به نام یگانه معبود بخشنده مهربان

شبکه های باور بیزین

Bayesian Belief Networks

تخمین چگالی

- در روشهای پارامتری، تا کنون تنها مدلهای توزیع ساده را درنظر میگرفتیم.
- اما در مواقع زیر مشکل خواهیم داشت:
  - توزیع داده پیچیده: مدلهای ساده نمیتوانند توزیعهای پیچیده را بخوبی نمایندگی کنند.
  - تعداد داده کم: برای تخمین دقیق پارامترها داده کافی موجود نیست.
- برای بعد ورودی زیاد، روشهای غیرپارامتری نیز معمولاً ناکارآمدند.
- چگونه توزیعهای پیچیده با ابعاد زیاد را یاد بگیریم؟
- یک راه حل:
  - تجزیه توزیع با استفاده از روابط استقلال مشروط ≡ تبدیل مسأله تخمین چگالی به یک مجموعه مسأله تخمین چگالی ساده تر
- تجزیه توزیع ها تحت فرض استقلال مشروط ایده زیربنایی شبکه‌های باور بیزین است.

√480

مدلسازی عدم قطعیت با احتمالات

- **توزیع توأم کامل (Full joint distribution):**
  - فرض کنید $\mathbf{X}=\{X_1,X_2,\ldots,X_d\}$ کلیه متغیرهای تصادفی تعریف کننده مسأله باشند. توزیع توأم کامل $P(X_1,X_2,\ldots,X_d)$ یا $P(\mathbf{X})$ است.
- **توزیع توأم کامل برای انجام هر نوع استنتاج احتمالاتی کافی است:**
  - یافتن احتمالات توأم یک زیرمجموعه از متغیرها: $P(X_1,X_2,X_3)$ یا $P(X_1,X_{10})$
  - یافتن احتمالات شرطی: $P(X_1 \mid X_2=True, X_3=False)$
- **مثال:**
  - بیماری: ذات الریه (pneumonia)
  - علایم بیمار (معاینه و آزمایش): تب (fever)، سرفه (cough)، رنگ پریدگی (paleness)، تعداد گلوبولهای سفید (WBC count) غیرطبیعی، درد قفسه سینه (chest pain)، ... .
  - بازنمایی بیمار: علایم و بیماری بصورت متغیرهای تصادفی بازنمایی میشوند.
  - هدف: توصیف یک توزیع چند متغیره که بیانگر ارتباط بین علایم و بیماری باشد. به بیان دیگر هدف طراحی روشهایی برای یادگیری این مدل چند متغیره و استنتاج با آن است.

حاشیه‌ای کردن

- توزیع توأم برای تعدادی متغیر:
  - تمام احتمالات انتساب مقادیر ممکن به متغیرها را بیان میکند.
- حاشیه ای کردن (Marginalization): حذف تعدادی از متغیرها از توزیع توأم با جمع بستن روی تمام مقادیر آن متغیرها
- مثال:

$\mathbf{P}(pneumonia, WBCcount)$

WBCcount

Pneumonia		*high*	*normal*	*low*	$\mathbf{P}(Pneumonia)$
	True	0.0008	0.0001	0.0001	0.001
	False	0.0042	0.9929	0.0019	0.999
	$\mathbf{P}(WBCcount)$	0.005	0.993	0.002	

√456

احتمال شرطی و استنتاج

- احتمالات توأم می‌توانند بر حسب احتمالات شرطی نوشته شوند:

$$P(A,B) = P(A \mid B)P(B) \quad \textbf{(product rule)}$$

$$P(X_1, X_2, \ldots X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots X_{i-1}) \quad \textbf{(chain rule)}$$

- احتمالات شرطی برای انواع استنتاج‌های احتمالاتی مفیدند. مثال:

$$P(Pneumonia = True \mid Fever = True, WBCcount = high, Cough = True)$$

- هر پرسشی می‌تواند با استفاده از توزیع توأم کامل پاسخ داده شود:
    - توزیع توأم زیرمجموعه‌ای از متغیرها با حاشیه ای کردن توزیع توأم کامل بدست می‌آید:

$$P(A=a, C=c) = \sum_i \sum_j P(A=a, B=b_i, C=c, D=d_j)$$

    - احتمال شرطی روی یک مجموعه از متغیرها، نسبت دو توزیع توأم جزئی است:

$$P(D=d \mid A=a, C=c) = \frac{P(A=a, C=c, D=d)}{P(A=a, C=c)}$$

$$= \frac{\sum_i P(A=a, B=b_i, C=c, D=d)}{\sum_i \sum_j P(A=a, B=b_i, C=c, D=d_j)}$$

√504

استنتاج

- هر احتمال توأم می‌تواند براساس قانون زنجیره‌ای بصورت حاصل ضرب احتمالات شرطی نوشته شود:

$$P(X_1, X_2, \ldots X_n) = P(X_n \mid X_{1,} \ldots X_{n-1}) P(X_{1,} \ldots X_{n-1})$$

$$= P(X_n \mid X_{1,} \ldots X_{n-1}) P(X_{n-1} \mid X_{1,} \ldots X_{n-2}) P(X_{1,} \ldots X_{n-2})$$

$$= \prod_{i=1}^{n} P(X_i \mid X_{1,} \ldots X_{i-1})$$

- تعریف توزیع برحسب احتمالات شرطی اغلب ساده تر است:

$$\mathbf{P}(Fever \mid Pneumonia = T)$$

$$\mathbf{P}(Fever \mid Pneumonia = F)$$

- توزیع توأم کامل: توزیع توأم روی تمام متغیرهای تعریف کننده دامنه مسأله
  - این توزیع برای نمایش کامل دامنه مسأله و انجام هر نوع استنتاج احتمالاتی کفایت میکند.

مدلسازی عدم قطعیت با احتمالات

- مسائل:
  - پیچیدگی حافظه: ذخیره درایه‌های توزیع توأم کامل $O(d^n)$ فضا نیاز دارد.
    n: تعداد متغیرهای تصادفی، **d**: تعداد مقادیر هر متغیر
  - پیچیدگی استنتاج: محاسبه پاسخ برخی از پرسشها $O(d^n)$ زمان نیاز دارد.
  - جمع آوری اطلاعات: تعیین مقادیر تمام درایه ها کاری دشوار است.
- مثال: برای مسأله بیماری ذات الریه
  - پیچیدگی حافظه: نیاز به تعریف 2*2*2*3*2=48 احتمال داریم.

Pneumonia (2 values: T,F), Fever (2: T,F), Cough (2: T,F), WBCcount (3: high, normal, low), paleness (2: T,F)

  - پیچیدگی زمانی: مثلاً برای محاسبه احتمال Pneumonia=T از روی توزیع توأم کامل نیاز به 2*2*3*2=24 عمل جمع داریم.

$$P(Pneumonia = T) =$$

$$= \sum_{i \in T,F} \sum_{j \in T,F} \sum_{k=h,n,l} \sum_{u \in T,F} P(Fever = i, Cough = j, WBCcount = k, Pale = u)$$

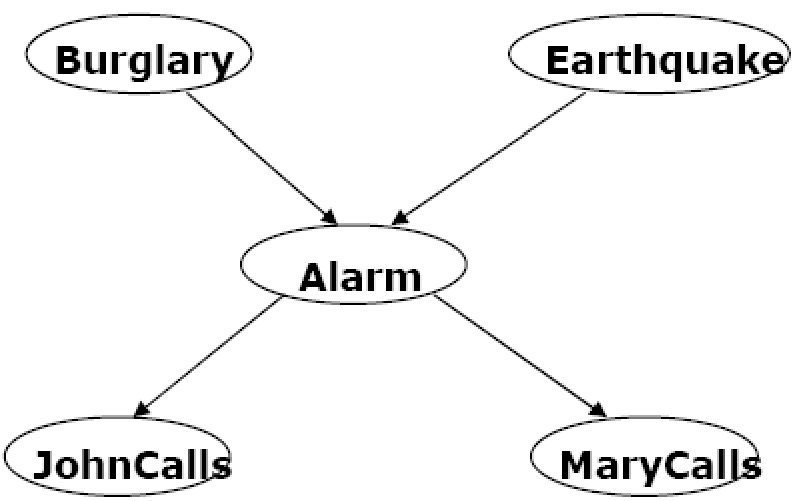
√484

شبکه‌های باور بیزین (BBNs)

- شبکه‌های (باور) بیزین:
  - توزیع توأم کامل روی متغیرهای تصادفی را بصورت فشرده تر، یعنی با تعداد پارامتر کمتر، نمایش میدهند.
  - از استقلال مشروط و حاشیه‌ای بین متغیرهای تصادفی بهره میبرند.
- استقلال A و B : $P(A,B) = P(A)P(B)$
- استقلال A و B مشروط بر C: $P(A,B \mid C) = P(A \mid C)P(B \mid C)$

$$P(A \mid C,B) = P(A \mid C)$$

- مثال: سیستم دزدگیر
  - یک سیستم هشدار ضدسرقت در منزل نصب کرده‌ایم، که ممکن است زمین لرزه های خفیف هم آن را به صدا درآورد. دو همسایه بنام John و Mary که یکدیگر را نمیشناسند با شنیدن صدای هشدار احتمالاً با ما تماس میگیرند.
  - میخواهیم توزیع احتمال رخداد ها را نمایش دهیم.



