Session 13

Fundamentals of Probability

Applied Mathematics for AI | Hobot Academy | Zahra Amini

Hobat Aacademy

Telegram & Instagram & YouTube: @hobotacademy & LinkedIn: @zahraamini-ai

www.hobotacademy.com

استقلال آماری

استقلال آماری چیست؟

در آمار و یادگیری ماشین، استقلال آماری (Statistical Independence) به حالتی گفته می‌شود که دو یا چند متغیر تصادفی هیچ ارتباطی با یکدیگر نداشته باشند. یعنی دانستن مقدار یکی، هیچ اطلاعاتی درباره مقدار دیگری ارائه ندهد.

به طور رسمی، دو متغیر تصادفی X و Y مستقل هستند اگر و فقط اگر توزیع توام آن‌ها برابر با حاصل ضرب توزیع‌های حاشیه‌ای باشد:

$$\mathrm{P(X,Y) = P(X)\ P(Y)}$$

برای متغیرهای پیوسته، این رابطه به صورت چگالی احتمال بیان می‌شود:

$$f_{\mathrm{X,Y}}(x,y) = f_{\mathrm{X}}(x) f_{\mathrm{Y}}(y)$$

اگر این رابطه برقرار نباشد، گفته می‌شود که متغیرها وابسته هستند.

توزیع احتمال مشترک P(X,Y): این مقدار نشان می‌دهد که دو متغیر تصادفی X و Y به‌طور همزمان چه احتمالی دارند.

مثال: اگر X تعداد روهای شیر در دو پرتاب سکه باشد و Y تعداد پشت‌ها، آنگاه P(X=۱,Y=۱) احتمال رخ دادن دقیقاً یک شیر و یک خط است.

توزیع احتمال حاشیه‌ای P(X) یا P(Y): این مقدار احتمال وقوع یک متغیر را بدون توجه به متغیر دیگر نشان می‌دهد.

مثال: P(X=۱) یعنی احتمال داشتن دقیقاً یک شیر، صرف‌نظر از تعداد خط‌ها.

اگر دو متغیر X و Y مستقل باشند:
- ✓ دانستن مقدار یکی از آن‌ها هیچ اطلاعاتی درباره دیگری ارائه نمی‌دهد.
- ✓ احتمالات آن‌ها تحت تأثیر هم قرار نمی‌گیرند.
- ✓ رابطه‌ی آن‌ها کاملاً تصادفی است.

اما اگر وابسته باشند:
- ✓ احتمال وقوع یکی با توجه به دیگری تغییر می‌کند.
- ✓ بین آن‌ها رابطه‌ی آماری مشخصی وجود دارد.

ی مستقل)، فرض کنید دو تاس شش‌وجهی منصفانه را پرتاب می‌کنیم. متغیرهای تصادفی زیر را در نظر بگیرید:

X عدد روی تاس اول

Y عدد روی تاس دوم

س دیگر وابسته نیست، می‌توان احتمال‌ها را به‌صورت جداگانه محاسبه کرد:

$$P(X = 3, Y = 5) = P(X = 3) \times P(Y = 5)$$

چون هر وجه از تاس احتمال ۱/۶ دارد:

$$P(X = 3, Y = 5) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

مثال ۲: نمره امتحان و میزان مطالعه (متغیرهای وابسته)، فرض کنید: X تعداد ساعت‌های مطالعه یک دانش‌آموز

Y نمره امتحان او

بدیهی است که هرچه ساعات مطالعه بیشتر باشد، نمره امتحان نیز معمولاً افزایش می‌یابد. بنابراین، X و Y وابسته هستند. اگر آن‌ها مستقل بودند، انتظار می‌رفت که:

$$P(X = 5, Y = 90) = P(X = 5)P(Y = 90)$$

اما در واقعیت، دانش‌آموزانی که بیشتر مطالعه می‌کنند معمولاً نمرات بالاتری دارند، بنابراین این معادله برقرار نیست و متغیرها وابسته‌اند.

دو متغیر X و Y در چه صورتی مستقل هستند؟

۱. توزیع احتمال مشترک P(X,Y) را محاسبه کنید.

۲. توزیع احتمال حاشیه‌ای P(X) و P(Y) را به دست آورید.

۳. بررسی کنید که آیا:

$$\textbf{Independence Property} \rightarrow P(X,Y) = P(X)P(Y) \quad \forall X, Y$$

✓ اگر این رابطه برقرار باشد، متغیرها مستقل هستند. در غیر این صورت، وابسته‌اند.

Example: Coin Flip

Suppose we flip two fair coins:
Let X be 1 if the first coin lands heads, 0 otherwise.
Let Y be 1 if the second coin lands heads, 0 otherwise.

