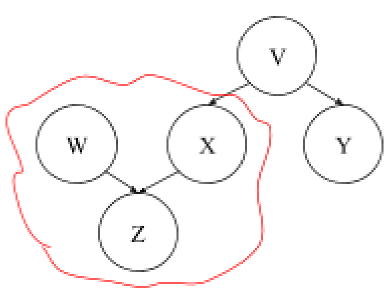


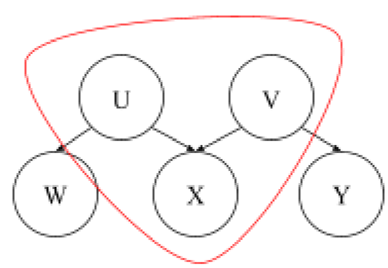
ب) نادرست. (مسیر فعال: U, X, V)

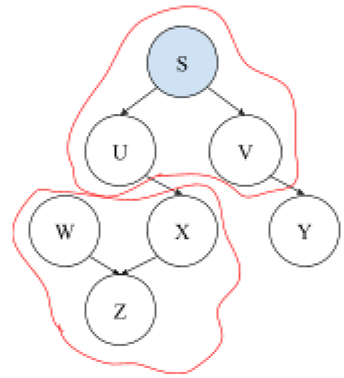


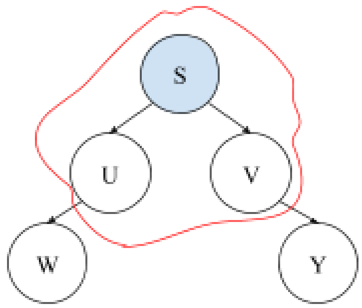
ج) درست.

برای تمام مسیرهای بین مبدا و مقصد (یعنی WوY)، یک سه‌تایی مشخص می‌کنیم که نشان بدهد مسیر غیرفعال است.



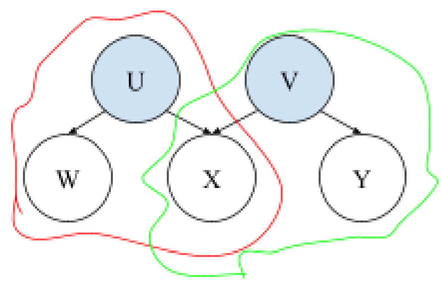


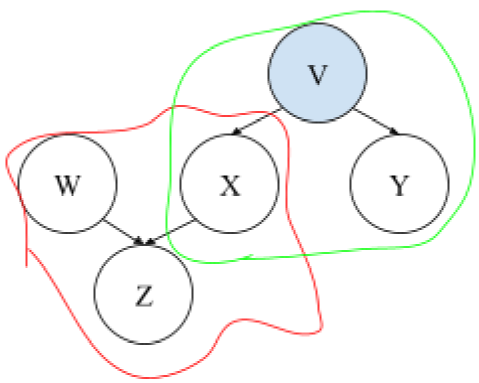


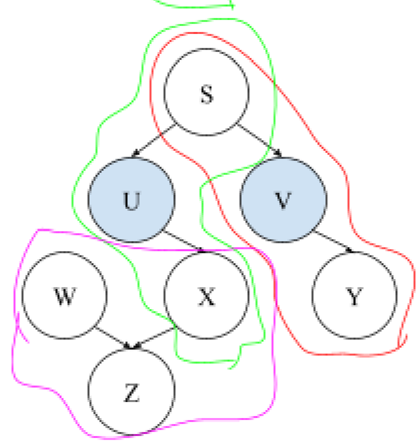


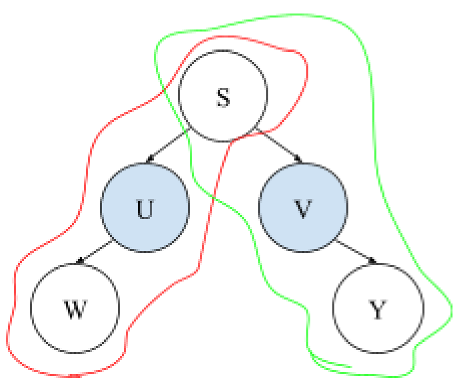
د)

برای تمام مسیرهای بین مبدا و مقصد (یعنی WوY)، یک سه‌تایی مشخص می‌کنیم که نشان بدهد مسیر غیرفعال است.