روابط علی

√480

شبکه‌های باور بیزین (BBNs)

1- یک گراف جهت دار بدون دور (DAG):

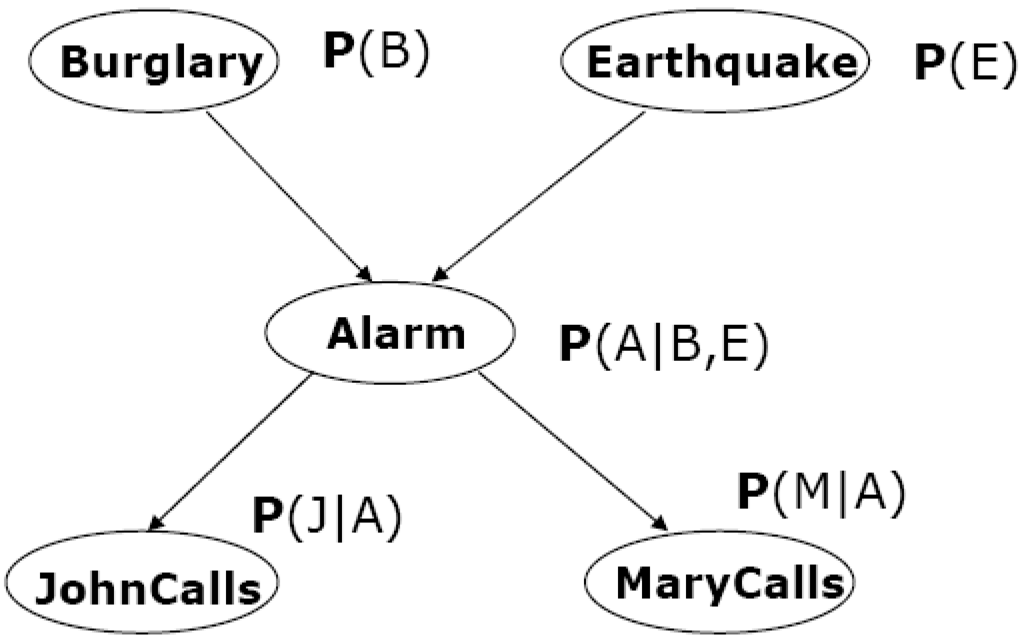
- گره‌ها (nodes): متغیرهای تصادفی

Burglary, Earthquake, Alarm, Mary calls and John calls

- یال‌ها (links): وابستگی‌های جهت دار (علی) بین متغیرها

احتمال اعلام هشدار از وقوع زمین لرزه تأثیر میپذیرد

احتمال تماس گرفتن یک همسایه از اعلام هشدار تأثیر میپذیرد



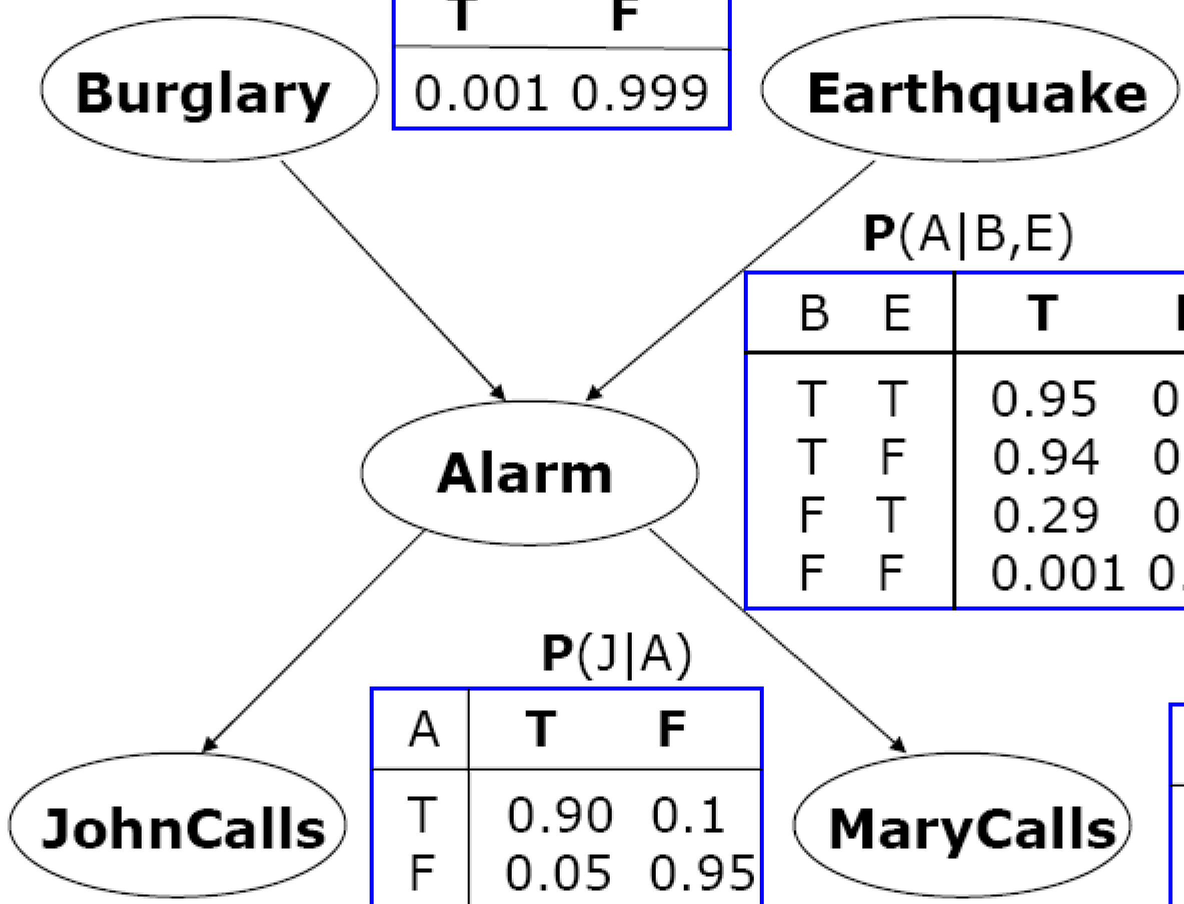
2- توزیع‌های شرطی محلی:

- متغیرها را با والدهایشان مرتبط میکنند.

$\sqrt{484}$

شبکه‌های باور بیزین (BBNs)

- توزیع‌های شرطی محلی:



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

P(A|B,E)

B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

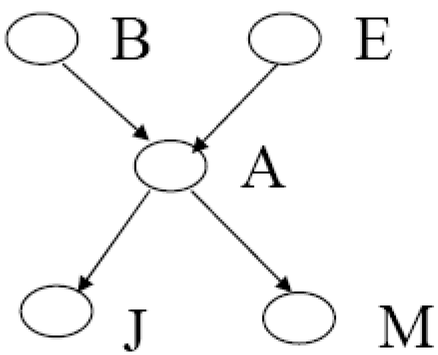
A	T	F
T	0.7	0.3
F	0.01	0.99

توزیع توأم کامل در شبکه‌های باور بیزین

- توزیع توأم کامل برحسب توزیعهای محلی شرطی بر اساس قاعده زنجیره‌ای با در نظر گرفتن استقلالهای مشروط بدست می‌آید:

$$\mathbf{P}(X_1, X_2, ..., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

- مثال:



  - انتساب‌های زیر از مقادیر به متغیرهای تصادفی مفروض است: $B = T, E = T, A = T, J = T, M = F$
  - احتمال توأم این انتسابها میشود:

$$P(B = T, E = T, A = T, J = T, M = F) =$$

$$P(B = T)P(E = T)P(A = T \mid B = T, E = T)P(J = T \mid A = T)P(M = F \mid A = T)$$

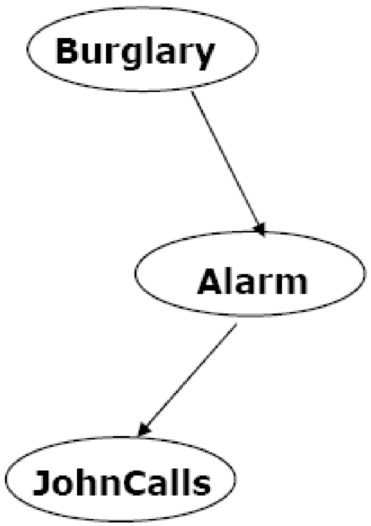
- شبکه بیزین توزیع توأم کامل روی متغیرهای تصادفی را بصورت فشرده‌تر، با استفاده از ضرب توزیعهای محلی شرطی، نمایش میدهد.
- ساختار گرافی شبکه استقلالهای شرطی و به تبع آن نحوه تجزیه توزیع توأم کامل را مشخص میکند.

√484

استقلال در شبکه‌های باور بیزین

3- ساختارهای پایه مرتبط با استقلال‌ها در گراف:

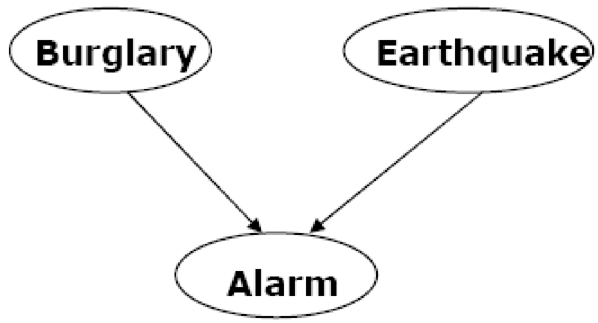
1.



$$P(J \mid A, B) = P(J \mid A)$$

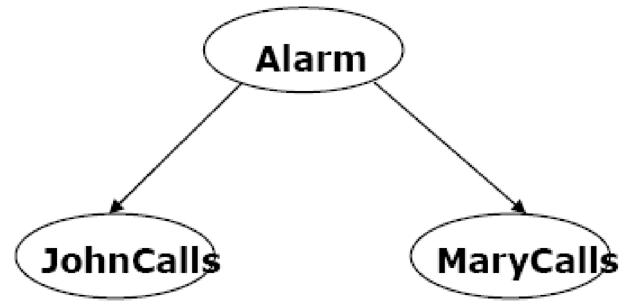
$$P(J, B \mid A) = P(J \mid A)P(B \mid A)$$

2.



$$P(B, E) = P(B)P(E)$$

3.



$$P(J \mid A, M) = P(J \mid A)$$

$$P(J, M \mid A) = P(J \mid A)P(M \mid A)$$

1. تماس جان به شرط اعلام خطر مستقل از سرقت است.
2. سرقت مستقل از زمین لرزه است (با ندانستن اعلام خطر).
   اما این دو به شرط اعلام خطر به هم وابسته میشوند.
3. تماس مری به شرط اعلام خطر مستقل از تماس جان است.

