مروری بر استنتاج احتمالاتی

احتمال و آمار پایه و بنیان هوش مصنوعی است. در این ارائه موارد زیر را مرور و جمع‌بندی می‌کنیم:

- متغیرهای تصادفی
- تابع احتمال توأم
- توابع احتمال حاشیه‌ای
- احتمال شرطی
- استنتاج احتمالاتی
- قانون بیز
- استدلال قهقرایی (استدلال معکوس)

متغیرهای تصادفی

در احتمال با آزمایش تصادفی سر و کار داریم. اما در کاربرد تنها به جنبه‌ای خاص از نتایج آزمایش‌ها توجه داریم. اما متغیرهای تصادفی چگونه کار را راحت می‌کنند؟

به مثالی که در ادامه می‌آید توجه کنید.

مثال از متغیرهای تصادفی

یک دیتاست را در نظر بگیرید که شامل ستون‌های زیر است:

- فشار خون: فشار خون دیاستولیک برحسب میلی‌جیوه
- BMI: شاخص توده بدنی
- سن
- تعداد دفعات حاملگی
- خروجی: 1 اگر شخص دیابت دارد و 0 در غیر این صورت

دیتاست فوق برگرفته از
https://www.kaggle.com/datasets/mathchi/diabetes-data-set

مربوط به خانم‌های با سن حداقل ۲۱ است.

مثال از متغیرهای تصادفی

یک دیتاست را در نظر بگیرید که شامل ستون‌های زیر است:

- فشار خون: فشار خون دیاستولیک برحسب میلی‌جیوه -> X1
- BMI: شاخص توده بدنی -> X2
- سن -> X3
- تعداد دفعات حاملگی -> X4
- خروجی: 1 اگر شخص دیابت دارد و 0 در غیر این صورت -> Y

دیتاست فوق برگرفته از
https://www.kaggle.com/datasets/mathchi/diabetes-data-set

مربوط به خانم‌های با سن حداقل ۲۱ است.

نکته

در تعریف رسمی، متغیرهای تصادفی توابعی هستند که اعضای فضای نمونه‌ی آزمایش تصادفی را به یک مجموعه‌ی عددی نگاشت می‌کنند.

اما در اسلایدها دیدیم که گاهی اعضای فضای نمونه به مجموعه‌ای از کلمات نگاشت می‌شوند!

مثلا متغیر تصادفی‌ای را در نظر بگیرید که روی دما تعریف می‌شود. خروجی این متغیر تصادفی کلمه‌ی ((سرد)) یا کلمه‌ی ((گرم)) است.

اما می‌توان سرد بودن را با 1 و گرم بودن را با 0 نشان داد!

یا برعکس، گرم بودن را با 1 و سرد بودن را با 0 نشان داد!

احتمالا در کارگاه عملی درس در بخش مرتبط با یادگیری ماشین با چنین موقعیت‌هایی برخورد کرده‌اید.

تابع احتمال توأم (JOINT PROBABILITY)

تابع جرم احتمال توأم دو متغیر تصادفی **گسسته** X و Y عبارت است از

$$f(x,y) = P_{XY} = P(X = x \ \text{and} \ Y = y) = P(X = x, Y = y)$$

مثال از تابع احتمال توأم

Y

X	1	2	3
1	0.32	0.03	0.01
2	0.06	0.24	0.02
3	0.02	0.03	0.27

Joint probability of when
$X = 3$ and $Y = 2$

دو نمایش یکسان از یک جدول

T	W	P
HOT	SUN	0.4
HOT	RAIN	0.1
COLD	SUN	0.2
COLD	RAIN	0.3

W/T	COLD	HOT
SUN	0.2	0.4
RAIN	0.3	0.1

توابع احتمال حاشیه‌ای (MARGINAL)

		Y			
		1	2	3	
X	1	0.32	0.03	0.01	0.36
	2	0.06	0.24	0.02	0.32
	3	0.02	0.03	0.27	0.32
		0.40	0.30	0.30	1

Marginals for X

$$g(x)=\sum_{y} f(x,y)$$

Marginals for Y

$$h(y)=\sum_{x} f(x,y)$$

$$\sum_{X}\sum_{Y} f(x,y)=1$$

نوشتن این مقادیر در حاشیه‌ی این جدول‌ها منشا اصطلاح تابع احتمال حاشیه‌ای (در برخی متون: کناری) است.

توابع احتمال حاشیه‌ای باعث تقلیل تعداد متغیرهای تصادفی می‌شوند. مثلا از دو متغیر تصادفی به یک متغیر تصادفی می‌رسیم.

به جداول به دید توزیع‌های احتمالاتی نگاه کنید که مقادیر داخل آنها احتمال‌ها هستند

این جدول توزیع احتمال توأم متغیر تصادفی T و W است.

T	W	P
HOT	SUN	0.4
HOT	RAIN	0.1
COLD	SUN	0.2
COLD	RAIN	0.3

نکته‌ی مهم: جمع مقادیر داخل جداول باید برابر 1 شود.

احتمال شرطی

می‌خواهیم احتمال آفتابی بودن هوا را محاسبه کنیم. **اما می‌دانیم هوای بیرون سرد است!**

T	W	P
HOT	SUN	0.4
HOT	RAIN	0.1
COLD	SUN	0.2
COLD	RAIN	0.3

$$P(W=SUN|T=COLD) = \frac{P(W=SUN,T=COLD)}{P(T=COLD)} = \frac{0.2}{0.5} = 0.4$$

P(T=COLD) همان احتمال حاشیه‌ای است که قبلتر بررسی شد.

توزیع شرطی

اگر بدانیم هوای بیرون سرد است، دیگر به گرم بودن آن کاری نداریم و جدول ما کوچکتر می‌شود.

