دانشگاه علم و صنعت
دانشکده مهندسی کامپیوتر

کمی‌سازی عدم قطعیت

«هوش مصنوعی: رهیافتی نوین»، فصل ۱۳

مدرس: آرش عبدی هجراندوست

نیم‌سال دوم ۱۴۰۱-۱۴۰۲

عدم قطعیت

- در منطق گزاره ها، امور، قطعی فرض می‌شدند.
- گاهی به دلایل مختلف، قطعیت وجود ندارد.
- احتمالات
- چگونه می‌توان از گزاره‌های احتمالا/تاحدودی درست استفاده کرد؟

یک مثال

- دستگاه تشخیص سرطان:
  - احتمال داشتن سرطان در افراد ۰.۰۰۱ است.
  - اگر سرطان باشد، با احتمال ۰.۹۹ نتیجه تست مثبت است.
  - اگر سرطان نباشد، با احتمال ۰.۹۹۵ نتیجه تست منفی است.
  - فرض کنید نتیجه تست شما مثبت شده:
    - چقدر احتمال دارد سرطان داشته باشید؟
      - زیر ۰.۵۰
      - بین ۰.۵۰ تا ۰.۸۰
      - بین ۰.۸۰ تا ۰.۹۵
      - بالای ۰.۹۵

تئوری تصمیم

- یک عامل منطقی وقتی با پدیده‌های غیر قطعی مواجه است، چگونه باید تصمیم بگیرد؟
- چگونه باید بین پدیده‌ها ترجیح دهد؟
- طبق تئوری تصمیم، یک عامل منطقی است اگر و فقط اگر عملی را انتخاب کند که **بیشترین سودمندی مورد انتظار** را در پی دارد
  - سودمندی مورد انتظار
    - میانگین سودمندی برای خروجی های مختلف هر عمل
    - میانگین وزن دار بر اساس احتمال وقوع هر خروجی
- مثال دیوانه و عقرب

تعاریف اولیه احتمال

- مجموع احتمالات برای تمام خروجی های یک متغیر تصادفی ۱ است
- احتمال، عددی نامنفی است.
- احتمالاتی نظیر P(x) احتمال غیر شرطی یا احتمال اولیه یا prior نام دارند.
  - میزان باور به وقوع پدیده‌ای وقتی هیچ اطلاعات دیگری در دست نیست.
- وقتی اطلاعات اضافی وجود دارد، احتمال P(X | Y) را احتمال شرطی یا پسین یا posterior می‌نامیم.
- قانون شرطی:

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)}$$

❖ هر متغیر تصادفی یک دامنه (مقادیر ممکن) دارد.

❖ احتمال هر مقدار از دامنه:

$$
\begin{aligned}
P(Weather = sunny) &= 0.6 \\
P(Weather = rain) &= 0.1 \\
P(Weather = cloudy) &= 0.29 \\
P(Weather = snow) &= 0.01\ ,
\end{aligned}
$$

❖ و برای سادگی:

$$\mathbf{P}(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$$

P = توزیع احتمال برای متغیر تصادفی Weather

❖ **P**(X | Y) = توزیع احتمال شرطی (برای مقادیر ممکن X و Y

احتمال و متغیر پیوسته

- تابع توزیع احتمال برای متغیر پیوسته نمی‌تواند تمام مقادیر ممکن را برشمرد
  - بی نهایت مقدار وجود دارد
- در این حالت احتمال متغیر تصادفی X را با تابعی پارامتردار/پیوسته برحسب X نشان می‌دهیم.
  - مثلا: P(NoonTemp=x)=Uniform_[18C,26C](x)
  - دما دارای توزیع یکنواخت بین بازه ۱۸ تا ۲۶ درجه است.
- به تابع مذکور، تابع چگالی احتمال probability density function – pdf می‌گوییم.

