

Công thức Bayes

Giảng viên: PGS.TS. Lê Sỹ Vinh
Khoa CNTT – Đại học Công Nghệ

Kiểm tra ung thư

- 1% phụ nữ bị ung thư vú (99% không bị)
- **Nếu bị ung thư vú**, kiểm tra X-ray phát hiện 80% (20% không phát hiện được).
- **Nếu không bị ung thư vú**, 9.6% người bị X-ray trả lời là có ung thư vú (90.4% đúng).

	Bị ung thư vú	Không bị
X-ray dương tính (positive)	80%	9.6%
X-ray âm tính (negative)	20%	90.4%

H: Bị ung thư; D: X-ray dương tính; $P(D|H)$

Kiểm tra ung thư

	Bị ung thư vú	Không bị
X-ray dương tính (positive)	80%	9.6%
X-ray âm tính (negative)	20%	90.4%

Nếu bạn nhận kết quả X-ray bị ung thư vú, xác suất bạn bị ung thư vú là bao nhiêu?

Công thức Bayes

$$P(H | D) = \frac{P(D | H) \times P(H)}{P(D)}$$

Giả sử:

D: Dữ liệu quan sát được (Data), ví dụ D=X-ray dương tính

H: Giả thuyết (Hypothesis), ví dụ H=Ung thư vú

Ta có:

$P(H | D)$: Xác suất hậu nghiệm (posterior probability)

$P(D | H)$: Xác suất của dữ liệu với điều kiện giả thuyết H;

hay được gọi là khả năng (likelihood) của giả thuyết H.

$P(H)$: Xác suất tiên nghiệm của giả thuyết, được xác định trước khi quan sát D

$P(D)$: Xác suất tiên nghiệm của dữ liệu, giống nhau cho mọi giả thuyết

$$P(D) = P(D | H) * P(H) + P(D | \text{phủ định } H) * P(\text{phủ định } H)$$

Kiểm tra ung thư

	Bị ung thư vú	Không bị
X-ray dương tính (positive)	80%	9.6%
X-ray âm tính (negative)	20%	90.4%

Nếu bạn nhận kết quả X-ray bị ung thư vú, xác suất bạn bị ung thư vú là bao nhiêu?

Bài tập 1

Nhà máy có 2 xưởng A và B với tổng sản phẩm lần lượt là 40% và 60%. Xác suất 1 sản phẩm hỏng từ xưởng A và B lần lượt là 2%, và 5%. Biết một sản phẩm bị hỏng, tính xác suất sản phẩm đó từ xưởng A của nhà máy.

Bài tập 2

Nhà trường có 3 Khoa ĐT, CH, và CNTT với số sinh viên tương ứng là 20%, 30% và 50%. Xác suất 1 sinh viên không tốt nghiệp đúng hạn từ khoa ĐT, CH, CNTT lần lượt là 25%, 35% và 30%. Biết một sinh viên X không tốt nghiệp đúng hạn, tính xác suất sinh viên đó thuộc khoa CNTT.

Bài tập 2

Nhà trường có 3 Khoa ĐT, CH, và CNTT với số sinh viên tương ứng là 20%, 30% và 50%. Xác suất 1 sinh viên không tốt nghiệp đúng hạn từ khoa ĐT, CH, CNTT lần lượt là 25%, 35% và 30%. Biết một sinh viên X không tốt nghiệp đúng hạn, tính xác suất sinh viên đó thuộc khoa CNTT.

H: Sinh viên khoa CNTT

X: Sinh viên không tốt nghiệp

$$P(\text{CNTT} | X) = P(X | \text{CNTT}) * P(\text{CNTT}) / P(X) = 0.3 * 0.5 / P(X) = 0.49$$

$$\begin{aligned} P(X) &= P(X | \text{CNTT}) * P(\text{CNTT}) + P(X | \text{DT}) * P(\text{DT}) + P(X | \text{CH}) * P(\text{CH}) \\ &= 0.5 * 0.3 + 0.2 * 0.25 + 0.3 * 0.35 = 0.305 \end{aligned}$$