Programme Name: BCA Semester V
Course Code: BCA57203(T)
Course Name: Artificial Intelligence
Class: BCA 2023
Academic Session: 2025-26

Study Material

Table of Contents

Module No.	**Module Name**	**Content**
Module 2	**Knowledge Representation Scheme and Reasoning Concept**	**Bayes' theorem**

Bayes’ Theorem in Artificial Intelligence

Motivation: Why Bayes’ Theorem?

In Artificial Intelligence, we often deal with **uncertainty**. AI systems rarely know the truth with certainty. Instead, they gather **evidence** (data) and must decide which **hypothesis** is most likely true.

Examples in AI:

- Is this email spam or not?
- Does a patient have a disease based on symptoms?
- Is the detected object in an image a cat or a dog?

Bayes’ Theorem gives us a systematic way to update our beliefs when new evidence arrives.

1 Conditional Probability

Conditional probability tells us the probability of an event A happening given that another event B has already occurred:

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

It gives a way to update our beliefs when we have partial information.
Suppose we randomly select a student from a class.

- 40% of the students are female.
- 30% of the students study Computer Science (CS).
- 20% are female *and* study CS.

Let event A = "student is female" and event B = "student studies CS". We want $P(A|B)$, the probability that a randomly chosen CS student is female.

From the definition:

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

We know:

$$P(A \cap B) = 0.20, \quad P(B) = 0.30.$$

So:

$$P(A|B) = \frac{0.20}{0.30} = \frac{2}{3} \approx 0.67.$$

That is, about 67% of CS students are female.

2 Deriving Bayes' Theorem

Suppose we want $P(A|B)$. From conditional probability:

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

Rearranging,

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

This is **Bayes' theorem**.

- $P(A)$: **Prior** (belief before evidence)
- $P(B|A)$: **Likelihood** (how likely is evidence if hypothesis is true)
- $P(B)$: **Evidence / Normalizing factor**
- $P(A|B)$: **Posterior** (updated belief after evidence)

The Intuition (No Math Background Needed)

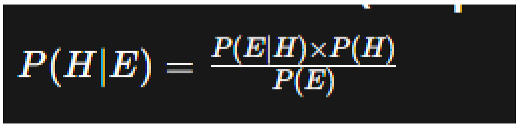
Bayes’ theorem is about **updating beliefs**:

New Belief = How well evidence matches hypothesis × Old Belief ÷ Overall Evidence

- **Old Belief (Prior)**: What we believed before seeing evidence.
- **Evidence Likelihood**: How likely the evidence is if the hypothesis were true.
- **New Belief (Posterior)**: What we believe now after seeing evidence.

Think of it like detective work: we start with a hunch (prior), find a clue (evidence), and update our suspicion level (posterior).

The Formula (Simple Form)

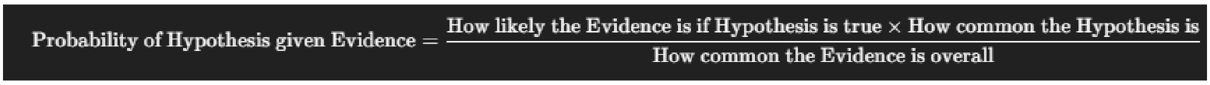


$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Where:

- **H** = Hypothesis (e.g., "The email is spam")
- **E** = Evidence (e.g., "The email contains 'lottery'")
- **P(H)** = Prior probability (before evidence)
- **P(E|H)** = Likelihood (how often we see this evidence if hypothesis is true)
- **P(H|E)** = Posterior probability (after evidence)
- **P(E)** = Total probability of evidence across all hypotheses

Even without math background, here’s the idea:



$$\text{Probability of Hypothesis given Evidence} = \frac{\text{How likely the Evidence is if Hypothesis is true} \times \text{How common the Hypothesis is}}{\text{How common the Evidence is overall}}$$

In words:

New Belief = Evidence × Prior Belief ÷ Normalization

3 Denominator: Law of Total Probability

The basic form of Bayes' theorem is:

$$P(Y|X) = \frac{P(X|Y)\,P(Y)}{P(X)}.$$

Law of Total Probability

The denominator $P(X)$ can be expressed using the law of total probability:

$$P(X) = \sum_{y\in\mathcal{Y}} P(X|Y=y)\,P(Y=y).$$

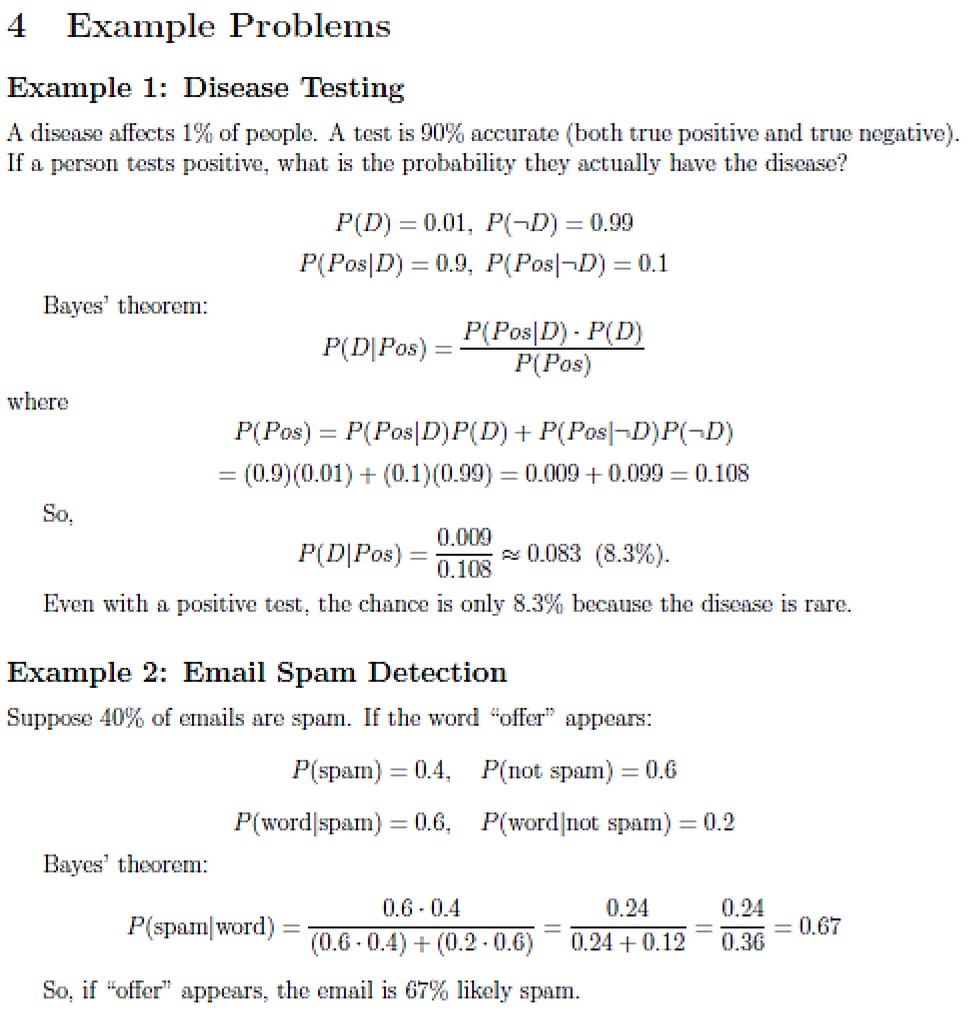
Thus, Bayes' theorem becomes:

$$P(Y|X) = \frac{P(X|Y)\,P(Y)}{\sum_{y\in\mathcal{Y}} P(X|Y=y)\,P(Y=y)}.$$

The denominator in Bayes' theorem is:

$$P(\text{Evidence}) = \sum_{y} P(\text{Evidence}|y)\cdot P(y).$$

This ensures probabilities add up to 1 across all possible classes.



4 Example Problems

Example 1: Disease Testing

A disease affects 1% of people. A test is 90% accurate (both true positive and true negative). If a person tests positive, what is the probability they actually have the disease?

$$P(D) = 0.01,\ \ P(\neg D) = 0.99$$
$$P(Pos|D) = 0.9,\ \ P(Pos|\neg D) = 0.1$$

Bayes' theorem:

$$P(D|Pos) = \frac{P(Pos|D) \cdot P(D)}{P(Pos)}$$

where

$$P(Pos) = P(Pos|D)P(D) + P(Pos|\neg D)P(\neg D)$$
$$= (0.9)(0.01) + (0.1)(0.99) = 0.009 + 0.099 = 0.108$$

So,

$$P(D|Pos) = \frac{0.009}{0.108} \approx 0.083 \ \ (8.3\%).$$

Even with a positive test, the chance is only 8.3% because the disease is rare.

Example 2: Email Spam Detection

Suppose 40% of emails are spam. If the word "offer" appears:

$$P(\text{spam}) = 0.4, \quad P(\text{not spam}) = 0.6$$

$$P(\text{word}|\text{spam}) = 0.6, \quad P(\text{word}|\text{not spam}) = 0.2$$

Bayes' theorem:

$$P(\text{spam}|\text{word}) = \frac{0.6 \cdot 0.4}{(0.6 \cdot 0.4) + (0.2 \cdot 0.6)} = \frac{0.24}{0.24 + 0.12} = \frac{0.24}{0.36} = 0.67$$

So, if "offer" appears, the email is 67% likely spam.

Numerical Example

1: Medical Diagnosis

Problem:

- 1% of people have a rare disease.
- A test detects the disease correctly 99% of the time if you have it.
- The test gives a false positive in 5% of healthy people.

If you test **positive**, what is the probability you actually have the disease?