تابع چگالی احتمال

- تابع چگالی احتمال معنای متفاوتی نسبت به تابع توزیع احتمال گسسته دارد.
  - احتمال آنکه دما در زمانی معین، دقیقا ۲۰ درجه باشد؟ صفر
    - حتی اگر باشد هم ابزار دقیقی که بسنجد این واقعیت را نداریم.
  - تمامی percept های تجربی ما از مفاهیم پیوسته دنیا، نادقیق است و ما هیچ چیزی از دنیا را دقیق نمی‌دانیم!
    - آیا چنین است؟
- معنای شهودی P(X=x) یا P(x) (تابع چگالی احتمال)
  - احتمال آنکه متغیر X در بازه‌ای کوچک در اطراف x بیفتد، تقسیم بر طول بازه مذکور

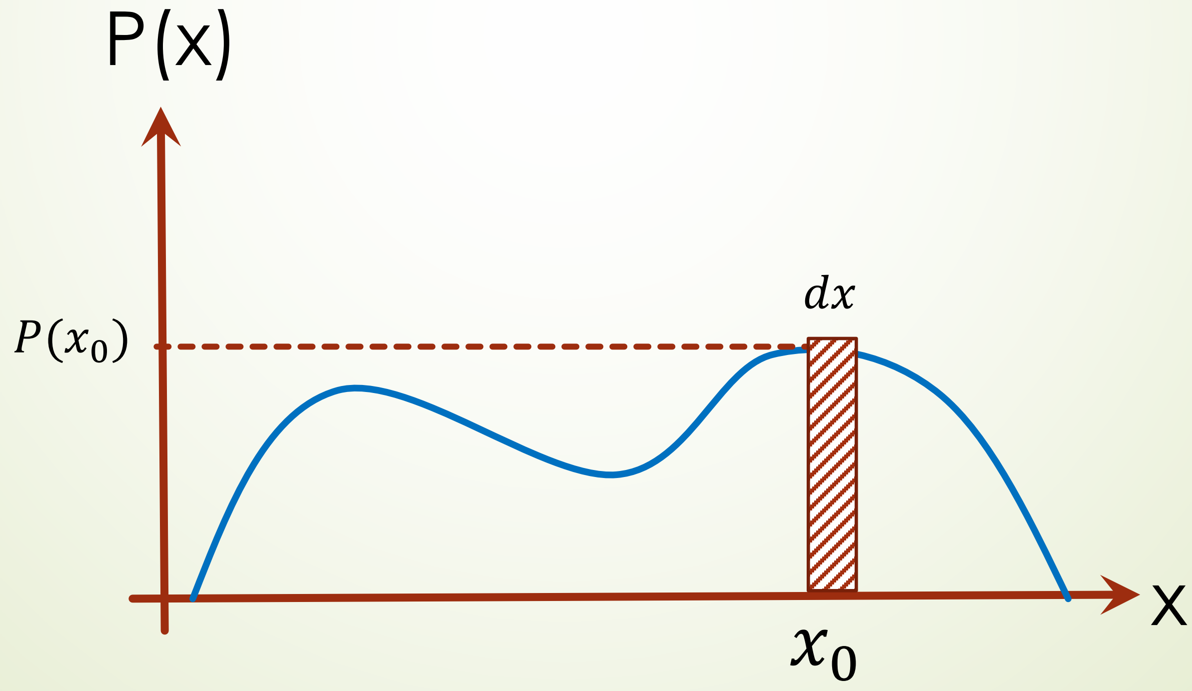
$$P(x) = \lim_{dx \to 0} P(x \leq X \leq x + dx)/dx$$

$$P(NoonTemp = x) = Uniform_{[18C,26C]}(x) = \begin{cases} \frac{1}{8C} \text{ if } 18C \leq x \leq 26C \\ 0 \text{ otherwise} \end{cases}$$

- احتمال آنکه دما ۲۰.۱۸ باشد = صفر
- P(NoonTemp=20.18)=1/8C

- $P(x) = \frac{P(x \le X \le x+dx)}{dx}$
- $P(x).dx = P(x \le X \le x+dx)$

- مساحت زیر نمودار = ۱



احتمال توام

- توزیع احتمال چند متغیر تصادفی
- $P(X,Y)$ = احتمال تمام ترکیب‌های ممکن برای مقادیر X و Y
- احتمال توام را می‌توان بدین شکل نوشت:
  - $P(X,Y)=P(X|Y).P(Y)$
- احتمال توام کامل: احتمال برای تمام متغیرهای موجود (بدون در نظر نگرفتن هیچ متغیری)