√484

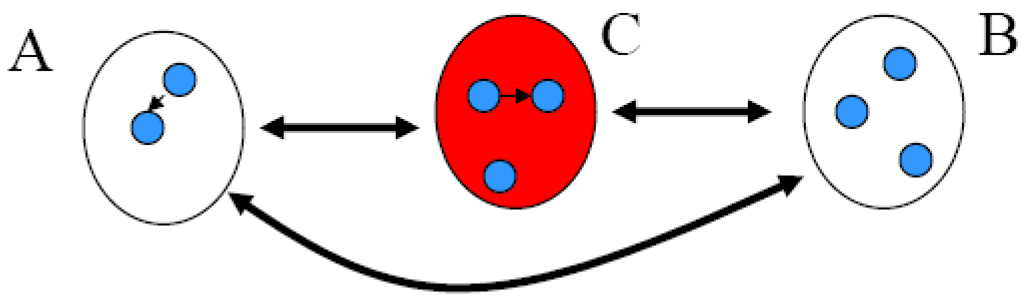
استقلال در شبکه‌های باور بیزین

- شبکه‌های بیزین استقلال مشروط بین متغیرها یا مجموعه متغیرهای دور از هم را مدل میکنند.
- بدین منظور از معیاری بنام جدایی جهت دار (d-separation) در گراف استفاده میشود.
- جدایی جهت دار در گراف
  - فرض کنید X، Y، و Z سه مجموعه از گره های گراف باشند.
  - اگر X و Y بوسیله Z بصورت جهت دار از هم جدا شوند آنگاه X و Y مشروط بر Z از یکدیگر مستقل خواهند بود.
- جدایی جهت دار (d-separation)
  - A از B با داشتن C بصورت جهت دار جدا میشود، اگر هر مسیر بدون جهت بین A و B مسدود (blocked) باشد.
- انسداد مسیر (path blocking)
  - در سه حالت تعریف میشود، که توسعه یافته ساختارهای پایه استقلال در گراف هستند.

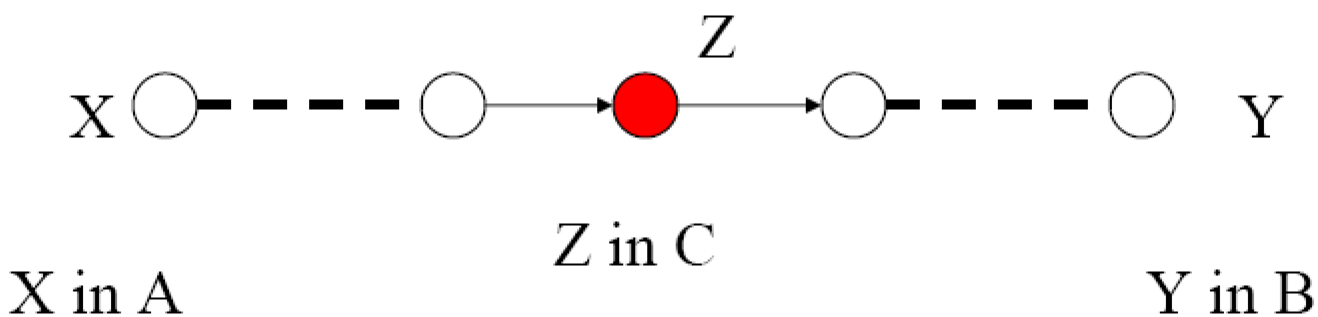
√484

مسدود بودن مسیرهای بدون جهت

- A از B با داشتن C بصورت جهت دار جدا میشود، اگر هر مسیر بدون جهت بین A و B مسدود (blocked) باشد. به گره‌های C مشاهدات (evidence) نیز گفته میشود.

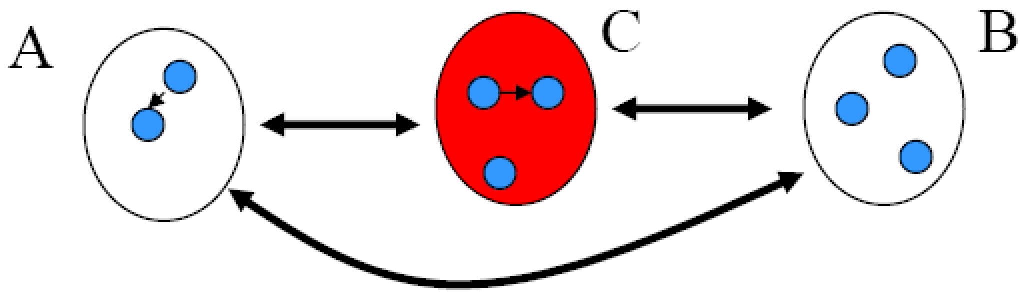


  - حالت 1 – انسداد مسیر با یک زیرساختار خطی (head-to-tail)

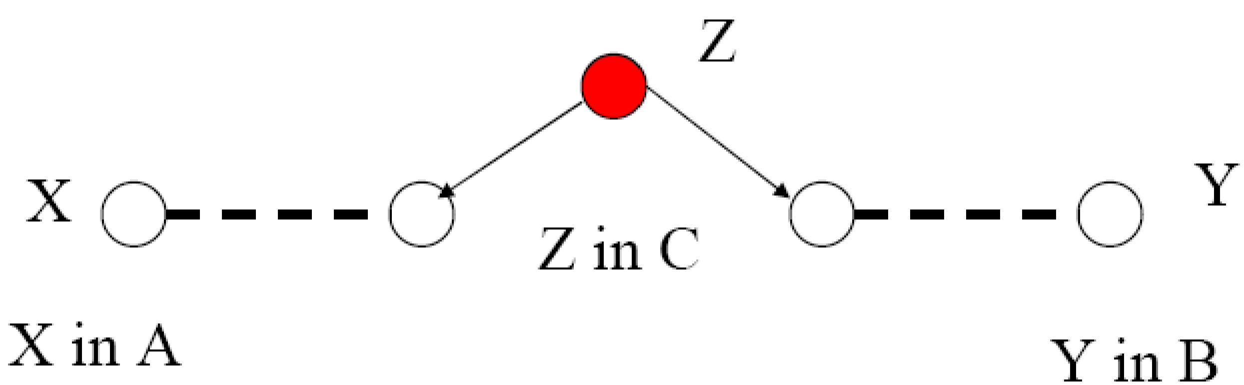


مسدود بودن مسیرهای بدون جهت

- **A از B با داشتن C بصورت جهت دار جدا میشود، اگر هر مسیر بدون جهت بین A و B** **مسدود (blocked)** **باشد.**

