Bayes Theorem

Vibha Mane
The COSINE Lab
Department of Electrical and Computer Engineering

Topics

- Bayes Theorem
- Bayes Theorem Illustration
- Some Examples

Bayes' Theorem

- Recall from the conditional probability definition:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A).$$

- **Bayes' theorem** is given by the following equation:

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

- Using the result from **total probability theorem** in the denominator, $\mathrm{P(B)} = P(B|A)P(A) + P(B|A^C)P(A^C)$, we can rewrite this as:

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

Bayes' Theorem Interpretation

- The various probabilities in the Bayes' Theorem have the following interpretation:
  - $P(A)$ is the **prior probability** or the initial degree of belief in the event A.
  - B is data or measured event, and $P(B|A)$ is the **likelihood** of the data given the event A.
  - $P(A|B)$ is the **posterior probability** of the event A given the prior belief and the data.
- Omitting the denominator, we have:

$$\underbrace{P(A|B)}_{\text{posterior}} \propto \underbrace{P(B|A)}_{\text{likelihood}}\underbrace{P(A)}_{\text{prior}}.$$

- Therefore, Bayes' Theorem helps us update the probability of an event, given prior and likelihood.

Bayes' Theorem Illustration (1 of 4)

Josh Tenenbaum, 1999

- Consider a simple example of **concept learning**, called the number game. We choose some simple arithmetical concept, such as even numbers or multiples of 3. You are then given a randomly chosen sample $D = \{x_1, x_2, \ldots x_N\}$ of positive examples from C, and asked to classify a new test case x, that is, does it belong to C or not.

- Let us assume that all numbers are between 1 and 100. Let H be the **hypothesis space** of concepts, such as:
  - odd numbers
  - even numbers
  - Squares
  - multiples of
  - powers of
  - …

Bayes' Theorem Illustration (2 of 4)

- Suppose you are given data $D = \{16, 8, 2, 64\}$. You may guess that the hidden concept is **"powers of 2"**. However, there are other concepts, such as **"even numbers"**, which are also consistent with the data.

- The key intuition is that we want to avoid **suspicious coincidences**. In order to explain why we chose $h_{two} \equiv$ "powers of 2" and not $h_{even} \equiv$ "even numbers", we look at the **likelihood, $p(D|h)$**. This is the probability of data D given hypothesis h.

- Since $h_{two} = \{2, 4, 8, 16, 32, 64\}$ and $h_{even} = \{2, 4, 6, \ldots 100\}$, with 4 samples we have the likelihoods,

  $p(D|h_{two}) = (1/6)^4 = 7.7 \times 10^{-4}$ and

  $p(D|h_{even}) = (1/50)^4 = 1.6 \times 10^{-7}$.

  This is a likelihood ratio of $5000:1$ in favor of h_{two}.

Bayes’ Theorem Illustration (3 of 4)

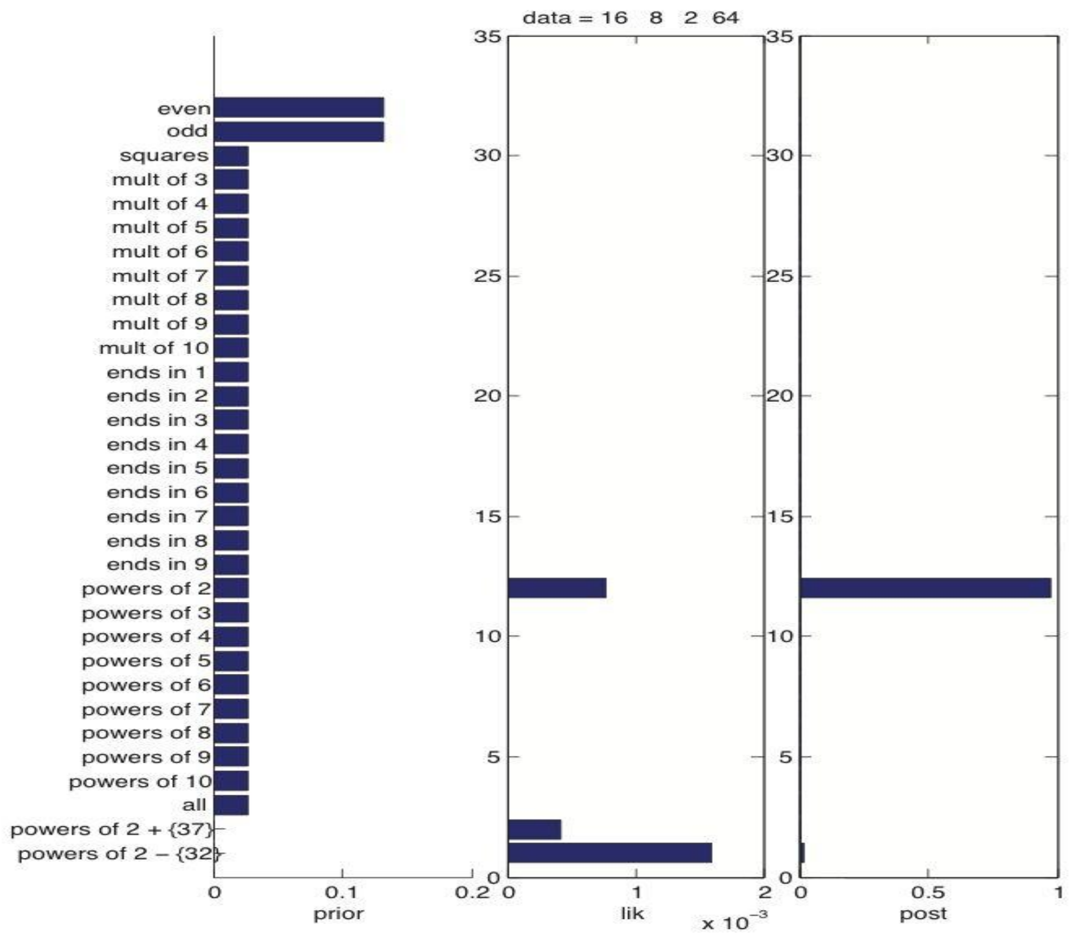
- Given the data $D = \{16, 8, 2, 64\}$, the hypothesis **“powers of 2 except 32”** is more likely than **“powers of 2”**. However, “powers of 2 except 32” seems **conceptually unnatural**.
- We can capture such intuition by assigning low **prior probability** to unnatural concepts.
- Therefore, **predictive posterior distribution** is proportional to the likelihood times the prior:

$$P(A|B) \propto P(B|A)P(A).$$

posterior likelihood prior

Bayes' Theorem Illustration (4 of 4)

This figure shows the prior, likelihood and posterior for data $D = \{16, 8, 2, 64\}$.



Example - Defective Products (1 of 3)

Example

We have two batches of products, where some of the products are defective. We are given the following information for each batch, where T represents total and D represents defective:

Batch B_1: $600\ T,\ 200\ D$

Batch B_2: $1000\ T,\ 500\ D$

We draw a product at random from one of the two batches and find that it is defective. What is the probability that it is drawn from batch B_2? We are given that the probability of drawing from either batch is equal.

Example - Defective Products (2 of 3)

Solution:

We **know the outcome:** Defective Product, and we want to determine which batch?

Therefore, we utilize **Bayes' Theorem**,

$$P(drawn\ from\ B_2|D) = \frac{P(D|drawn\ from\ B_2)\,P(drawn\ from\ B_2)}{P(D)}$$

$P(drawn\ from\ B_2) = {}^{1}/_{2}$, since **equal probability** of drawing from each batch.

$P(D|drawn\ from\ B_2) = {}^{500}/_{1000} = {}^{1}/_{2}$, since 500 out of 1000 in B_2 are defective.

Example - Defective Products (3 of 3)

We next compute the denominator $P(D)$ from **total probability theorem**.

$$P(D) = P(D|\, B_1)P(\, B_1) + P(D|\, B_2)P(\, B_2)$$

$$= \frac{200}{600}\frac{1}{2} + \frac{500}{1000}\frac{1}{2} = 0.416.$$

Therefore, $P(drawn\ from\ B_2|D) = \frac{{}^1/_2 \times {}^1/_2}{0.416} \cong 0.6$

The denominator in Bayes' Theorem is computed from total probability theorem.

Example - False Positives (1 of 4)

Example

A rare disease affects one person in 10^5. A test for the disease shows positive with a probability $^{99}/_{100}$ when applied to an ill person, and it shows positive with a probability $^{3}/_{100}$ when applied to a healthy person. What is the probability that a person has the disease given that the test shows positive?

Solution:

We **know the outcome:** Positive Test, and we want to determine if the person has disease or not?

Therefore, we utilize **Bayes' Theorem**

$$P(D|S) = \frac{P(S|D)P(D)}{P(S)}$$

Example - False Positive (2 of 4)

where D is the event that the person has disease and S is the event that the test shows positive for a person.

We have the following values for the probabilities:

Probability of disease in population $= P(D) = \frac{1}{10^5}$

Probability of healthy in population $= P(D^C) = 1 - \frac{1}{10^5}$

Probability of positive given disease $= P(S|D) = \frac{99}{100}$

Probability of positive given healthy $= P(S|D^C) = \frac{3}{100}$

Example - False Positive (3 of 4)

We compute the denominator $P(S)$ from the **total probability theorem**:

$$P(S) = P(S|D)P(D) + P(S|D^C)P(D^C)$$

$$= \frac{99}{100}\frac{1}{10^5} + \frac{3}{100}\frac{99999}{10^5}$$

And the required probability is

$$P(D|S) = \frac{\frac{99}{100}\frac{1}{10^5}}{\frac{99}{100}\frac{1}{10^5} + \frac{3}{100}\frac{99999}{10^5}} \cong \frac{3}{10000}$$

Example - False Positive (4 of 4)

This example illustrates that in this scenario for a **rare disease**, the probability that a person has a disease is very small even if they test positive.

References

1. Charles Boncelet, Probability, Statistics and Random Signals, Oxford University Press, 2016.
2. Sheldon Ross, A First Course in Probability, Macmillan Publishing Company, 1988.
3. R. D. Yates, et al., Probability and Stochastic Processes, John Wiley, 2005.
4. Josh Tenenbaum, A Bayesian framework for concept learning, PhD thesis, MIT, 1999.