منشا احتمال

- مناقشه بر سر اینکه منشا احتمال چیست:
  - نگاه فراوانی محور:
    - احتمال از روی آزمایش و تجربه به دست می‌آید و بر اساس فراوانی مشاهدات
  - نگاه عینیت‌گرا:
    - احتمالات، مفاهیمی واقعی مربوط به دنیا هستند (جنبه‌هایی از دنیای واقعی هستند) که عبارت است از گرایش ذاتی اشیاء به داشتن یک رفتار(خروجی) خاص.
    - احتمال، صرفا توصیفی از میزان باور مشاهده کننده خارجی نسبت به اشیاء نیست.
  - نگاه ذهنیت‌گرا:
    - احتمال راهی برای بیان باور مشاهده کننده است نسبت به یک پدیده، بی آنکه معنا و مفهوم فیزیکی خارجی برای آن متصور باشد.

- بحث:
  - احتمال آمدن خط در پرتاب سکه ۰.۵ است. (یعنی چه؟)
  - احتمال بارندگی فردا ۰.۷ است؟
    - نگاه فراوانی محور: اگر ۱۰۰ بار فردا بشود در شرایطی که دقیقا امروز حاکم است، ۷۰ بارش باران می‌بارد؟
    - چنین آزمایشی امکان پذیر است؟ قبلا تجربه شده است؟
  - به احتمال ۰.۸۵ تیم ایران از گروه خود صعود نخواهد کرد!
    - قبلا شرایط مشابه تجربه شده؟ نگاه فراوانی محور؟ عینیت گرا؟ ذهنیت گرا؟
  - آیا در پرتاب سکه، عنصر تصادف وجود دارد؟ (متغیر تصادفی؟)

استنتاج با توزیع توام کامل

- استنتاج احتمالاتی: یافتن احتمال پسین برای گزاره‌های پرسش شده با پدیده‌های مشاهده شده
- با توزیع توام کامل به عنوان پایگاه دانش می‌توان پاسخ هر پرسشی را داد.
- فرض کنیم ۳ متغیر بولین داریم: دندان درد، کرم‌خوردگی دندان، گیرکردن میله دندانپزشک در دندان
- توزیع توام کامل:

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(cavity \lor toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

❖ احتمال حاشیه‌ای- Marginal Probability

$$P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

❖ حاشیه‌سازی(!) - Marginalization

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z})$$

❖ شرطی سازی:

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z})$$

مثال استنتاج شرطی

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(cavity \mid toothache) = \frac{P(cavity \land toothache)}{P(toothache)} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \land toothache)}{P(toothache)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

- مخرج کسر نقش نرمال‌سازی را بازی می‌کند (جمع دو احتمال فوق را یک می‌کند)

$$\begin{aligned}\mathbf{P}(Cavity \mid toothache) &= \alpha\, \mathbf{P}(Cavity, toothache)\\ &= \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]\\ &= \alpha\, [\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle] = \alpha\, \langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle\ .\end{aligned}$$

- Cavity = متغیر بولین (هر دو مقدار ممکن)
- cavity = مقدار true برای متغیر Cavity (کرم خوردگی داشتن دندان)

استقلال

- اگر در مثال دندانپزشک، متغیر چهارمی به نام «وضعیت هوا» هم اضافه شود...

$$P(toothache, catch, cavity, cloudy)$$
$$= P(cloudy \mid toothache, catch, cavity) P(toothache, catch, cavity)$$

$$P(cloudy \mid toothache, catch, cavity) = P(cloudy)$$

$$P(toothache, catch, cavity, cloudy) = P(cloudy) P(toothache, catch, cavity)$$

$$\mathbf{P}(Toothache, Catch, Cavity, Weather) = \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