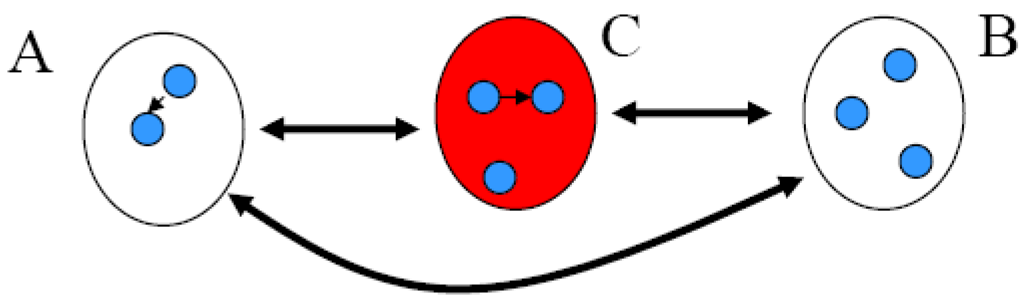


  - **حالت 2 – انسداد مسیر با یک زیرساختار ^ گونه (tail-to-tail)**

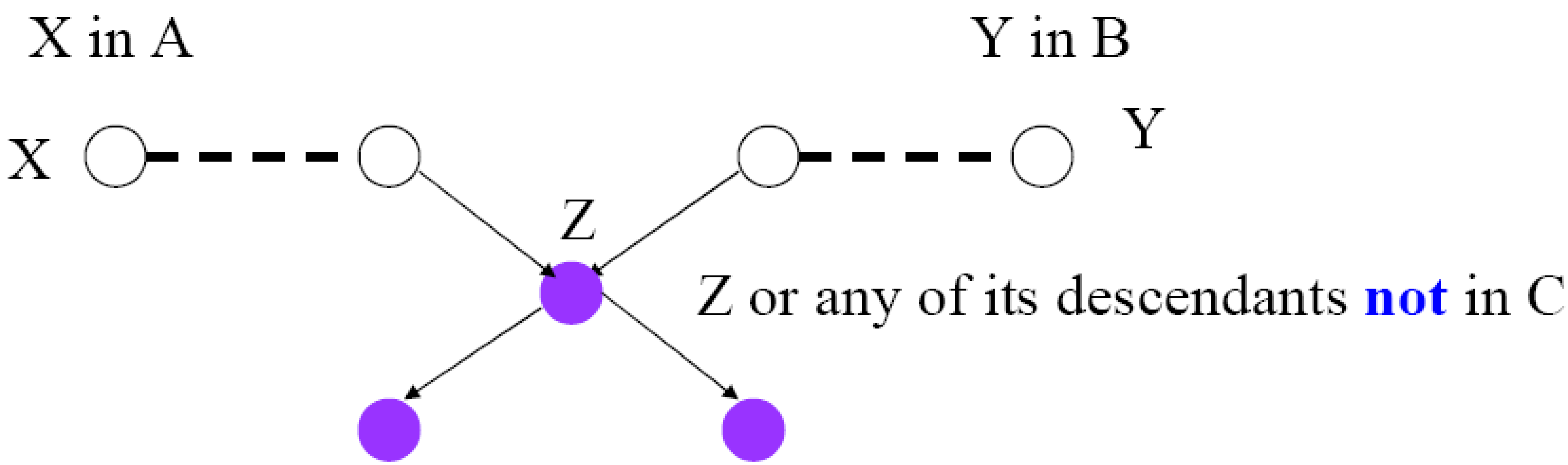


مسدود بودن مسیرهای بدون جهت

- A از B با داشتن C بصورت جهت دار جدا میشود، اگر هر مسیر بدون جهت بین A و B مسدود (blocked) باشد.



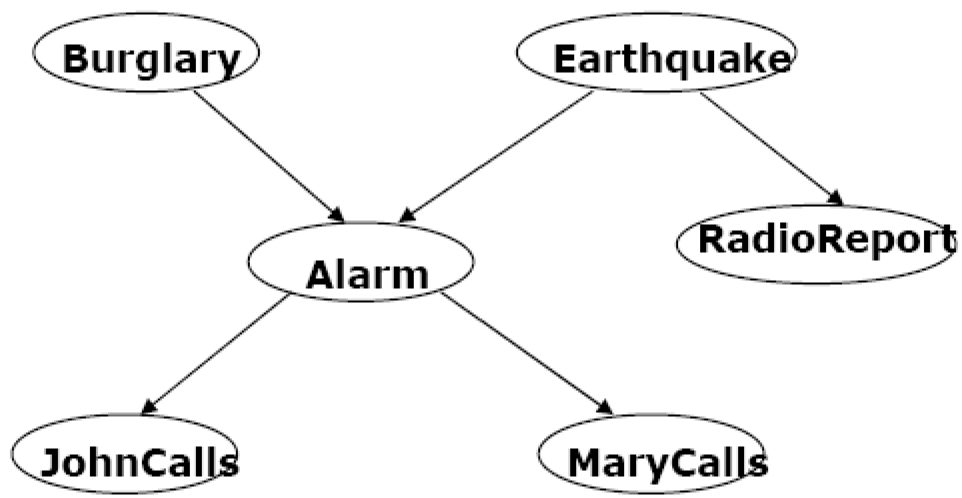
  - حالت 3 – انسداد مسیر با یک زیرساختار v گونه (head-to-head)



√504

استقلال در BBN ها

- مثال:



- آیا زمین لرزه و سرقت با دانستن تماس مری مستقل هستند؟ خیر
- آیا سرقت و تماس مری مستقل هستند؟ (با ندانستن اعلام خطر) خیر
- آیا سرقت و خبر رادیو با دانستن زمین لرزه مستقل هستند؟ بله
- آیا سرقت و خبر رادیو با دانستن تماس مری مستقل هستند؟ خیر

√456

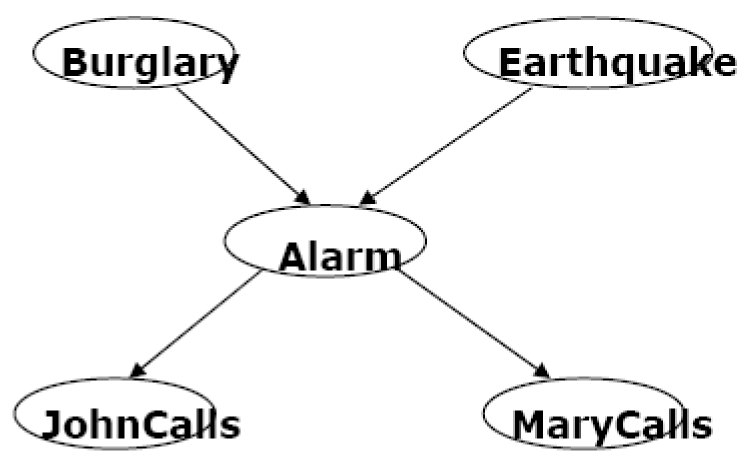
مسأله پیچیدگی پارامترها

- در BBN توزیع توأم کامل بصورت زیر تعریف میشود:

$$\mathbf{P}(X_1, X_2, \ldots, X_n) = \prod_{i=1,\ldots n} \mathbf{P}(X_i \mid pa(X_i))$$

- چه تعداد پارامتر باید ذخیره شود؟
- در مثال دزدگیر (پنج متغیر دودویی):
  - تعداد پارامترهای توزیع توأم کامل؟ $2^5 = 32$
  - یک پارامتر آزاد است (بنابر خاصیت احتمالاتی). پس داریم: $2^5 - 1 = 31$
  - تعداد پارامترهای BBN؟ (اسلاید بعد) $2^3 + 2(2^2) + 2(2) = 20$
  - یک پارامتر (در هر سطر) برای هر توزیع شرطی محلی آزاد است. پس داریم:

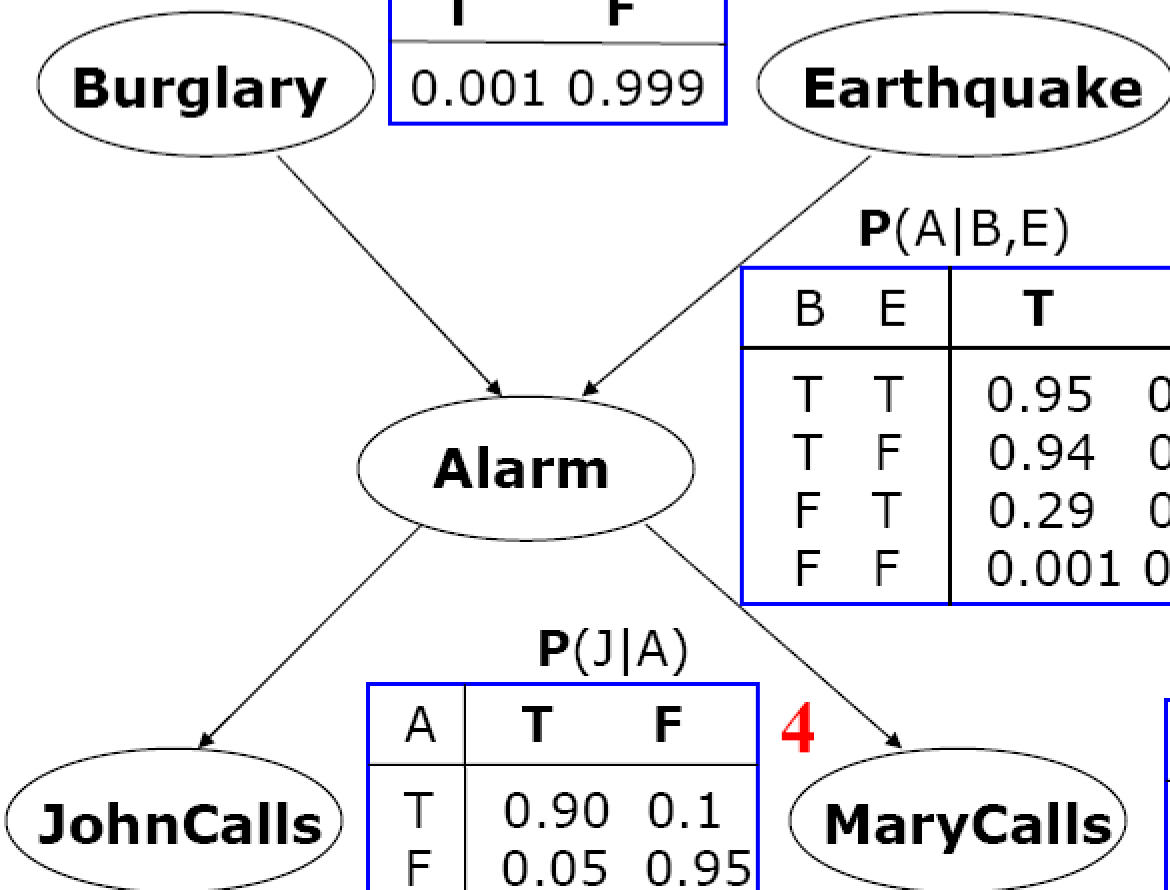
$$2^2 + 2(2) + 2(1) = 10$$



شبکه‌های باور بیزین (BBNs)

اسلاید قبل

- توزیع‌های شرطی محلی:



P(B) — 2

T	F
0.001	0.999

P(E) — 2

T	F
0.002	0.998

P(A|B,E) — 8

B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

P(J|A) — 4

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A) — 4

A	T	F
T	0.7	0.3
F	0.01	0.99

√450

مسأله جمع آوری اطلاعات مدل

- ساختار BBN معمولاً منعکس کننده روابط علی موجود در مسأله است. لذا به آنها شبکه های علی (causal networks) نیز گفته میشود.

