

Bayes' Theorem

Introduction

Bayes theorem is an indispensable law of probability, allowing you to deductively quantify unknown probabilities. The theory rests upon conditional probability. Let's take a look at it in practice.

Objectives

You will be able to:

- Define Bayes' theorem in relation to conditional probabilities
- Identify examples of applications of Bayes' theorem

Bayes' formula

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

Breaking the formula apart

Bayes' theorem is quite intuitive, decomposing the conditional probability of 'A given B' in terms of the probability that both events are true divided by the probability that B is true. Bayes theorem takes this natural idea a step further, expressing the probability that both events are true as a conditional probability multiplied by the condition itself.

To recap:

Bayes' Theorem takes the definition of the conditional likelihood:

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

and rewrites the $P(A \cap B)$ as $P(B|A)P(A)$, which makes perfect sense; the probability of B given A is true, multiplied by the probability that A is true, gives us the probability that both are true.

Making this substitution, you have Bayes' Theorem:

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

A simple example

Let's take a simple theoretical example to demonstrate. Imagine there are two fish tanks at the local pet store. The small tank holds 10 Betta fish. The large tank has 200 goldfish and 35 Betta fish. Given that a fish is a Betta fish, what's the probability it comes from the small tank?

On the one hand, it seems that if you were to select a fish from the large tank, you'd probably end up with a goldfish. However, because these tanks are of such vastly different sizes, the probability that the fish came from the larger tank is actually more probable.

Using Bayes' theorem, you are looking to find the probability that the fish came from the small tank, given that it is a Betta fish:

$$P(\text{small_tank} | \text{Betta_fish}) = \frac{P(\text{Betta_fish} | \text{small_tank})P(\text{small_tank})}{P(\text{Betta_fish})}$$

Furthermore, you know:

$$P(\text{Betta_fish} | \text{small_tank}) = 1$$
$$P(\text{small_tank}) = \frac{\text{number of fish in small tank}}{\text{number of all fish}} = \frac{10}{245}$$

$$P(\text{small_tank}) = \frac{10}{45} = \frac{2}{9}$$

$$P(\text{Betta_fish}) = \frac{45}{245}$$

Giving you:

$$P(\text{small_tank} | \text{Betta_fish}) = \frac{1 \cdot \frac{10}{245}}{\frac{45}{245}}$$

$$P(\text{small_tank} | \text{Betta_fish}) = \frac{10}{45}$$

While concrete, this example fails to demonstrate the full power of Bayes' theorem since you had all of the underlying information, so you don't even need to use Bayes' theorem. You could have simply looked at the number of Betta fish in the small tank versus the number of Betta fish overall:

$$\frac{10}{45}$$

giving you exactly the same result.

An NLP example

With this simple example out of the way, let's examine a more practical example from the field of Natural Language Processing.

A common introductory example to Natural Language Processing or classification is detecting spam. While you may enjoy spam in a can, you probably don't enjoy getting spam in your inbox. Bayes' theorem can serve as a natural classification method in these scenarios. Assume that the word "offer" (as in Special Offer, We Have an Offer for You, or Don't Miss This Offer!) occurs in 73% of the spam messages you receive. In comparison, only 10% of your desired mail contains the word "offer". If 20% of the messages you receive are spam, and you receive another message with the word "offer", what is the probability that it is spam?

As you might have guessed, you can solve this using the Bayes' theorem!

First, set up the problem:

$$P(\text{Spam} | \text{Offer}) = \frac{P(\text{Offer} | \text{Spam})P(\text{Spam})}{P(\text{Offer})}$$

Then substituting some of the immediate knowledge we have from the scenario:

$$P(\text{Spam} | \text{Offer}) = \frac{.73 \cdot .20}{P(\text{Offer})}$$

Finally, the probability of receiving an email with the word "offer", $P(\text{Offer})$, can be evaluated by decomposing it into the two subsets spam and not spam:

$$P(\text{Offer}) = P(\text{Spam}) \cdot P(\text{Offer} | \text{Spam}) + P(\sim\text{Spam}) \cdot P(\text{Offer} | \sim\text{Spam})$$

$$P(\text{Offer}) = .20 \cdot .73 + .8 \cdot .10$$

$$P(\text{Offer}) = .146 + .08$$

$$P(\text{Offer}) = .226$$

Finally, substituting this into the original Bayes formula you have:

$$P(\text{Spam} | \text{Offer}) = \frac{.73 \cdot .20}{P(\text{Offer})}$$

$$P(\text{Spam} | \text{Offer}) = \frac{.73 \cdot .20}{.226}$$

$$P(\text{Spam} | \text{Offer}) = .6460$$

As you can see, while spam has a much higher occurrence of the word "offer", the presence of the word alone does not provide strong confidence that the message is spam. To provide more statistical power, you will eventually extend Bayes' theorem to multiple observations simultaneously using the relative probabilities of multiple words.

Summary

In this lesson, you were introduced to the Bayes' theorem, and saw how it can be used to quantify conditional probabilities. With that, let's turn to some more simple examples for you to practice and deepen your understanding.

that, let's turn to some more simple examples for you to practice and deepen your understanding.