هوش مصنوعی

تکلیف شماره‌ی ۹

کمی‌سازی عدم اطمینان و استدلال احتمالاتی

(۱)

مثال) سال 13-6 کتاب aima2e

با توجه به full joint distribution داده شده در جدول زیر، مقادیر زیر را محاسبه کنید:

	دندان درد		¬ دندان درد	
	toothace		¬ toothace	
	catch	¬ catch	catch	¬ catch
پوسیدگی دندان cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

متغیرهای تصادفی بولی

(a) $P(toothace)$ Toothace = true

$$= 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

(b) $P(cavity)$ Cavity = true

$$= 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

(c) $\mathbb{P}(Cavity)$ توزیع احتمال (بردار مقادیر احتمال) متغیر تصادفی Cavity

$$= \langle P(cavity), P(\neg cavity) \rangle$$

$$= \langle 0.2, 0.8 \rangle$$

(d) $\mathbb{P}(Toothace \mid cavity)$ احتمال دندان درد به شرط درستی کرم خوردگی

$$= \langle P(toothace \mid cavity), P(\neg toothace \mid cavity) \rangle$$

$$= \left\langle \frac{P(toothace \wedge cavity)}{P(cavity)}, \frac{P(\neg toothace \wedge cavity)}{P(cavity)} \right\rangle$$

$$= \left\langle \frac{0.108 + 0.012}{0.2}, \frac{0.072 + 0.008}{0.2} \right\rangle$$

$$= \langle 0.6, 0.4 \rangle$$

(e) $\mathbb{P}(Cavity \mid toothace \vee catch)$

$$= \left\langle \frac{P(cavity \wedge (toothace \vee catch))}{P(toothace \vee catch)}, \frac{P(\neg cavity \wedge (toothace \vee catch))}{P(toothace \vee catch)} \right\rangle$$

$$= \left\langle \frac{0.108 + 0.012 + 0.072}{0.416}, \frac{0.016 + 0.064 + 0.144}{0.416} \right\rangle$$

$$= \langle 0.4615, 0.5384 \rangle$$

مثال) مساله 8_13 کتاب aima2e

پس از انجام checkup، پزشک یک خبر خوب و یک خبر بد برای شما دارد:

خبر بد: تست شما مثبت بوده به این معنی که شما دچار یک بیماری جدی شده اید.

از طرفی این تست 99% دقیق است.

(احتمال اینکه شما بیمار باشید و جواب تست مثبت باشد =

احتمال اینکه شما بیمار نباشید و جواب تست منفی باشد = 0.99)

خبر خوب: این بیماری یک بیماری نادر است که از بین 10,000 نفر، تنها یک نفر بدان مبتلا می‌شود.

چرا این یک خبر خوب است؟

شانس اینکه واقعاً شما این بیماری را داشته باشید، چه قدر است؟

راه حل:

$$P(test \mid disease) = P(\neg test \mid \neg disease) = 0.99$$

$$P(disease) = 0.0001$$

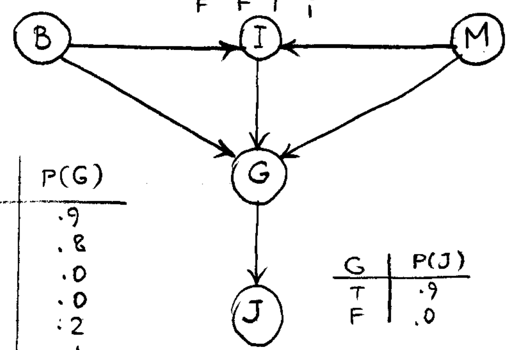
آنچه بیمار باید نگران آن باشد، $P(disease \mid test)$ است:

این خبر خوب از آن جهت خوب است که این مقدار وابسته به $P(disease)$ است (به طور مستقیم) که هرچه کمتر باشد این مقدار را کمتر می‌کند.

$$\begin{aligned}
P(disease \mid test) &= \frac{P(test \mid disease)P(disease)}{P(test)} \\
&= \frac{P(test \mid disease)}{P(test \mid disease)P(disease) + P(test \mid \neg disease)P(\neg disease)} \\
&= \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.01 \times 0.9999} \\
&= 0.009804
\end{aligned}$$