- این روابط معمولاً بسادگی توسط افراد خبره حوزه مورد نظر (domain experts) تعیین میشود.

- پارامترهای مدل معرف توزیع های شرطی مرتبط کننده یک متغیر تصادفی با والدهایش است.

- تعداد پارامترها نسبت به توزیع توأم کامل بسیار کمتر است. لذا تخمین آنها توسط افراد خبره یا یادگیری خودکار از روی نمونه‌های داده ساده تر است.

استنتاج در شبکه‌های بیزین

- معمولاً علاقه مند به انجام استنتاج های گوناگون در BBN هستیم:
  - استنتاج برای تشخیص (diagnostic) (رسیدن از معلول به علت):

$$\mathrm{P}(Burglary \mid JohnCalls = T)$$

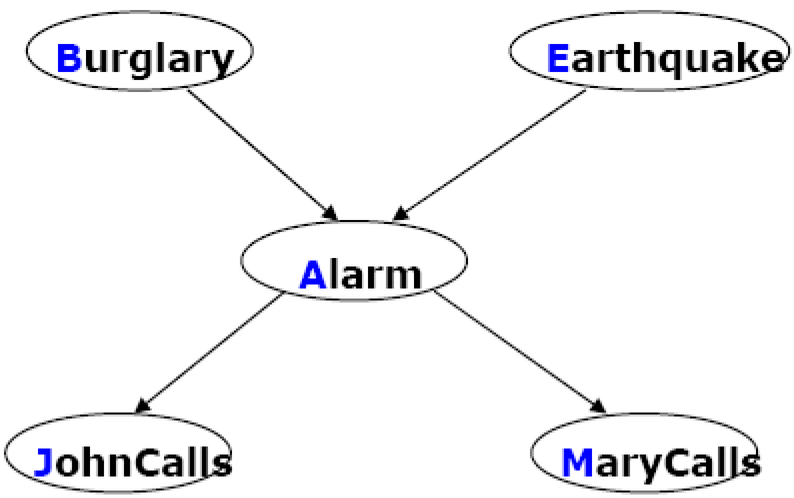
  - استنتاج برای پیشگویی (prediction) (رسیدن از علت به معلول):

$$\mathrm{P}(JohnCalls \mid Burglary = T)$$

  - سایر پرسشهای احتمالاتی (پرسشهایی روی توزیع توأم):

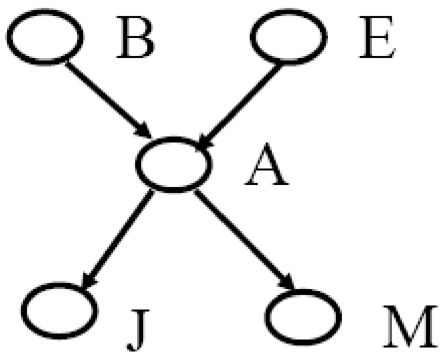
$$\mathrm{P}(Alarm)$$

- مسأله: آیا میتوان از استقلالها برای ساخت الگوریتمهای خاص برای تسریع استنتاج استفاده کرد؟
  - متأسفانه استنتاج دقیق در BBN ها یک مسأله NP مشکل است. استنتاج تقریبی نیز همین گونه است. اما در بسیاری مواقع میتوان بهبودهای قابل توجهی بدست آورد. مثال: محاسبه $P(J = T)$



√462

استنتاج در شبکه‌های بیزین



- مثال: محاسبه $P(J = T)$

رهیافت کورکورانه:

- جمع بستن توزیع توأم کامل روی تمام متغیرهای مقداردهی نشده
- بیان توزیع توأم کامل بصورت ضرب توزیعهای شرطی

$$P(J = T) =$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} \sum_{m\in T,F} P(B = b, E = e, A = a, J = T, M = m)$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} \sum_{m\in T,F} P(J = T \mid A = a) P(M = m \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e)$$

- هزینه محاسباتی:
  - تعداد جمع‌ها: **2*(2*(2*1+1)+1)+1=15**
  - تعداد ضرب‌ها: **16*4=64**

√504

استنتاج در شبکه‌های بیزین

- بعنوان مثال برای محاسبه عبارت زیر

$$P(B=T \mid J=T) = \frac{P(B=T, J=T)}{P(J=T)}$$

بخشی از محاسبات برای صورت و مخرج مشترک است که میتوانند برای صرفه جویی در محاسبات به اشتراک گذاشته شوند.

- برای برخی پرسشها نگهداری و ترتیب مناسب برای انجام ضربها و جمعها محاسبات را کاهش میدهد:

$$\mathbf{P}(B \mid J=T) = \frac{\mathbf{P}(B, J=T)}{P(J=T)} = \alpha \mathbf{P}(B, J=T)$$

- روش تجزیه بازگشتی (Recursive decomposition)
  - انجام محاسبات (جمعها و ضربها در میان یکدیگر) پیش از استنتاج
- روش حذف متغیر (Variable elimination)
  - انجام محاسبات (جمعها و ضربها در میان یکدیگر) با حذف کردن یک به یک متغیرها در حین استنتاج

√444

روش حذف متغیر

- مثال: محاسبه $P(J=T)$ با حذف یک به یک متغیرها
  - میتوان ترتیبهای مختلفی در نظر گرفت. با ترتیب M، E، B، A داریم:

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F}\sum_{m\in T,F} P(J=T\,|\,A=a)P(M=m\,|\,A=a)P(A=a\,|B=b,E=e)P(B=b)P(E=e)$$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F} P(J=T\,|\,A=a)P(A=a\,|B=b,E=e)P(B=b)P(E=e)\left[\sum_{m\in T,F} P(M=m\,|\,A=a)\right]$$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F} P(J=T\,|\,A=a)P(A=a\,|\,B=b,E=e)P(B=b)P(E=e)\quad 1$$

$$= \sum_{a\in T,F}\sum_{b\in T,F} P(J=T\,|\,A=a)P(B=b)\left[\sum_{e\in T,F} P(A=a\,|\,B=b,E=e)P(E=e)\right]$$

$$= \sum_{a\in T,F}\sum_{b\in T,F} P(J=T\,|\,A=a)P(B=b)\,\tau_1(A=a,B=b)$$

$$= \sum_{a\in T,F} P(J=T\,|\,A=a)\left[\sum_{e\in T,F} P(B=b)\,\tau_1(A=a,B=b)\right]$$

[redacted]

$$= \sum_{a\in T,F} P(J=T\,|\,A=a)\quad \tau_2(A=a)$$

√448

استنتاج در شبکه‌های بیزین

- **استنتاج دقیق:**
  - **حذف متغیر (VE)**
  - **تجزیه بازگشتی (RD)**
  - **استنتاج نمادین (SI)**
  - **الگوریتم انتشار باور (BP)**
  - **روش خوشه بندی (clustering) و درخت توأم (Jtree)**
  - **وارون‌سازی پیکانها (AR)**

- **استنتاج تقریبی:**
  - **روشهای مونت کارلو (Monte Carlo methods) (تصادفی) (فصل 11)**
    مبتنی بر نمونه برداری
  - **روشهای تغییری (Variational methods) (معین) (فصل 10)**

یادگیری شبکه‌های بیزین

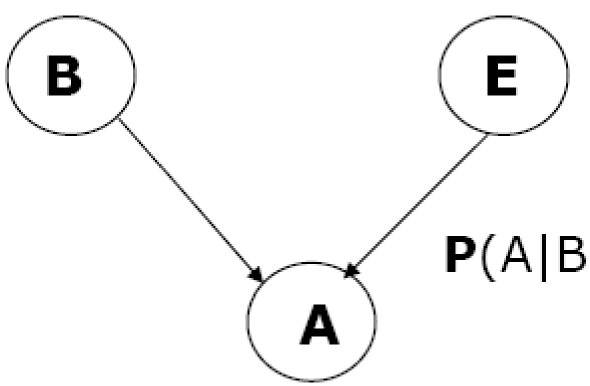
- **یادگیری:**
  - یادگیری پارامترهای احتمالات شرطی
  - یادگیری ساختار شبکه

- **متغیرها:**
  - قابل مشاهده (observable): مقدارشان در تمام نمونه‌های داده موجود است.
  - مخفی (hidden): مقدارشان در هیچ یک از نمونه‌های داده موجود نیست.
  - مقادیر مفقود (missing values): مقدارشان در برخی از نمونه‌های داده موجود و در برخی دیگر موجود نیست.