- و در حالت کلی برای گزاره های a, b یا متغیرهای **مستقل** تصادفی X,Y داریم:

$$P(a \mid b) = P(a) \quad \text{or} \quad P(b \mid a) = P(b) \quad \text{or} \quad P(a \wedge b) = P(a)P(b)$$

$$\mathbf{P}(X \mid Y) = \mathbf{P}(X) \quad \text{or} \quad \mathbf{P}(Y \mid X) = \mathbf{P}(Y) \quad \text{or} \quad \mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$$

قاعده بیز (Bayes)

$$P(a \wedge b) = P(a | b) P(b)$$

$$P(a \wedge b) = P(b | a) P(a)$$

- قاعده بیز:

$$P(b | a) = \frac{P(a | b) P(b)}{P(a)}$$

- و در حالت کلی:

$$\mathbf{P}(Y | X) = \frac{\mathbf{P}(X | Y) \mathbf{P}(Y)}{\mathbf{P}(X)}$$

- اگر یک مشاهده زمینه‌ای e داشته باشیم:

$$\mathbf{P}(Y | X, \mathbf{e}) = \frac{\mathbf{P}(X | Y, \mathbf{e}) \mathbf{P}(Y | \mathbf{e})}{\mathbf{P}(X | \mathbf{e})}$$

کاربرد قاعده بیز

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)}$$

- در نگاه اول، قاعده بیز، یک عبارت را بر حسب سه عبارت دیگر که پیچیدگی یکی از آنها در حد پیچیدگی عبارت اول است بیان می‌کند.
  - چرا باید کاربردی باشد؟
- اگر از این زاویه به مساله نگاه کنیم، قدری متفاوت است:

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

- $P(effect|cause)$: زاویه دید عِلّی (Causal)
- $P(cause|effect)$: زاویه دید تشخیصی (Diagnostic)

- اطلاعات در سمت راست معادله فوق، ارزان‌تر از اطلاع سمت چپ به دست می‌آیند!

کاربرد قاعده بیز

- پزشک می‌داند بیماری مننژیت (m) در ۷۰٪ موارد باعث سفتی گردن (s) می‌شود.
- پزشک می‌داند احتمال داشتن مننژیت در مردم ۱/۵۰۰۰۰ است.
- (پزشک می‌داند ۱٪ مردم دچار سفتی گردن هستند.)
- مراجعه کننده با عارضه سفتی گردن مراجعه کرده است:

$$
\begin{aligned}
P(s \mid m) &= 0.7 \\
P(m) &= 1/50000 \\
P(s) &= 0.01 \\
P(m \mid s) &= \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.7 \times 1/50000}{0.01} = 0.0014
\end{aligned}
$$

❖ اگر نرمال‌سازی که قبلا اشاره شد را در اینجا به کار بگیریم:

$$\mathbf{P}(M \mid s) = \alpha \langle P(s \mid m)P(m), P(s \mid \neg m)P(\neg m) \rangle$$

❖ در این صورت به جای دانستن $P(s)$ باید $P(s|\neg m)$ را بدانیم.

❖ جمع $P(s|m)P(m) + P(s|\neg m)P(\neg m)$ که برابر با معکوس α است، $P(s)$ را تشکیل می‌دهد.

❖ در عمل گاهی اولی ارزان تر/در دسترس است و گاهی دومی.

❖ شکل کلی قاعده بیز با نرمال‌سازی:

$$\mathbf{P}(Y \mid X) = \alpha \mathbf{P}(X \mid Y)\mathbf{P}(Y)$$

چرا قاعده بیز کاربردی است؟

- دلیل اول: ارزان‌تر بودن و دسترسی بیشتر به اطلاعات عِلّی نسبت به اطلاعات تشخیصی
- دلیل دوم: اطلاعات تشخیصی حتی در صورت موجود بودن، شکننده هستند.
  - اگر در اثر یک اپیدمی، بیماری مننژیت گسترش یابد، احتمال پیشین $P(m)$ افزایش می‌یابد.
  - در این صورت پزشکی که $P(m|s)$ را بر اساس اطلاعات آماری گذشته خود (یا کتاب خود) می‌دانسته، چگونه باید آن را به‌روز کند؟
    - در انتظار نسخه بعدی کتاب؟ یا جمع آوری دوباره اطلاعات از افراد دارای سفتی گردن؟
  - اما پزشکی که $P(m|s)$ را با قاعده بیز محاسبه کرده، می‌داند که این احتمال نسبت مستقیمی با $P(m)$ دارد.