۳

مثال) شبکه بیزی زیر را در نظر بگیرید :



B	M	$P(I)$
T	T	.9
T	F	.5
F	T	.5
F	F	1

$P(B)$
.9

$P(M)$
.1

B	I	M	$P(G)$
T	T	T	.9
T	T	F	.8
T	F	T	.0
T	F	F	.0
F	T	T	.2
F	T	F	.1
F	F	T	.0
F	F	F	.0

G	$P(J)$
T	.9
F	.0

a) بدون در نظر گرفتن CPT ها کدام یک از عبارات زیر از ساختار شبکه نتیجه می شود؟

(i) $P(B,I,M) = P(B)P(I)P(M)$

(ii) $P(J|G) = P(J|G,I)$ ✓

(iii) $P(M|G,B,I) = P(M|G,B,I,J)$ ✓

b) مقدار زیر را محاسبه کنید :

$$P(b,i,\neg m,g,j) = P(b)P(i|b,\neg m)P(\neg m)P(g|i,b,\neg m)P(j|g)$$

$$= 0.9 \times 0.9 \times 0.5 \times 0.8 \times 0.9 = 0.2916$$

۴

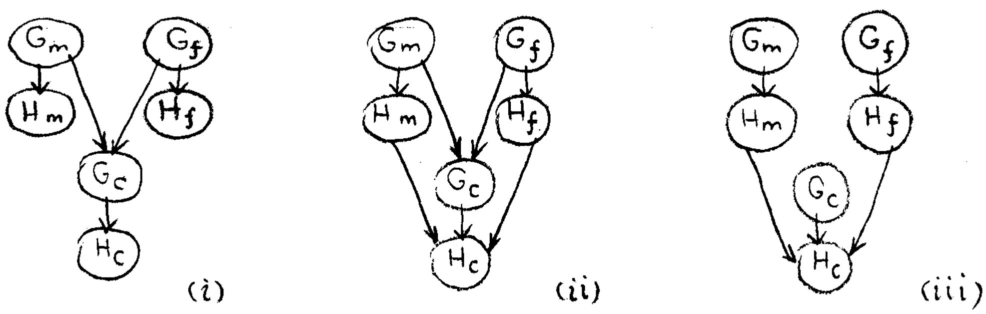
مثال)

Handedness

فرض کنید H_x یک متغیر تصادفی باشد که به راست دستی یا چپ دستی شخص x با مقادیر ممکن l یا r اشاره دارد. یک فرض معمول این است که راست دستی یا چپ دستی با یک مکانیزم ساده به ارث برده می‌شوند، یعنی احتمالاً ژن G_x وجود دارد که (با مقادیر r یا l) و با احتمال s موجب راست دستی یا چپ دستی می‌شود. به علاوه خود این ژن احتمالاً باید از یکی از والدهای شخص با احتمال کوچک m برای جهش در راست/چپ دستی به ارث برده شده باشد.

(a) کدام یک از شبکه‌های بیزی زیر ادعا می‌کند که

$$\mathbb{P}(\underbrace{G_f}_{\text{father}}, \underbrace{G_m}_{\text{mother}}, \underbrace{G_c}_{\text{child}}) = \mathbb{P}(G_f)\,\mathbb{P}(G_m)\,\mathbb{P}(G_c)$$



هیچ یالی بین آنها نیست.

(b) کدام یک از این شبکه‌ها ادعاهای استقلال سازگار با فرضیه را دارد؟ (i) و (ii)

تأکید بر عدم وجود یال هاست؛ (iii) تأکید بر استقلال ژن‌ها می‌کند که تناقض با فرضیه است.

(c) کدام یک از این شبکه‌ها بهترین توصیف فرضیه است؟

(i) : زیرا (ii) بین متغیرهای H پیوند برقرار کرده که در فرضیه وجود ندارد.

(d) CPT را برای G_{child} در شبکه‌های (i) یا (ii) رسم کنید و مقادیر آن را پر کنید.