T	W	P
HOT	SUN	0.4
HOT	RAIN	0.1
COLD	SUN	0.2
COLD	RAIN	0.3

T	W	P
COLD	SUN	0.2
COLD	RAIN	0.3

جمع مقادیر داخل جدول ۱ نیستند!

توزیع شرطی

اگر بدانیم هوای بیرون سرد است، دیگر به گرم بودن آن کاری نداریم و جدول ما کوچکتر می‌شود.

T	W	P
HOT	SUN	0.4
HOT	RAIN	0.1
COLD	SUN	0.2
COLD	RAIN	0.3

T	W	P
COLD	SUN	0.2/0.5=0.4
COLD	RAIN	0.3/0.5=0.6

نرمال‌سازی می‌کنیم. توجه داریم که

$P(T = COLD) = 0.5 = 0.2 + 0.3$

استنتاج احتمالاتی

در استنتاج احتمالاتی سه نوع متغیر تصادفی داریم:

- **شواهد (EVIDENCE):** متغیر(های) تصادفی‌ای که درمورد آن‌ها اطلاعات داریم و مقادیرشان را می‌دانیم. این‌ها همان متغیرهایی هستند که در احتمال شرطی از آن‌ها استفاده می‌کردیم تا فضای احتمالاتی را محدودتر و جدول را کوچکتر کنیم.
- **کوئری (QUERY):** متغیری تصادفی که می‌خواهیم اطلاعاتی درمورد آن به دست آوریم. مثلا چقدر احتمال دارد هوا بارانی باشد؟
- **متغیرهای نهان (HIDDEN):** متغیرهایی که در توزیع احتمالاتی در دست ما تعریف شده‌اند، اما نه جزو شواهد ما هستند و نه کوئری ما است. یعنی نه اطلاعاتی از مقادیر آن‌ها داریم، و نه می‌خواهیم که اطلاعاتی درموردشان کسب کنیم. این متغیرها را با رابطه‌ی توابع احتمال حاشیه‌ای کنار می‌گذاریم.

درک قانون بیز مهم است!

LIKELIHOOD
The probability of "B" being True, given "A" is True

PRIOR
The probability "A" being True. This is the knowledge.

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

POSTERIOR
The probability of "A" being True, given "B" is True

MARGINALIZATION
The probability "B" being True.

استدلال قهقرایی / استدلال معکوس

قضیه‌ی بیز شیوه‌ای برای استدلال معکوس، و به بیانی دیگر، استدلال از معلول (effect) به علت (cause) است. توجه دارید که استدلال از علت به معلول غالبا سرراست و روشن است. آنچه اهمیت دارد، استدلال از معلول به علت (استدلال قهقرایی/استدلال معکوس) است. به عنوان مثال مصرف قند فراوان می‌تواند باعث دیابت شود، اما اگر شخصی دیابت بگیرد آیا علت مصرف قند فراوان بوده است؟

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

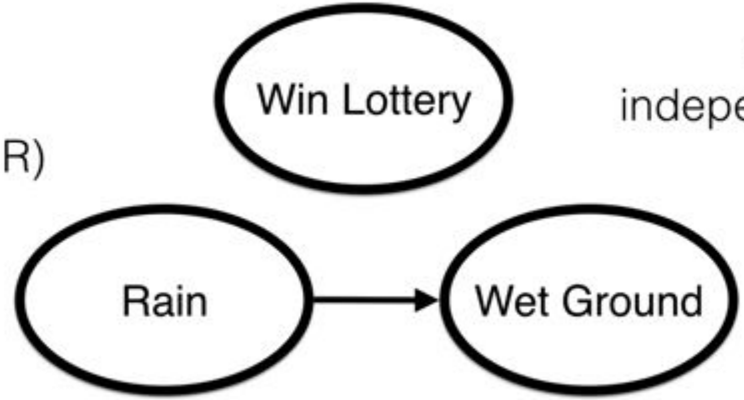
مدل‌های گرافیکی احتمالاتی

- روشی برای نشان دادن توزیع احتمالاتی کل متغیرهای تصادفی که متناظر با رئوس گراف هستند فراهم می‌کنند.
- روشی برای استنتاج فراهم می‌کنند.
- وابستگی و استقلال بین متغیرها را می‌توان از روی این مدل‌ها درک کرد.

Bayesian Networks

$P(L, R, W)$

$= P(L)\, P(R)\, P(W \mid R)$



conditional independence structure

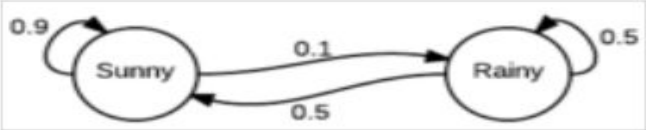
مدل مارکوف

در هر زمان وضعیت بعدی فقط به وضعیت کنونی وابسته است. با گذشته کاری نداریم.

Markov Model

The weather on day 0 (today) is known to be sunny. This is represented by an initial state vector in which the "sunny" entry is 100%, and the "rainy" entry is 0%:

$$\mathbf{x}^{(0)} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$



The weather on day 1 (tomorrow) can be predicted by multiplying the state vector from day 0 by the transition matrix:

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} P = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \end{bmatrix}$$

Thus, there is a 90% chance that day 1 will also be sunny. The weather on day 2 (the day after tomorrow) can be predicted in the same way, from the state vector we computed for day 1:

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} P = \mathbf{x}^{(0)} P^2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}^2 = \begin{bmatrix} 0.86 & 0.14 \end{bmatrix}$$

or

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} P = \begin{bmatrix} 0.9 & 0.1 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.86 & 0.14 \end{bmatrix}$$

Steady state of the weather $\mathbf{q} = \lim_{n \to \infty} \mathbf{x}^{(n)}$