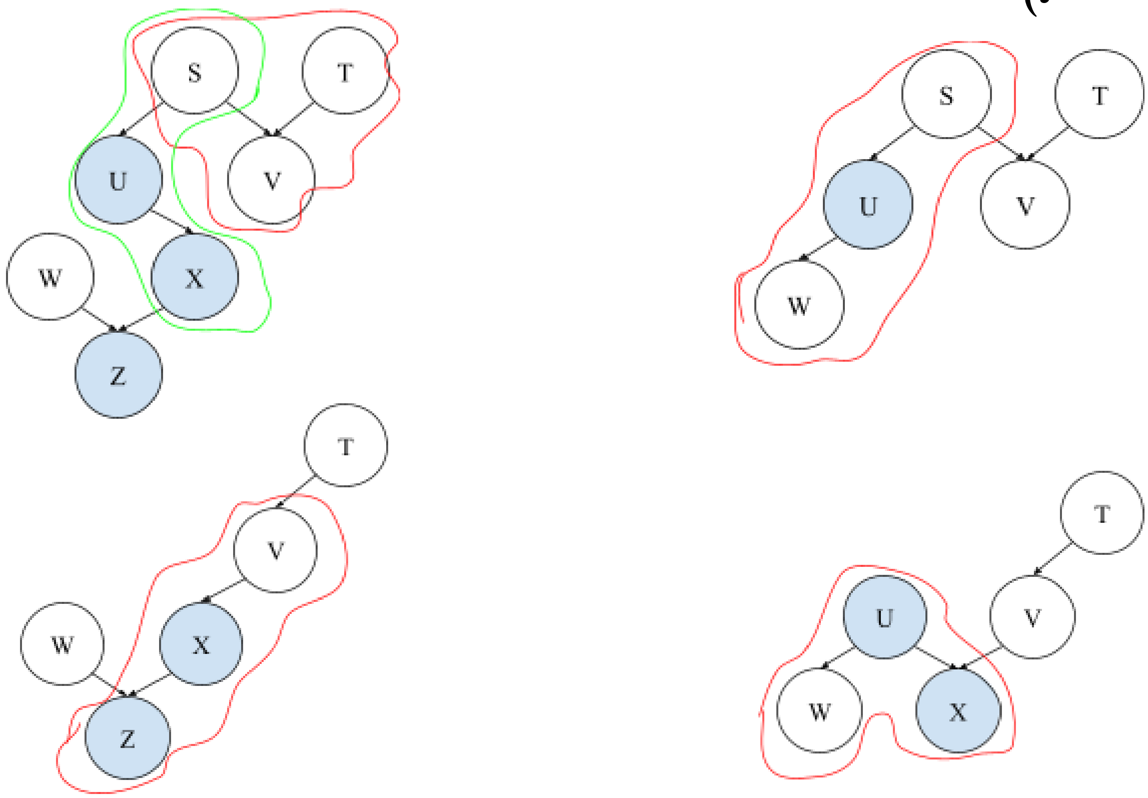








ه)



سوال ۶.

آ) Rejection Sampling

هدف ایجاد Sampleهاست. عیب این روش آن است که ممکن است نمونه‌های زیادی Reject شوند و با توجه به تعداد کم نمونه‌ها، به جواب درست نرسیم. بنابراین برای دقت بالای این روش نیاز به داده‌های بزرگ داریم.

اکنون نمونه‌ها را با توجه به اعداد تصادفی داده شده به‌دست می‌آوریم. برای هر راس، اگر $P(p_i|Parents(P_i) \geq r_j$ باشد، مقدار P_i را در نمونه اضافه کرده و در غیر این‌صورت $\neg p_i$ را در نظر می گیریم. اگر شرط‌های Observe شده‌ی Query در نمونه نقض شود، آن را Reject می کنیم.

$r_1 < P(p_1)$	$r_2 < P(p_2 \vert p_1)$	$r_3 > P(p_3 \vert p_2)$	$r_4 > P(p_4 \vert p_2)$	
p_1	p_2	$\neg p_3$	$\neg p_4$	$\rightarrow (p_1, p_2, \neg p_3, \neg p_4)$
$r_5 > P(p_1)$	$r_6 > P(p_2 \vert \neg p_1)$			
$\neg p_1$	$\neg p_2$			$\rightarrow\ Rejection$
$r_7 < P(p_1)$	$r_8 < P(p_2 \vert p_1)$	$r_9 > P(p_3 \vert p_2)$	$r_{10} > P(p_4 \vert p_2)$	
p_1	p_2	$\neg p_3$	$\neg p_4$	$\rightarrow (p_1, p_2, \neg p_3, \neg p_4)$
$r_{11} < P(p_1)$	$r_{12} > P(p_2 \vert p_1)$			
p_1	$\neg p_2$			$\rightarrow\ Rejection$
$r_{13} > P(p_1)$	$r_{14} < P(p_2 \vert \neg p_1)$	$r_{15} > P(p_3 \vert p_2)$	$r_{16} < P(p_4 \vert p_2)$	
$\neg p_1$	p_2	$\neg p_3$	p_4	$\rightarrow (\neg p_1, p_2, \neg p_3, p_4)$
$r_{17} > P(p_1)$	$r_{18} > P(p_2 \vert \neg p_1)$			
$\neg p_1$	$\neg p_2$			$\rightarrow\ Rejection$
$r_{19} < P(p_1)$	$r_{20} < P(p_2 \vert p_1)$			
p_1	p_2			$\rightarrow\ No\ more\ Data!$