* جهش‌های $r \to l$ و $l \to r$ وقتی والدها ژن‌های مختلف دارند، لغو می‌شود:

G_{mother}	G_{father}	$P(G_{child}=l \mid \dots)$	$P(G_{child}=r \mid \dots)$
l	l	$1-m$	m
l	r	0.5	0.5
r	l	0.5	0.5
r	r	m	$1-m$

(e) فرض کنید که $P(G_{father}=1) = P(G_{mother}=1) = x$ باشد.
در شبکه (i) یا (ii) عبارتی را برای $P(G_{child}=1)$ برحسب m و x تنها، با شرط‌گذاری بر روی گره‌های والدان استخراج نمایید.

$$
\begin{aligned}
P(G_{child}=1) &= \sum_{g_m, g_f} P(G_{child}=1 \mid g_m, g_f) P(g_m, g_f) \\
&= \sum_{g_m, g_f} P(G_{child}=1 \mid g_m, g_f) P(g_m) P(g_f) \\
&= (1-m)x^2 + 0.5\,x(1-x) + 0.5\,(1-x)x + m(1-x)^2 \\
&= x^2 - mx^2 + x - x^2 + m - 2mx + mx^2 \\
&= x + m - 2mx
\end{aligned}
$$

(f) تحت شرط تعادل ژنتیکی، انتظار داریم که توزیع ژن‌ها در خلال نسل‌ها مشابه باشد.
با استفاده از این مقدار x را محاسبه کنید و
با دانستن راست/چپ دستی در انسان‌ها توضیح دهید که چرا فرضیه توصیف شده در ابتدای این سؤال بایستی نادرست باشد.

تعادل یعنی $P(G_{child}=1)$ (بدون اطلاعات والد) باید برابر باشد با $P(G_{mother}=1) = P(G_{father}=1)$
یعنی:

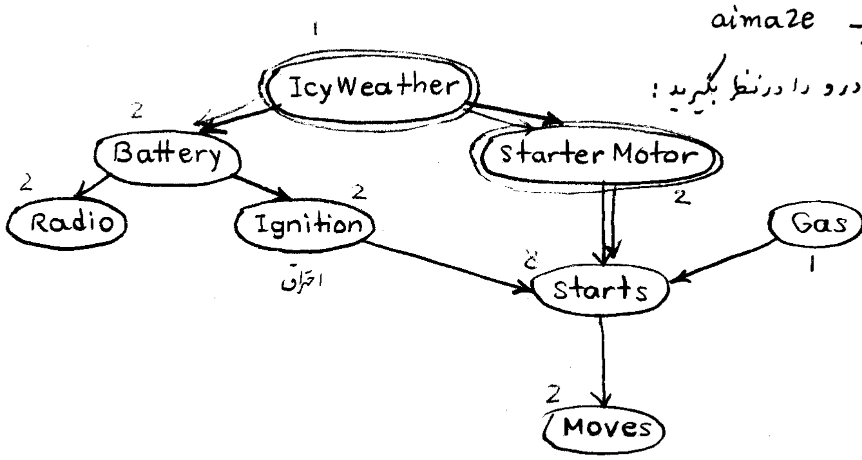
$$x + m - 2mx = x \Rightarrow x = 0.5$$

اما x در عمل $x \approx 0.08$ است.
بنابراین اشتباه در مدل تقارن وراثت است.
مطالعات فعلی نشان می‌دهد که راست/چپ دستی نتیجه‌ی یک ژن نیست و می‌تواند دراثر عوامل فرهنگی نیز باشد.

۵

مثال) مساله ۱۴_۱ کتاب aima2e

شبکه بیزی زیر برای عیب‌یابی خودرو را درنظر بگیرید:



(a) شبکه را با متغیرهای بولی IcyWeather و StarterMotor توسعه دهید.

← چون هوای یخی معمول هیچ یک از متغیرهای وابسته به خودرو نیست، والدی نخواهد داشت.

هوای یخی به طور مستقیم بر باتری و موتور استارت زننده تاثیر می‌گذارد.

StarterMotor یک پیش‌فرض دیگر برای starts است.

شبکه جدید روی شبکه قبلی با رنگ صورتی اضافه شده است.

(b) برای همه گره‌ها، جداول احتمال شرطی قابل استدلال به دست آورید.

احتمالات مستقل وابسته به نوع خودرو می‌توانند متفاوت باشند.