We can create a probability table:

X	Y	$P(X,Y)$	$P(X)P(Y)$
0	0	0.25	$0.5 \times 0.5 = \mathbf{0.25}$
0	1	0.25	$0.5 \times 0.5 = \mathbf{0.25}$
1	0	0.25	$0.5 \times 0.5 = \mathbf{0.25}$
1	1	0.25	$0.5 \times 0.5 = \mathbf{0.25}$

Since $P(X,Y) = P(X)P(Y)$ for all cases, the two coin flips are independent.

استقلال شرطی

گاهی دو متغیر X و Y با توجه به متغیر سوم Z مستقل می‌شوند:

$$P(X,Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$

مثال:

X: میزان مصرف فست‌فود
Y: احتمال ابتلا به بیماری قلبی
Z: سطح کلسترول

اگر سطح کلسترول Z را بدانیم، مصرف فست‌فود ممکن است دیگر تأثیری بر پیش‌بینی بیماری قلبی نداشته باشد.

در این حالت، X و Y با توجه به Z مستقل می‌شوند.

✓ استقلال آماری یعنی احتمال وقوع دو متغیر برابر حاصل‌ضرب احتمال‌های جداگانه آن‌ها باشد.

✓ استقلال شرطی یعنی دو متغیر با دانستن مقدار متغیر سوم مستقل می‌شوند.

۱. یک کارخانه قطعات معیوب و سالم تولید می‌کند. احتمال معیوب بودن یک قطعه ۵٪ و احتمال تولید آن از دستگاه A ۵۰٪ است. اگر این دو رویداد مستقل باشند، احتمال اینکه یک قطعه هم از دستگاه A باشد و هم معیوب چقدر است؟

۲. اگر دستگاه A ۷٪ قطعات معیوب و دستگاه B ۳٪ قطعات معیوب تولید کند، آیا رویدادهای "معیوب بودن" و "از دستگاه A بودن" هنوز مستقل هستند؟ چرا؟

احتمال شرطی

احتمال شرطی به احتمال وقوع یک رویداد A زمانی که اطلاعاتی در مورد رویداد دیگری B داریم، اشاره دارد. به عبارت دیگر،

احتمال شرطی P(A|B) احتمال وقوع A را با توجه به این که رویداد B رخ داده است، نشان می‌دهد.

فرمول احتمال شرطی به‌صورت زیر است:

Conditional Probability:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{if} \quad P(B) > 0$$

که در آن:
- P(A∩B) احتمال وقوع همزمان رویدادهای A و B.
- و P(B) احتمال وقوع رویدادB.

در یک مدرسه، دانش‌آموزان به طور متوسط ساعت‌های متفاوتی مطالعه می‌کنند و نمرات مختلفی دریافت می‌کنند. می‌خواهیم بررسی کنیم

که احتمال این که یک دانش‌آموز نمره‌ای بالاتر از ۱۵ بگیرد، مشروط بر این که بیش از ۵ ساعت در روز مطالعه کند، چقدر است.

تعاریف رویدادها:
رویداد A: دانش‌آموزی که نمره بالای ۱۵ در امتحان می‌گیرد.
رویداد B: دانش‌آموزی که بیش از ۵ ساعت در روز مطالعه می‌کند.
احتمال مورد نظر: P(A|B) یعنی احتمال این که یک دانش‌آموز نمره بالای ۱۵ بگیرد، مشروط به این که بیش از ۵ ساعت در روز مطالعه کرده باشد.

در بین ۱۰۰ دانش‌آموز:
۴۰ نفر بیش از ۵ ساعت در روز مطالعه می‌کنند. یعنی: $P(B) = \frac{40}{100} = 0.4$

از میان این ۴۰ دانش‌آموز، ۳۰ نفر نمره بالای ۱۵ می‌گیرند. یعنی: $P(A \cap B) = \frac{30}{100} = 0.3$

برای محاسبه احتمال شرطی داریم:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \rightarrow$$

$$P(A \mid B) = \frac{0.3}{0.4}$$

$$P(A \mid B) = 0.75$$

یکی دیگر از کاربردهای مهم قضیه بیز و احتمال شرطی، تشخیص ناهنجاری‌ها است. در اینجا، هدف شناسایی رفتارهای غیرمعمول یا ناهنجار در داده‌ها است.

به‌طور مثال، در سیستم‌های تشخیص تقلب، از فیلتر بیزی برای شبیه‌سازی رفتارهای عادی و ناهنجار استفاده می‌شود و

سپس داده‌های جدید برای ارزیابی احتمال ناهنجاری به‌کار می‌روند.