در نهایت سه نمونه داریم که در دوتا از آن‌ها p_1 داریم و شروط مسئله رعایت شده است. پس:

$P(p_1 | p_2, \neg p_3) = \frac{2}{3}$

ب) Likelihood Weighting

این روش نیز تقریبا مشابه حالت قبل عمل می‌کند. اما به‌جای اینکه احتمال داده‌های شرط را نیز با اعداد تصادفی مقایسه کرده و یک نمونه را Reject کنیم، شرط‌ها را Observe شده و ثابت در نظر می‌گیریم. با توجه به آن‌ها برای حالات مختلف Query یک Weight به‌دست می‌آوریم و در نهایت در محاسبه‌ی پاسخ نهایی از آن استفاده می‌کنیم. مزیت این روش نسبت به روش قبلی، دقت بالاتر و عدم نیاز به داده‌های بزرگ است. اما همچنان متغیرهای پدر، مقداردهی ثابت شده‌اند که مستقل از عملیات مشاهده است و باعث کمتر شدن دقت تخمین می‌شود.

اکنون در این سوال، P_2, P_3 ثابت شده‌اند. بنابراین اعداد رندوم را برای به‌دست آوردن مقادیر P_1, P_4 استفاده می کنیم. بنابراین اعداد رندوم با شماره‌ی فرد را به متغیر P_1 تخصیص می‌دهیم. اگر عدد تصادفی مربوطه از احتمال $P(p_1)$ کمتر باشد، مقدار متغیر را p_1 در نظر گرفته و در غیر این‌صورت $\neg p_1$ می گیریم.

$P(p_1) = 0.4$

$r_1 < P(p_1)$ $r_3 > P(p_1)$ $r_5 > P(p_1)$ $r_7 < P(p_1)$ $r_9 < P(p_1)$ $r_{11} < P(p_1)$

$r_{13} > P(p_1)$ $r_{15} > P(p_1)$ $r_{17} > P(p_1)$ $r_{19} < P(p_1)$

$\rightarrow C(p_1) = 5, C(\neg p_1) = 5$

اکنون وزن‌های مربوطه را به‌دست می‌آوریم.

$W_1 = P(\neg p_3|p_2) \times P(p_2|p_1) \Rightarrow W_1 = 0.8 \times 0.8 = 0.64$

$W_2 = P(\neg p_3|p_2) \times P(p_2|\neg p_1) \Rightarrow W_1 = 0.8 \times 0.5 = 0.4$

$P(p_1|p_2, \neg p_3) = \frac{W_1 \times C(p_1)}{W_1 \times C(p_1) + W_2 \times C(\neg p_1)} = \frac{0.64 \times 5}{0.64 \times 5 + 0.4 \times 5} = \frac{0.64}{1.04} = 0.615$

ج) Gibbs Sampling

در این روش مقدار متغیرهای Observe شده را ثابت در نظر می گیریم. متغیرهای دیگر را به‌طور تصادفی (از اعداد رندوم داده شده) مقداردهی اولیه می کنیم. اکنون احتمال‌های زیر را حساب می کنیم: (به‌دست آوردن روابط بسیار ساده است و در کلاس اثبات شده.)

$P(p_1|p_2, \neg p_3, p_4) = \frac{P(p_1) \times P(p_2|p_1)}{\sum_{p_1} P(p_1) \times P(p_2|p_1)} = \frac{0.4 \times 0.8}{0.4 \times 0.8 + 0.6 \times 0.5} = \frac{0.32}{0.62} = 0.52$

$P(p_4|p_1, p_2, \neg p_3) = \frac{P(p_4|p_2)}{\sum_{p_4} P(p_4|p_2)} = 0.8$

حال با احتمال 0.52 $P_1 = p_1$ و در غیر این‌صورت $P_1 = \neg p_1$.

اکنون از اعداد تصادفی داده شده $(r_2 - r_{20})$، اعداد با شماره‌ی فرد را مربوط به متغیر P_1 در نظر گرفته و باقی را مربوط به متغیر P_4 می گیریم. به ازای هرکدام از اعداد با شماره‌ی فرد اگر کوچک‌تر از 0.52 باشد، مقدار P_1 را p_1 و در غیر این‌صورت $\neg p_1$ می گیریم. حال داریم:

$P(p_1) = 0.52$

$r_3 > P(p_1)$ $r_5 > P(p_1)$ $r_7 < P(p_1)$ $r_9 < P(p_1)$ $r_{11} < P(p_1)$ $r_{13} > P(p_1)$

$r_{15} > P(p_1)$ $r_{17} < P(p_1)$ $r_{19} < P(p_1)$

$\Rightarrow P(p_1|p_2, \neg p_3) = \frac{5}{9} = 0.56$

در این روش محاسبات با دقت بیش‌تری انجام می‌شود.