√448

تخمین پارامترهای شبکه بیزین

- یادگیری پارامترهای BBN با فرض قابل مشاهده بودن تمام متغیرها
  - ایده: تجزیه مسأله تخمین پارامترهای توزیع توأم کامل به یک مجموعه مسایل تخمین کوچکتر متناظر با احتمالات شرطی محلی
  - فرضی که ما را قادر به انجام این تجزیه میکند: پارامترهای توزیعهای شرطی از هم مستقل هستند.
  - پارامترهای هر توزیع شرطی (یک پارامتر برای هر انتساب مقدار به متغیرهای والد) میتوانند مستقلاً یادگرفته شوند.
  - مثال: **A** و **B** و **E** متغیرهای دودویی با مقادیر **True** و **False** :



$$\mathbf{P}(A|B,E)\begin{cases}\mathbf{P}(A|B=T,E=T)\\ \mathbf{P}(A|B=T,E=F)\\ \mathbf{P}(A|B=F,E=T)\\ \mathbf{P}(A|B=F,E=F)\end{cases}$$

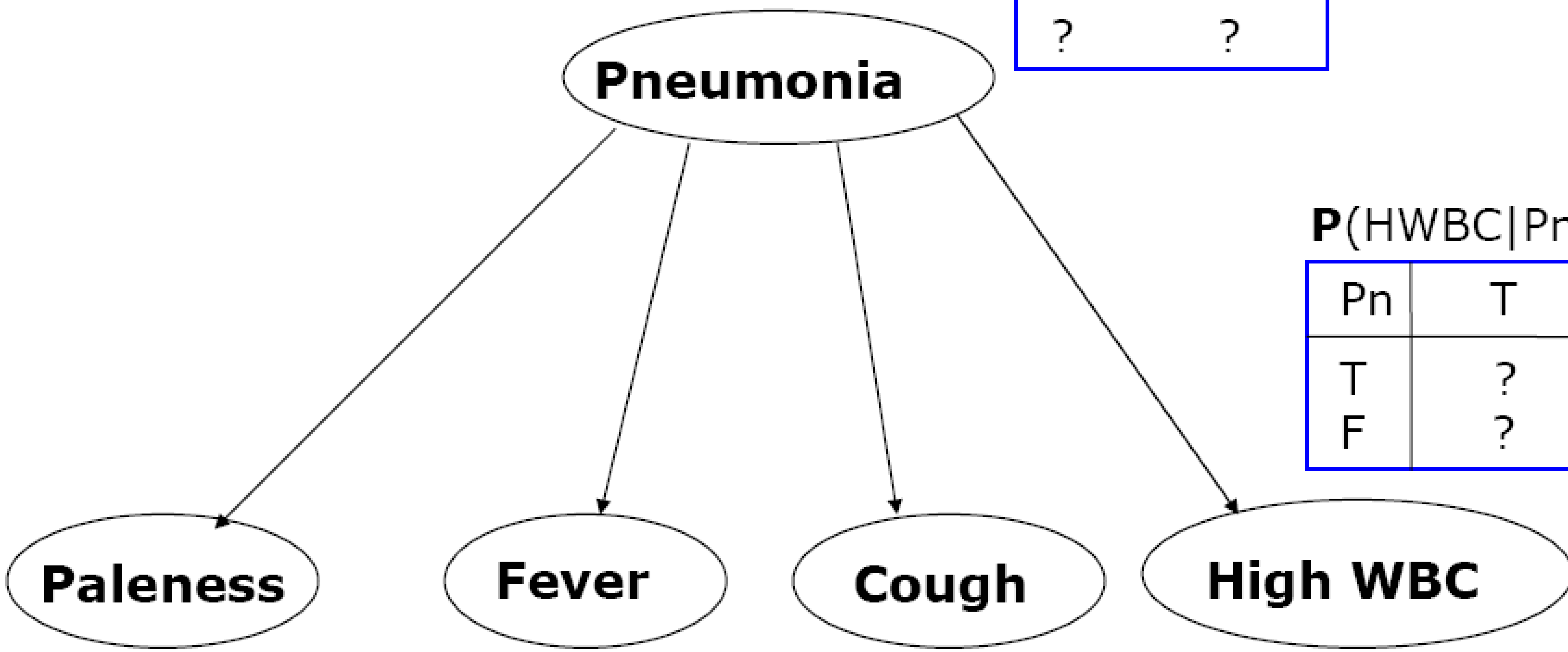
$\leftarrow$ چهار مسأله تخمین

تخمین پارامترهای شبکه بیزین

- مثال: بیماری ذات الریه

P(Pneumonia)

T	F
?	?



P(HWBC|Pneum)

Pn	T	F
T	?	?
F	?	?

P(Palen|Pneum)

?

P(Fever|Pneum)

?

P(Cough|Pneum)

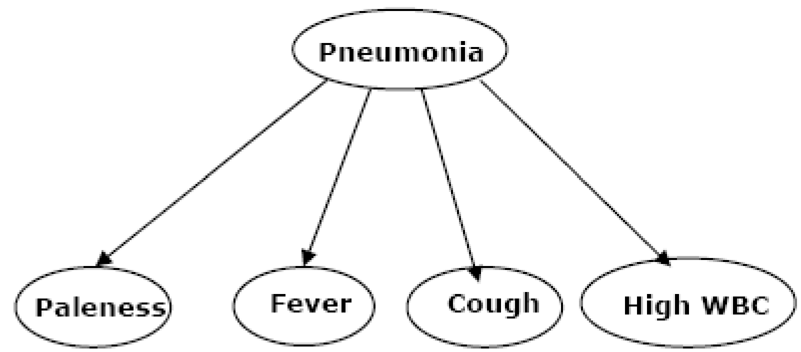
?

√480

تخمین پارامترهای شبکه بیزین

- **مثال:** بیماری ذات الریه
  - داده $\mathbf{D}$ (مربوط به بیماران مختلف):

Pal	Fev	Cou	HWB	Pneu
T	T	T	T	F
T	F	F	F	F
F	F	T	T	T
F	F	T	F	T
F	T	T	T	T
T	F	T	F	F
F	F	F	F	F
T	T	F	F	F
T	T	T	T	T
F	T	F	T	T
T	F	F	T	F
F	T	F	F	F



  - شبیه مسأله پرتاب چند سکه
  - یک زیر مسأله یادگیری:

    یادگیری دقیقاً یک توزیع شرطی، مثل:

$$\mathbf{P}(Fever \mid Pneumonia = T)$$

  - سئوال: چگونه از نمونه‌ها برای یادگیری یک توزیع شرطی استفاده کنیم؟

تخمین پارامترهای شبکه بیزین

- مثال: بیماری ذات الریه
  - سئوال: چگونه از نمونه‌ها برای یادگیری یک توزیع شرطی استفاده کنیم؟

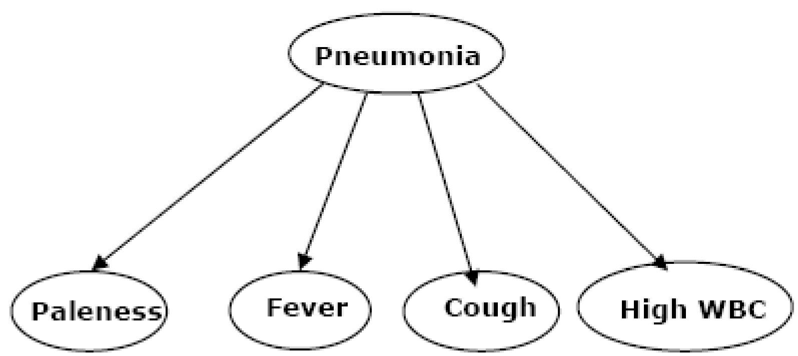
$$\mathbf{P}(Fever \mid Pneumonia = T)$$

  - پاسخ:

1 – انتخاب نمونه‌های با Pneumonia=T (و صرف نظر از بقیه)

2 – تمرکز بر (انتخاب) تنها مقادیر متغیر تصادفی تعریف کننده توزیع (Fever)

3 – یادگیری پارامترهای توزیع شرطی به همان صورت که پارامترهای توزیع را در یک مسأله پرتاب سکه یا تاس بدست می آوردیم.



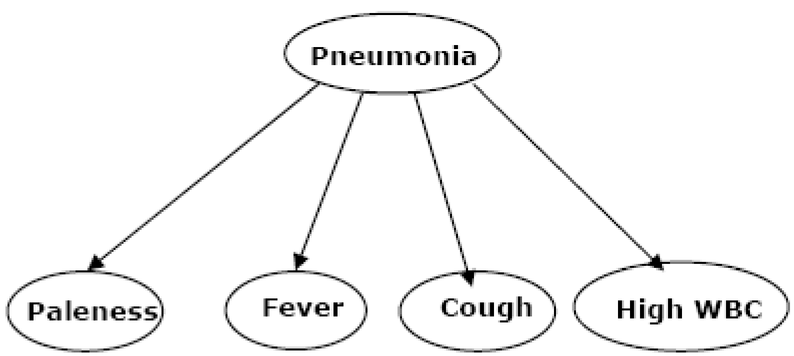
تخمین پارامترهای شبکه بیزین

- **مثال:** بیماری ذات الریه
  - یادگیری $\mathbf{P}(Fever \mid Pneumonia = T)$

Pal	Fev	Cou	HWB	Pneu
T	T	T	T	F
T	F	F	F	F
F	F	T	T	T
F	F	T	F	T
F	T	T	T	T
T	F	T	F	F
F	F	F	F	F
T	T	F	F	F
T	T	T	T	T
F	T	F	T	T
T	F	F	T	F
F	T	F	F	F

قدم 1 – انتخاب نمونه‌های با Pneumonia=T
و صرف نظر از بقیه

Pal	Fev	Cou	HWB	Pneu
F	F	T	T	T
F	F	T	F	T
F	T	T	T	T
T	T	T	T	T
F	T	F	T	T



تخمین پارامترهای شبکه بیزین

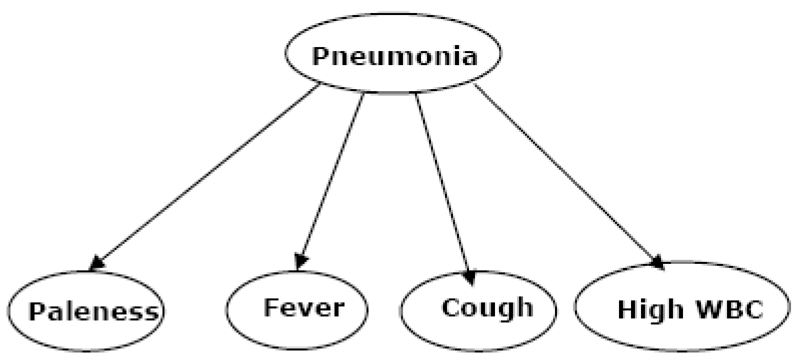
- مثال: بیماری ذات الریه
  - یادگیری $P(Fever \mid Pneumonia = T)$