برای مثال:

prior احتمال پیشین

$P(icyWeather) = 0.05$

$P(battery \mid icyweather) = 0.95$, $P(battery \mid \neg icyweather) = 0.997$

$P(startermotor \mid icyweather) = 0.98$, $P(startermotor \mid \neg icyweather) = 0.999$

$P(radio \mid battery) = 0.9999$, $P(radio \mid \neg battery) = 0.05$

$P(ignition \mid battery) = 0.998$, $P(ignition \mid \neg battery) = 0.01$

$P(gas) = 0.995$

$P(starts \mid ignition \wedge startermotor \wedge gas) = 0.9999$, other entries 0.0.

$P(moves \mid starts) = 0.998.$

(c) با فرض اینکه هیچ رابطه‌ی استقلال شرطی بین متغیرهای تصادفی وجود نداشته باشد،
چند مقدار مستقل در توزیع احتمال پیوندی 8 متغیر بولی وجود دارد؟

با ۸ متغیر بولی، توزیع احتمال پیوندی دارای $2^8 - 1 = 255$ درایه‌ی مستقل دارد.

(d) چند مقدار احتمال مستقل در جداول شبکه‌ی فوق وجود دارد؟

با توجه به توپولوژی نشان داده شده در شکل ۱، مجموع تعداد درایه‌های مستقل CPT برابر است:

$$1 + 2 + 2 + 2 + 2 + 1 + 8 + 2 = 20$$

(e) توزیع شرطی برای Starts می‌تواند به صورت یک توزیع noisy-AND توصیف شود.
این خانواده را به طور کلی تعریف کنید و به توزیع noisy-OR آن را مرتبط کنید.

CPT برای Starts مجموعه‌ای از شرایط نزدیک به لازم که با هم تقریباً کافی هستند؛ را توصیف می‌کند.
یعنی همه‌ی درایه‌ها نزدیک به صفر هستند، بجز درایه‌ای که در آن همه‌ی شرایط true هستند.
این درایه کاملاً یک نیست (زیرا همیشه نقص‌هایی وجود دارند که به آنها فکر نکرده‌ایم)...
اما با اضافه کردن شرایط بیشتر به یک نزدیکتر می‌شود.
اگر یک Leak را به عنوان یک والد اضافی آن بیفزاییم، آنگاه این احتمال یک است اگر همه‌ی والدها درست باشند.
با استفاده از قانون دمورگان $A \wedge B \equiv \neg(\neg A \vee \neg B)$ می‌توانیم noisy-AND را به noisy-OR ربط کنیم.
یعنی noisy-AND مشابه noisy-OR است بجز اینکه قطبیت والد و فرزند معکوس می‌شود.
در noisy-OR داریم:

$$P(Y = true \mid x_1, x_2, \ldots, x_k) = 1 - \prod_{\{i : x_i = true\}} q_i$$

که در آن q_i احتمال این است که حضور نا‌امن والد موفق نشود در true شدن فرزند تأثیر بگذارد.

در noisy-AND می‌توانیم بنویسیم:

$$P(Y = true \mid x_1, x_2, \ldots, x_k) = \prod_{\{i : x_i = false\}} r_i$$

که در آن r_i احتمال این است که غیاب نا‌امن والد موفق نشود در false شدن فرزند تأثیر بگذارد.

۶

مثال) مسأله 2-14 کتاب aimaze

در یک ایستگاه انرژی هسته‌ای، زنگ خطری وجود دارد که در هنگام تجاوز درجه‌ی دما از یک آستانه‌ی داده شده به صدا در می‌آید. درجه، دمای هسته را اندازه‌گیری می‌کند.

متغیرهای بولی A (صداهای زنگ خطر)، F_A (زنگ خطر خراب است) و F_G (درجه خراب است) و متغیرهای چند مقداری G (مقدار خوانده شده از درجه) و T (دمای واقعی هسته) را در نظر بگیرید.

(a) یک شبکه‌ی بیزی برای این حوزه رسم کنید، با توجه به این که درجه به احتمال بیشتر خراب می‌شود هرگاه دمای هسته بسیار بالا رود.

(b) آیا این شبکه یک polytree است؟

(c) فرض کنید تنها دو دمای اندازه‌گیری شده و واقعی وجود دارد: عادی normal و بالا high

احتمال اینکه درجه دمای درستی را نشان دهد هنگامی که کار می‌کند x و

هنگامی که خراب است y است.