Step 1: Translate into probabilities

- P(Disease) = 0.01
- P(No Disease) = 0.99
- P(Test+|Disease) = 0.99

- P(Test+|No Disease) = 0.05

Step 2: Compute total chance of positive test

- True positive = 0.99 × 0.01 = 0.0099
- False positive = 0.05 × 0.99 = 0.0495
- Total positive = 0.0099 + 0.0495 = 0.0594

Step 3: Apply Bayes’ theorem

P(Disease|Test+)=0.00990.0594≈0.167P(Disease|Test+) = \frac{0.0099}{0.0594} ≈ 0.167

Answer: Only **16.7% chance** you actually have the disease, even though the test is "99% accurate".

2: Spam Filter

Problem:

- 40% of emails are spam.
- In spam emails, 80% contain the word "lottery".
- In normal emails, 5% contain "lottery".

If an email contains "lottery", what’s the chance it is spam?

Step 1: Translate into probabilities

- P(Spam) = 0.40
- P(Not Spam) = 0.60
- P(Lottery|Spam) = 0.80
- P(Lottery|Not Spam) = 0.05

Step 2: Compute total chance of seeing "lottery"

- Spam contribution = 0.40 × 0.80 = 0.32
- Not spam contribution = 0.60 × 0.05 = 0.03
- Total = 0.32 + 0.03 = 0.35

Step 3: Apply Bayes’ theorem

P(Spam|Lottery)=0.320.35≈0.914P(Spam|Lottery) = \frac{0.32}{0.35} ≈ 0.914

Answer: About **91% chance** it’s spam.

3: Weather Prediction

Problem:

- Weather forecast says: 30% chance of rain on any given day.
- If it rains, 90% of people carry umbrellas.
- If it does not rain, 20% of people still carry umbrellas (just in case).

You see a person with an umbrella. What's the probability it's raining?

Step 1: Translate into probabilities

- P(Rain) = 0.30
- P(No Rain) = 0.70
- P(Umbrella|Rain) = 0.90
- P(Umbrella|No Rain) = 0.20

Step 2: Compute total chance of seeing umbrella

- Rain contribution = 0.30 × 0.90 = 0.27
- No rain contribution = 0.70 × 0.20 = 0.14
- Total = 0.27 + 0.14 = 0.41

Step 3: Apply Bayes' theorem

P(Rain|Umbrella)=0.270.41≈0.658P(Rain|Umbrella) = \frac{0.27}{0.41} ≈ 0.658

Answer: About **66% chance** it's raining if you see someone carrying an umbrella.

Classification Examples

Each example below shows how to **compute posterior probabilities** using Bayes' theorem and then **classify** a new data point. For stability we show both probability-space calculations and **log-score** computations (unnormalized). At the end of each example we make a final class decision.

Example A — Categorical Naïve Bayes (Email Classification)

Scenario: You have two classes: **Spam** and **Ham** (not spam). From training data you observe these frequencies:

- Prior: P(Spam)=0.40, P(Ham)=0.60
- Word probabilities (estimated from training):
    - P("free"|Spam)=0.70, P("win"|Spam)=0.50
    - P("free"|Ham)=0.02, P("win"|Ham)=0.01

New email: contains the words {"free", "win"} (both present). Assume conditional independence (Naïve Bayes).

Step 1 — Compute likelihoods P(Evidence|Class)

- P(E|Spam) = P("free"|Spam) × P("win"|Spam) = 0.70 × 0.50 = **0.3500**
- P(E|Ham) = P("free"|Ham) × P("win"|Ham) = 0.02 × 0.01 = **0.0002**

Step 2 — Compute numerators (prior × likelihood)

- Spam: 0.40 × 0.3500 = **0.1400**
- Ham: 0.60 × 0.0002 = **0.00012**
- Evidence total = 0.1400 + 0.00012 = **0.14012**

Step 3 — Posterior probabilities

- P(Spam|E) = 0.1400 / 0.14012 ≈ **0.9991**
- P(Ham|E) = 0.00012 / 0.14012 ≈ **0.0009**

Decision: Classify as **Spam** (very high posterior).

Applications of Bayes’ Theorem in AI

- **Spam filtering** (decide if email is spam or not)
- **Medical diagnosis** (predict disease from symptoms)
- **Speech recognition** (guessing words from noisy sounds)
- **Computer vision** (recognizing objects under uncertainty)
- **Robotics** (localization: where am I given sensor readings?)

Practice Questions

1. If 20% of students fail a class, and 70% of failing students skip lectures, while only 10% of passing students skip lectures, what is the probability a student who skips lectures is failing?
2. In an email system, 50% of emails are spam. 60% of spam emails contain “discount”, while 2% of normal emails contain “discount”. If an email has “discount”, what is the chance it’s spam?
3. A factory produces 1% defective items. A test detects defectives with 95% accuracy, but has 10% false positives. If the test says defective, what is the probability the item is truly defective?