استفاده از Bayes با چند مشاهده

- اطلاعات در مسیر علِّی معمولا موجود است.
- در مسیر تشخیصی، اگر چند مشاهده (evidence) وجود داشته باشد:
  - می‌توان با داشتن جدول توزیع توام احتمالات پاسخ را پیدا کرد:

$$\mathbf{P}(Cavity \mid toothache \wedge catch) = \alpha \langle 0.108, 0.016 \rangle \approx \langle 0.871, 0.129 \rangle$$

  - اما داشتن جدولی بزرگ برای متغیرهای زیاد غیر ممکن/دشوار است.
  - می‌توان از قاعده بیز استفاده کرد:

$$\begin{aligned}&\mathbf{P}(Cavity \mid toothache \wedge catch)\\ &\quad = \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)\end{aligned}$$

  - اما لازم است احتمال شرطی ترکیب عطفی چند متغیر را بدانیم و با تعداد متغیر بالا فرض دانستن چنین اطلاعاتی به سختی داشتن جدول توزیع توام است.
  - می‌توان از برخی از متغیرهای کم اهمیت‌تر صرف نظر کرد تا جوابی **تقریبی** ولی کم هزینه‌تر یافت!
    - ایموجی مربوطه!

استفاده از Bayes با چند مشاهده

- آیا از مفهوم استقلال می‌توان استفاده کرد؟

$$\alpha\,\mathbf{P}(toothache \wedge catch \mid Cavity)\,\mathbf{P}(Cavity)$$

  - آیا toothache و catch مستقلند؟
    - اگر باشند، احتمال شرطی فوق به دو بخش ارزان خواهد شکست.
    - اما مستقل نیستند!
    - اگر میله دندانپزشک گیر کند، احتمال اینکه فرد دندان درد هم داشته باشد بیشتر می‌شود و برعکس
      - در جامعه آماری دندان دردَنده ها(!)، فراوانی گیرَنده(!)های میله پزشکی بیشتر است نسبت به جامعه آماری نرمال
- toothache و catch مستقلند، اگر وضعیت Cavity را بدانیم!
  - هر دو متغیر **مستقیما** از یک ریشه مشترک تاثیر می‌گیرند، بی آنکه تاثیر **مستقیمی** روی هم بگذارند.
    - دندان درد مستقیما از وضعیت عصب‌های دندان اثر می‌گیرد و گیرکردن میله تابع مهارت دندانپزشک است.

$$\mathbf{P}(toothache \wedge catch \mid Cavity) = \mathbf{P}(toothache \mid Cavity)\mathbf{P}(catch \mid Cavity)$$

- این دو متغیر نسبت به هم دارای **استقلال شرطی** هستند.

استقلال شرطی

- دو متغیر X و Y دارای استقلال شرطی هستند:

$$\mathbf{P}(X, Y \mid Z) = \mathbf{P}(X \mid Z)\mathbf{P}(Y \mid Z)$$

$$\mathbf{P}(X \mid Y, Z) = \mathbf{P}(X \mid Z) \quad \text{and} \quad \mathbf{P}(Y \mid X, Z) = \mathbf{P}(Y \mid Z)$$

- بیان آماری استقلال شرطی:
  - در یک جامعه آماری ۱۰۰ نفره، همه افراد دارای کرم خوردگی دندان (به یک اندازه ثابت) هستند
  - ۳۰٪ دندان درد دارند
  - ۴۰٪ catch=true هستند
  - هم بستگی بین این ۳۰ نفر و آن ۴۰ نفر؟
    - اگر آن ۴۰ نفر را جدا کنیم، آیا می‌توان گفت در بین آنها شیوع دندان درد بیشتر یا کمتر از ۳۰٪ است؟
    - خیر ← استقلال شرطی