جدول احتمال شرطی شرط با G را رسم کنید.

(d) فرض کنید که زنگ خطر به درستی کار می‌کند، مگر اینکه خراب باشد. در این حالت صدایی نخواهد داشت.

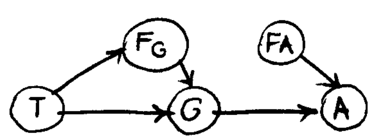
جدول احتمال شرطی شرط با A را رسم کنید.

(e) فرض کنید که زنگ خطر و درجه هر دو کار می‌کند و زنگ خطر به صدا در آید.

عبارتی را برای احتمال اینکه دمای هسته بسیار بالاست برحسب احتمالات شرطی مختلف در این شبکه محاسبه کنید.

راه حل:

(a)



نکته کلیدی: گره‌های خرابی، والد های گره‌های سنسوری هستند.

و گره‌ی دما والد گره‌های درجه و خرابی درجه است.

(b) این شبکه نمی‌تواند polytree باشد، زیرا دما به دو طریق بر روی درجه اثر می‌گذارد.

(c) CBT برای G در زیر نشان داده شده است.

بیان مسأله اندکی زیرکانه است: F_G یعنی "کار نمی‌کند" و $\neg F_G$ یعنی "کار می‌کند"

	T = normal		T = high	
	F_G	$\neg F_G$	F_G	$\neg F_G$
G = normal	y	x	$1-y$	$1-x$
G = high	$1-y$	$1-x$	y	x

(d) CBT برای A :

	G = normal		G = high	
	F_A	$\neg F_A$	F_A	$\neg F_A$
A	0	0	0	1
$\neg A$	1	1	1	0

(e) خلاصه‌نویسی T=high با t (یا T)

G=high با g (یا G)

هدف محاسبه‌ی احتمال $P(T|A, \neg F_G, \neg F_A)$ است.

$A \wedge \neg F_G \wedge \neg F_A \Longrightarrow T$

از آنجا که رفتار زنگ خطر قطعی است، می‌توانیم استدلال کنیم که اگر زنگ خطر کار کند و به صدا درآید، G باید high باشد.

از آنجا که F_A و A، از T، d-separated هستند، تنها لازم است که $P(T|\neg F_G, G)$ را محاسبه کنیم. راههای مختلفی برای این کار وجود دارد. "خوش بیانه"ترین راه توجه به این است که درایه‌های CPT به ما مقدار $P(G|T, \neg F_G)$ را می‌دهند که آن پیشنهاد می‌کند از قاعده‌ی بیز تعمیم یافته استفاده کنیم تا G و T را با $\neg F_G$ به صورت پس زمینه جایگزین نماییم:

$$P(T|\neg F_G, G) \propto P(G|T, \neg F_G)\, P(T|\neg F_G)$$

سپس مجدداً بر روی جمله‌ی آخر قاعده‌ی بیز را به کار می‌گیریم:

$$P(T|\neg F_G, G) \propto P(G|T, \neg F_G)\, P(\neg F_G|T)\, P(T)$$

رابطه‌ی مشابهی برای $\neg T$ برقرار است:

$$P(\neg T|\neg F_G, G) \propto P(G|\neg T, \neg F_G)\, P(\neg F_G|\neg T)\, P(\neg T)$$

با نرمال‌سازی به دست می‌آوریم:

$$P(T|\neg F_G, G) = \frac{P(G|T, \neg F_G)P(\neg F_G|T)P(T)}{P(G|T, \neg F_G)P(\neg F_G|T)P(T) + P(G|\neg T, \neg F_G)P(\neg F_G|\neg T)P(\neg T)}$$

روش سیستماتیک :

$$P(T|\neg F_G , G) = \frac{P(T , \neg F_G , G)}{P(\neg F_G , G)} = \frac{P(T, \neg F_G , G)}{P(T, G , \neg F_G) + P(\neg T , G , \neg F_G)}$$

اما طبق قاعده‌ی زنجیری

$$P(T, \neg F_G , G) = P(T)\, P(\neg F_G | T)\, P(G | T, \neg F_G)$$

$$P(\neg T, \neg F_G , G) = P(\neg T) P(\neg F_G | \neg T) P(G | \neg T, \neg F_G)$$

