BAYES RULES IN MAMMOGRAM EXAMPLE

SYRACUSE UNIVERSITY
School of Information Studies

BAYES' THEOREM

Bayes' theorem lets us swap the order of the dependence between events.

Two events A and B

We know that P(A,B) = P(B|A)P(A)

Since P(A,B) = P(B,A), we also know that P(B,A) = P(B|A)P(A)

Therefore:

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

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

BACK TO THE MAMMOGRAM EXAMPLE IN PROFESSOR STROGATZ'S ARTICLE

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

Event A: A patient has cancer

Event B: A patient's mammogram test result is positive

P(B|A) = Among patients who have cancer, how many have positive test results?

We can calculate it from past data.

P(A|B) = For a patient with a positive test result, what is the chance of cancer?

This is **prediction**!

Bayes' theorem provides an approach to predict the probability of future events based on prior experience.

THE MAMMOGRAM EXAMPLE

The probability that a woman has breast cancer is … ?

If a woman has breast cancer, the probability that she will have a positive mammogram is … ?

If a woman does not have breast cancer, the probability that she will still have a positive mammogram is … ?

BACK TO THE MAMMOGRAM EXAMPLE

The probability that a woman has breast cancer is
P(cancer) = 0.008

If a woman has breast cancer, the probability that she will have a positive mammogram is
P(positive|cancer) = 0.9

If a woman does not have breast cancer, the probability that she will still have a positive mammogram is
P(positive|no cancer) = 0.07

TRANSLATE THEM INTO PROBABILITY NOTATIONS

Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer?

Which probability is greater:
P(cancer | positive) or P(no cancer | positive)?

To calculate P(cancer | positive), need to calculate:

- P(positive)
- P(cancer)
- P(positive | cancer)

To calculate P(no cancer | positive), need to calculate:

- P(positive)
- P(no cancer)
- P(positive | no cancer)

TRANSLATE THEM INTO PROBABILITY NOTATIONS

The probability that a woman has positive mammogram is:

P(positive)=Unknown

The probability that a woman has breast cancer is 0.008.

P(cancer) = 0.008, P(no cancer) = 0.992, (add up to 1)

If a woman has breast cancer, the probability that she will have a positive mammogram is 0.9.

P(positive | cancer) = 0.9

If a woman does not have breast cancer, the probability that she will still have a positive mammogram is 0.07.

P(positive | no cancer) = 0.07

TRANSLATE THEM INTO PROBABILITY NOTATIONS

All **prior probabilities** we have calculated:

P(cancer) = 0.008

P(no cancer) = 0.992

All **conditional probabilities** we have calculated:

P(positive | cancer) = 0.9

P(positive | no cancer) = 0.07

The **posterior probabilities** to be calculated:

P(cancer | positive) = ?   P(no cancer | positive) = ?

SO, OUR PREDICTION IS …

$$P(cancer \mid positive)$$

$$= \frac{P(positive \mid cancer) \cdot \ P(cancer)}{P(positive)} = \frac{0.9 \cdot \ 0.008}{P(positive)} = \frac{0.0072}{P(positive)}$$

$$P(no\_cancer \mid positive)$$

$$= \frac{P(positive \mid no\_cancer) \cdot \ P(no\_cancer)}{P(positive)} = \frac{0.07 \cdot \ 0.992}{P(positive)} = \frac{0.069}{P(positive)}$$

NO CANCER!

$$P(cancer \mid positive) = \frac{P(positive \mid cancer) \cdot P(cancer)}{P(positive)} = \frac{0.9 \cdot 0.008}{P(positive)} = \frac{0.0072}{P(positive)}$$

$$P(no\_cancer \mid positive) = \frac{P(positive \mid no\_cancer) \cdot P(no\_cancer)}{P(positive)} = \frac{0.07 \cdot 0.992}{P(positive)} = \frac{0.069}{P(positive)}$$

Although we don't know P(positive), it does not matter. We just need to know which posterior probability is greater.

THIS DIAGNOSIS IS DETERMINED BY THE MAMMOGRAM RESULT ONLY

$$P(cancer \mid positive)$$
$$= \frac{P(positive \mid cancer) \cdot P(cancer)}{P(positive)} = \frac{0.9 \cdot 0.008}{P(positive)} = \frac{0.0072}{P(positive)}$$

$$P(no\_cancer \mid positive)$$
$$= \frac{P(positive \mid no\_cancer) \cdot P(no\_cancer)}{P(positive)} = \frac{0.07 \cdot 0.992}{P(positive)} = \frac{0.069}{P(positive)}$$

WHAT IF THE DIAGNOSIS IS DETERMINED BY MORE FACTORS THAN JUST THE MAMMOGRAM RESULT?

Attribute 1: Positive mammogram? Yes or no
Attribute 2: Family history? Yes or no
Attribute 3: Alcohol? Yes or no

How many posteriors to calculate?
Two posteriors for each possible combination of the attributes.
In this case, $2 * 2^3 = 16$

SYRACUSE UNIVERSITY
School of Information Studies

ALL POSTERIOR PROBABILITIES FOR THREE BINARY ATTRIBUTES

$$P(positive, yes, yes \mid cancer)$$
$$P(positive, yes, no \mid cancer)$$
$$P(positive, no, yes \mid cancer)$$
$$P(positive, no, no \mid cancer)$$
$$P(negative, yes, yes \mid cancer)$$
$$P(negative, yes, no \mid cancer)$$
$$P(negative, no, yes \mid cancer)$$
$$P(negative, no, no \mid cancer)$$

$$P(positive, yes, yes \mid cancer)$$
$$P(positive, yes, no \mid no\_cancer)$$
$$P(positive, no, yes \mid no\_cancer)$$
$$P(positive, no, no \mid no\_cancer)$$
$$P(negative, yes, yes \mid no\_cancer)$$
$$P(negative, yes, no \mid no\_cancer)$$
$$P(negative, no, yes \mid no\_cancer)$$
$$P(negative, no, no \mid no\_cancer)$$