استفاده از استقلال شرطی و قاعده بیز

- دندانپزشک لازم دارد بداند احتمال کرم خوردگی (و نیاز به ترمیم) با چند مشاهده مختلف چقدر است؟
  - ۵ متغیر، مثلا درد+رادیولوژی+مسواک میزنه یا نه+رژیم غذایی بد داره یا نه +... ۳۲ حالت مختلف
- جدول کامل احتمالات توام را نمی‌توان داشت
- با قاعده بیز، احتمال کرم خوردگی از بقیه احتمالات جدا می‌شود (مقدارش را داریم)
- احتمال توام بقیه متغیرها به دلیل استقلال شرطی شان، حاصل ضرب احتمال تک تک متغیرها به شرط ریشه است
- احتمال تک تک متغیر ها به شرط ریشه را به راحتی می‌توان در گوگل هم پیدا کرد!
  - این جزء شناسنامه یک بیماری است که چند درصد باعث بروز فلان علامت میشود و چند درصد بهمان علامت.

میزان کاهش هزینه

- با ۳ متغیر تصادفی دودوئی (شامل ریشه)، در جدول توزیع احتمال، هشت عدد (احتمال) وجود دارد که جمعشان ۱ است
  - بنابراین، عملا هفت عدد مستقل وجود دارد.
- با استفاده از استقلال شرطی و قاعده بیز:
  - یک عدد مستقل برای ریشه (کرم خوردگی) داریم. (جمع احتمال کرم خوردگی و کرم ناخوردگی یک است ← یک عدد مستقل)
  - یک عدد مستقل برای دندان درد با فرض وجود کرم خوردگی داریم (با این فرض، جمع احتمال داشتن و نداشتن دندان درد، یک است)
  - یک عدد مستقل برای دندان درد با فرض عدم وجود کرم خوردگی داریم (با این فرض، جمع احتمال داشتن و نداشتن دندان درد، یک است)
  - دو عدد مستقل دیگر به همین ترتیب برای catch مورد نیاز است.
- بنابراین اگر از استقلال شرطی استفاده نشود، $2^n - 1$ عدد احتمال نیاز است.
  - رشد **<u>نمایی</u>** نیازمندی با افزایش تعداد متغیر
- اگر از استقلال شرطی و قاعده بیز استفاده شود $2n - 1$ عدد نیاز است
  - رشد **<u>خطی</u>** نیازمندی با افزایش تعداد متغیر
- اگر متغیرها k مقدار مختلف داشته باشند (دودوئی نباشند) چطور؟

دسته‌بندی مبتنی بر بیز

- مدل بیز ساده انگارانه (Naive Bayes):

$$P(Cause, Effect_1, \dots Effect_n) = P(Cause)\prod_i P(Effect_i|Cause)$$

- چرا ساده انگارانه:
  - ممکن است effect ها واقعا دارای استقلال شرطی نباشند.
- به مدل فوق دسته‌بندی کننده بیز نیز گفته می‌شود.
  - دسته بندی؟
    - تعدادی effect دیده‌ایم. دسته خروجی (cause) را میخواهیم تعیین کنیم.

$$P(Cause|Effect_1, \dots Effect_n) = \alpha P(Cause, Effect_1, \dots Effect_n) = \alpha P(Cause)\prod_i P(Effect_i|Cause)$$

    - ضریب α اهمیتی ندارد، زیرا احتمال مقادیر مختلف برای Cause با یکدیگر مقایسه می‌شوند و مقدار با بیشترین احتمال، خروجی دسته‌بند خواهد بود.
    - $P(Cause)$ به صورت آماری قابل تحصیل است.
    - $P(Effect_i|Cause)$ نیز هم!
- این مدل با وجود ساده لوحی، در عمل کارایی خیلی خوبی دارد.
- یک مزیت مهم: کنار آمدن با Missed Value ها
  - کافی است برای $m < n$ (یعنی برای value های موجود) مدل بالا را محاسبه کنیم: $P(Cause|Effect_1, \dots Effect_m)$
  - وقتی برخی ویژگی‌ها از یک نمونه موجود نیست، بسیاری از دسته‌بندها دچار مشکل می‌شوند
    - مثلا درخت تصمیم؟