با فرض $P(T)=p$ ، $P(F_G|T)=g$ و $P(F_G|\neg T)=h$ خواهیم داشت :

$$P(T|\neg F_G , G) = \frac{p(1-g)(1-x)}{p(1-g)(1-x) + (1-p)(1-h)x}$$

(۷)

مثال) مسأله 18-13 کتاب aima2e

دسته‌بندی متون، عمل انتساب یک سند داده شده به یکی از دسته‌های مشخص، براساس محتوای متن است.

برای این کار اغلب از مدل‌های Naive Bayes استفاده می‌شود.

در این مدل‌ها، متغیر پرس‌وجو (query variable)، دستهٔ سند و

متغیرهای effect، حضور یا عدم حضور هر یک از کلمات زبان است.

فرض این است که کلمات به طور مستقل در اسناد ظاهر می‌شوند که فراوانی آنها به‌وسیلهٔ دستهٔ سند تعیین می‌شود.

(a) به طور دقیق تشریح کنید که چگونه چنین مدلی می‌تواند ساخته شود.

فرض کنید که "داده‌های آموزشی" training data به صورت مجموعه‌ای از اسناد که تاکنون به دسته‌ها نسبت داده شده‌اند، داده شده‌اند.

(b) به طور دقیق توضیح دهید که چگونه یک سند جدید دسته‌بندی می‌شود.

(c) آیا فرض استقلال، منطقی است؟ توضیح دهید.

راه حل:

(a) این مدل از موارد زیر تشکیل می‌شود:

$\mathbb{P}(Category)$ احتمالات پیشین رده‌ها

$\mathbb{P}(Word_i | Category)$ که در آن

$Word_i = true$ اگر و فقط اگر سند مورد نظر حاوی i امین کلمه‌ی لغت‌نامه باشد.

برای هر دسته‌ی c، $\mathbb{P}(Category = c)$ به صورت کسری از همهٔ اسنادکه در دستهٔ c قرار دارند برآورد می‌شود.

به طور مشابه $\mathbb{P}(Word_i = true | Category = c)$ یا همان $\mathbb{P}(word_i | c)$

به صورت کسری از اسناد دسته‌ی c که حاوی کلمه‌ی i هستند، برآورد می‌شود.

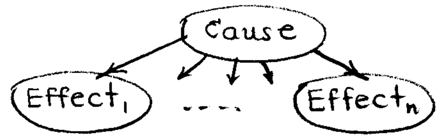
(b) با استفاده از ترتیب ساختار استقلال شرطی مدل خواهیم داشت:

$$\mathbb{P}(Category) = \alpha\, \mathbb{P}(Category, word_1, \ldots, word_n)$$
$$= \alpha\, \mathbb{P}(Category) \prod_i \mathbb{P}(word_i | Category)$$

(c) فرض استقلال به وضوح در عمل نقض می‌شود.

برای مثال زوج کلمات "artificial intelligence" در متون مرتبط با هم ظاهر می‌شوند.

Naive Bayes Model



$$\mathbb{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbb{P}(Cause) \prod_i \mathbb{P}(Effect_i \mid Cause)$$

Bayesian Networks

شبکه‌ی بیزی: نمادگذاری گرافیکی برای تأکید بر استقلال شرطی

و بنابراین برای بیان فشرده‌ی توزیع های توأم کامل

نحو: به ازای هر متغیر یک گره

گراف جهت دار بدون دور (هر اتصال ≈ تأثیر مستقیم)

$\mathbb{P}(X_i \mid Parent(X_i))$ توزیع شرطی برای هر گره با داشتن والدهای آن (CPT)

Conditional Probability Table

CBT برای متغیر بولی X_i با k والد بولی، 2^k سطر برای ترکیب های مختلف والدها دارد.

هر سطر نیاز به یک عدد p برای $X_i=true$ دارد. (عدد برای $X_i=false$ ، $1-p$ خواهد بود)

اگر حداکثر تعداد والد گره ها برابر با k باشد، شبکه‌ی کامل به $O(n 2^k)$ عدد نیاز دارد.

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