قدم 2 – تمرکز بر (انتخاب) تنها مقادیر متغیر تصادفی تعریف کننده توزیع (Fever) و صرف نظر از بقیه

Pal	Fev	Cou	HWB	Pneu
F	F	T	T	T
F	F	T	F	T
F	T	T	T	T
T	T	T	T	T
F	T	F	T	T

$\longrightarrow$

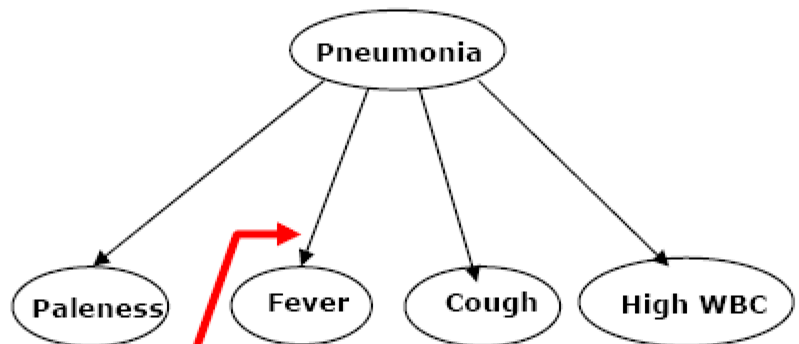
Fev
F
F
T
T
T



تخمین پارامترهای شبکه بیزین

- مثال: بیماری ذات الریه
  - یادگیری $\mathrm{P}(Fever \mid Pneumonia = T)$

قدم 3 – (روش اول) یادگیری تخمین ML



Fev
F
F
T
T
T

$$\mathrm{P}(Fever \mid Pneumonia = T)$$

T	F
0.6	0.4

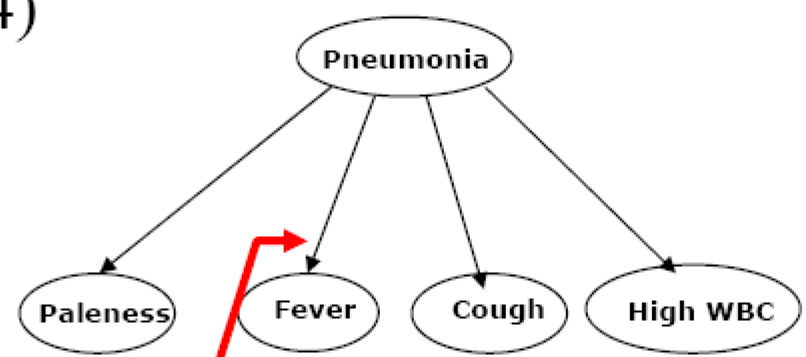
تخمین پارامترهای شبکه بیزین

- **مثال:** بیماری ذات الریه
  - یادگیری $\mathbf{P}(Fever \mid Pneumonia = T)$

قدم 3 – (روش دوم) یادگیری تخمین بیزین، با فرض توزیع پیشین زیر:

$$\theta_{Fever \mid Pneumonia = T} \sim Beta(3,4)$$

Fev
F
F
T
T
T



Posterior:

$$\theta_{Fever \mid Pneumonia = T} \sim Beta(6,6)$$

√448

مقادیر مفقود

- یک مجموعه متغیر تصادفی: $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$
- مجموعه داده: $D = \{D_1, D_2, .., D_N\}$

  - اما برخی از مقادیر برخی از نمونه های مفقود است:

$D_i = (x_1^i, x_3^i, \ldots x_n^i)$ Missing value x_2^i

$D_{i+1} = (x_3^{i+1}, \ldots x_n^{i+1})$ Missing values x_1^{i+1}, x_2^{i+1}

Etc.

- مثال: پرونده های پزشکی

- میخواهیم با وجود مقادیر مفقود چگالی توأم $P(\mathbf{X})$ را تخمین بزنیم

√4

تخمین چگالی

- هدف: تخمین یک مجموعه از پارامترها $\hat{\Theta}$
- معیارهای تخمین:
  - ML: $\max_{\Theta} \; p(D \mid \Theta, \xi)$
  - بیزین: $p(\Theta \mid D, \xi)$
- روشهای بهینه سازی ML:
  - صعود در امتداد گرادیان، مزدوج گرادیان، نیوتن-رفسون، ...
- مسأله:
  - عدم استفاده از ساختار شبکه باور نظیر وقتی مقادیر مشاهده نشده از متغیرها وجود دارد
- روش میانگین گیری-بیشینه سازی (EM)
  - یک روش بهینه سازی دیگر
  - مناسب برای مواقعی که مقادیر مفقود یا مخفی داریم
  - از ساختار شبکه باور استفاده میکند

√4

الگوریتم EM

- ایده : محاسبه تخمین پارامتر بصورت تکراری با انجام دو قدم زیر:

قدم میانگین گیری (expectation step): تکمیل تمام متغیرهای مخفی و مفقود با میانگین گیری برای مقدار جاری پارامترها Θ'

قدم بیشینه سازی (maximization step): محاسبه تخمین جدید پارامترها Θ برای داده تکمیل شده

توقف زمانی که بهبودی حاصل نمیشود

الگوریتم EM

- فرض کنید H مجموعه مقادیر مخفی یا مفقود باشد

$$P(H,D \mid \Theta,\xi) = P(H \mid D,\Theta,\xi)P(D \mid \Theta,\xi)$$

$$\log P(H,D \mid \Theta,\xi) = \log P(H \mid D,\Theta,\xi) + \log P(D \mid \Theta,\xi)$$

$$\log P(D \mid \Theta,\xi) = \log P(H,D \mid \Theta,\xi) - \log P(H \mid D,\Theta,\xi)$$

Log-likelihood of data

میانگین گیری از دو طرف روی H (با $P(H \mid D,\Theta',\xi)$ برای یک Θ')

$$E_{H|D,\Theta'} \log P(D \mid \Theta,\xi) = E_{H|D,\Theta'} \log P(H,D \mid \Theta,\xi) - E_{H|D,\Theta'} \log P(H \mid \Theta,\xi)$$

$$\underbrace{\log P(D \mid \Theta,\xi)}_{\text{Log-likelihood of data}} = Q(\Theta \mid \Theta') + H(\Theta \mid \Theta')$$

- قدم E: محاسبه Q
- قدم M: بیشینه سازی Q

√224

تحلیل مؤلفه اصلی

Principal Component Analysis (PCA)

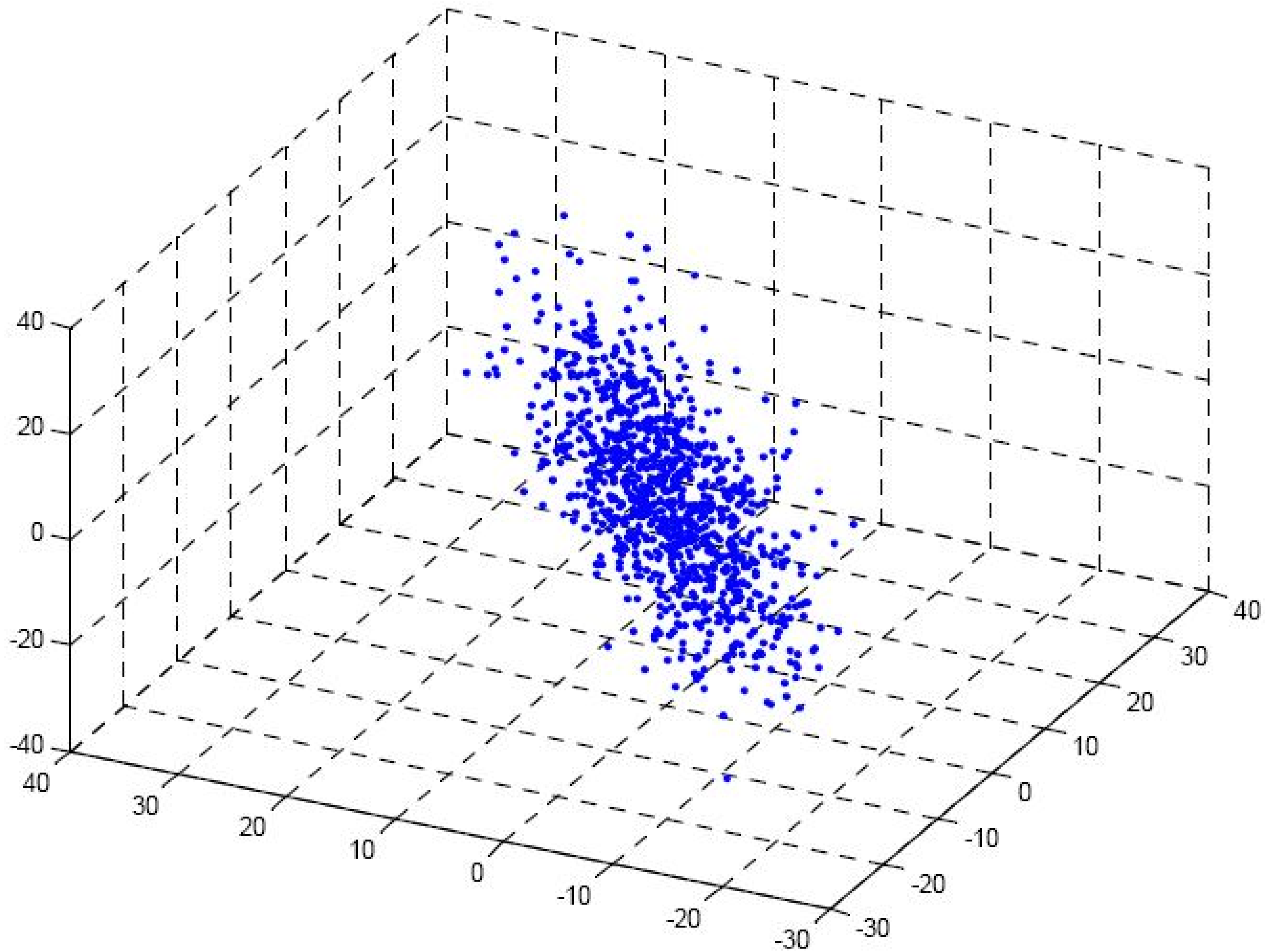
تحلیل مؤلفه اصلی (PCA)

- **هدف: جایگزین کردن یک ورودی دارای بعد بالا با یک مجموعه کوچک از ویژگیها (که از ترکیب ورودیها بدست می‌آید).**
  - **این روش جزو روشهای استخراج ویژگی است نه انتخاب ویژگی**

- **تحلیل مؤلفه اصلی (PCA):**
  - **یک تبدیل خطی از ورودی d بعدی x به بردار ویژگی M بعدی z که $M<d$، به نحوی که تحت این تبدیل واریانس حفظ شونده داده بیشینه باشد.**
  - **بطور معادل، این تبدیل عبارت از یک تصویر سازی یا افکنش (projection) خطی است که مجموع مربعات خطای بازسازی را کمینه میکند.**

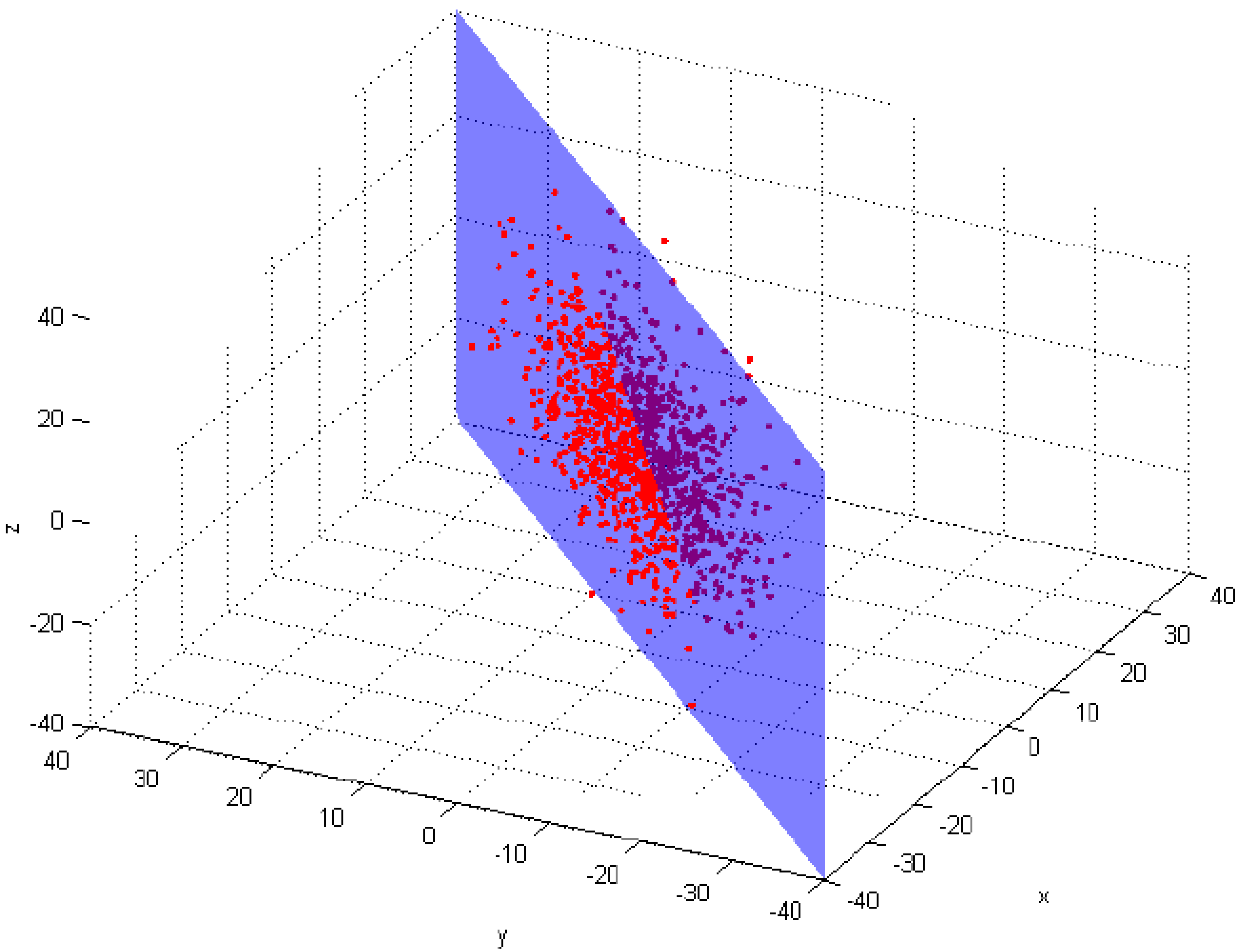
تحلیل مؤلفه اصلی (PCA)

- مجموعه داده



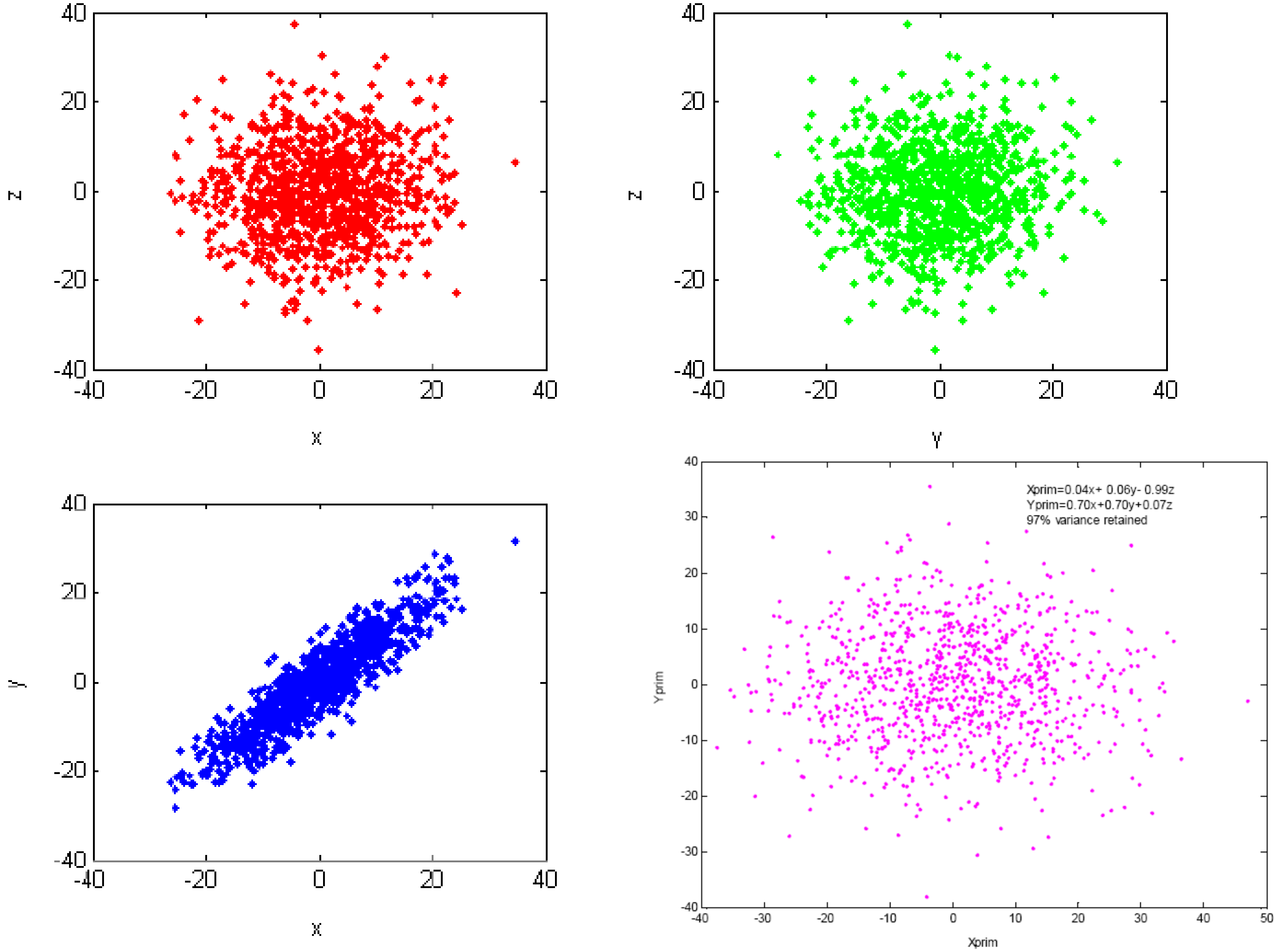
تحلیل مؤلفه اصلی (PCA)

- زیرفضای اصلی مجموعه داده



تحلیل مؤلفه اصلی (PCA)

- افکنش بر روی صفحات مختصات (قرمز، سبز، و آبی)
- افکنش بر روی زیر فضای اصلی (صورتی)



√484