UNIT 5

I. Probability: Define Probability, Joint Probability, Conditional Probability With Example:

Probability

Probability is a way to quantify uncertainty and represents the likelihood that a particular event will occur. It is expressed as a number between 0 and 1, where 0 indicates that the event will not happen, and 1 indicates that it will certainly happen.

Example: For instance, in a weather prediction model, if there is a 70% probability of rain tomorrow, it means that out of 100 similar days, it would rain on approximately 70 of those days. In AI, such probabilities are essential for making informed decisions, like whether to carry an umbrella.

Joint Probability

Joint probability refers to the likelihood of two (or more) events occurring at the same time. It combines the probabilities of different events and is useful for understanding the relationships between them.

Example: Consider a spam detection system. Let's define:

Event A: The email contains the word "free."
Event B: The email is classified as spam.

The joint probability $P(A \cap B)$ represents the likelihood that an email both contains the word "free" and is categorized as spam. For

example, if 30 out of 100 spam emails contain the word "free," the joint probability might be $P(A \cap B) = 30/100 = 0.3$
This information is valuable for the spam filter to identify spam emails based on multiple indicators.

Conditional Probability

Conditional probability measures the likelihood of one event occurring, given that another event has already happened. It helps update our beliefs based on new information.

Example: In a movie recommendation system, let's define:

Event A: A user likes action movies.
Event B: A user will like a specific action movie (e.g., "Fast & Furious").

The conditional probability $P(A|B)$ would be the likelihood that a user who enjoys action movies will also like "Fast & Furious." If data shows that 80 out of 100 users who like action movies also enjoy "Fast & Furious," then $P(A|B) = 80/100 = 0.8$

This allows the recommendation system to suggest movies that are more likely to align with user preferences.

II. Markov Processes: Markov Processes, Markov's Decision Process, Hidden Markov Model.

Markov Processes:

A Markov process is a mathematical system that undergoes transitions from one state to another based on certain probabilistic rules. It has the "memoryless" property, meaning the next state depends only on the current state, not on the sequence of events that preceded it.

Key Features:

- **States**: The different conditions or situations the process can be in (e.g., weather can be sunny, rainy, or cloudy).
- **Transitions**: The probabilities of moving from one state to another. For example, if it's sunny today, there might be a 70% chance it will be sunny tomorrow and a 30% chance it will be rainy.
- **Markov Property**: The future state depends only on the current state, not the past states.

Example:
Consider a simple weather model:

States: Sunny, Rainy
Transition probabilities:

Current State	Next State	Transition Probability
Sunny	Sunny	0.8
Sunny	Rainy	0.2
Rainy	Sunny	0.4
Rainy	Rainy	0.6

DRAW DIAGRAM ON YOUR OWN

Markov's Decision Process (MDP)

A **Markov Decision Process** is an extension of the Markov process that incorporates decision-making. It is used in situations where an agent must choose actions to maximize a reward over time while operating in a stochastic (random) environment.

Key Components:

- **States (S):** Possible situations the agent can be in.
- **Actions (A):** Choices available to the agent in each state.
- **Transition Model (T):** Defines the probability of moving to a new state after taking an action in the current state.
- **Reward Function (R):** Provides feedback to the agent, assigning a value (reward) for each action taken in a state.
- **Policy (π):** A strategy that specifies the action to take in each state to maximize rewards.

Example:

Imagine a robot navigating a grid:

- **States**: Each cell in the grid.
- **Actions**: Move up, down, left, right.
- **Transition Model:** Moving up from a cell might lead to another cell with a 90% chance, while a 10% chance could lead to a wall (and no movement).
- **Reward Function:** The robot gets +1 for reaching a goal and -1 for hitting a wall.

The robot uses a policy to determine its best moves based on the expected rewards.

Hidden Markov Model (HMM)

A **Hidden Markov Model** is a statistical model where the system being modeled is assumed to be a Markov process with hidden (unobserved) states. Unlike a standard Markov process, you cannot see the current state directly; instead, you see some observable outcomes that depend on the hidden states.

Key Features:

- **Hidden States:** The actual states of the system that are not directly observable.
- **Observations:** The visible outputs that depend on the hidden states.
- **Transition Probabilities:** The probabilities of moving from one hidden state to another.
- **Emission Probabilities:** The probabilities of observing a particular output from a hidden state.

Example:

Consider a weather system where you can't see the weather directly, but you can hear a weather report:

- **Hidden States:** Sunny, Rainy
- **Observations:** "Good weather" or "Bad weather" based on the actual state.
- **Transition Probabilities:** Similar to the Markov process.
- **Emission Probabilities:** If it's sunny, there might be an 80% chance of hearing "Good weather," and if it's rainy, a 70% chance of hearing "Bad weather."

We use HMMs to infer the most likely weather (hidden state) based on the reports you hear (observations).

III. Bayes Theorem:

Bayes' Theorem is a fundamental theorem in probability and statistics that describes how to update the probability of a hypothesis based on new evidence. The theorem shows how prior knowledge (initial beliefs) and new information can be combined to form a revised probability.

Formula:
$P(A|B)=P(B|A)\cdot P(A)/P(B)$

Where:

- **P(A|B):** The probability of event A occurring given that B has occurred (posterior probability).
- **P(B|A):** The probability of event B given A has occurred (likelihood).
- **P(A):** The initial probability of A occurring (prior probability).
- **P(B):** The probability of B occurring (evidence).

Uses:
- **Medical Diagnosis:** Bayes' Theorem is used to update the probability of a disease given test results.
- **Spam Detection:** It helps determine the probability of an email being spam based on certain words in the message.
- **Machine Learning:** Used in Bayesian inference methods to update model parameters as new data becomes available.

Limitations:
- **Dependence on Prior:** If the prior probability P(A) is not well-estimated, the results can be misleading.
- **Computational Complexity:** For complex models with many variables, calculating Bayes' Theorem becomes computationally intensive.
- **Assumptions of Independence**: In some applications (e.g., Naive Bayes), independence assumptions may not hold, leading to inaccurate results.

IV. Bayesian Belief Networks (BBNs)

A **Bayesian Belief Network (BBN),** or Bayesian Network, is a graphical model that represents probabilistic relationships among a set of variables. It allows for reasoning under uncertainty by encoding dependencies between variables using directed edges.

Structure:

Nodes: Represent random variables.

Edges: Directed edges between nodes represent conditional dependencies.

Conditional Probability Tables (CPTs): Each node has a table specifying the probability of each possible value, given its parent nodes

Uses:

Bayesian Networks are widely used in fields such as:

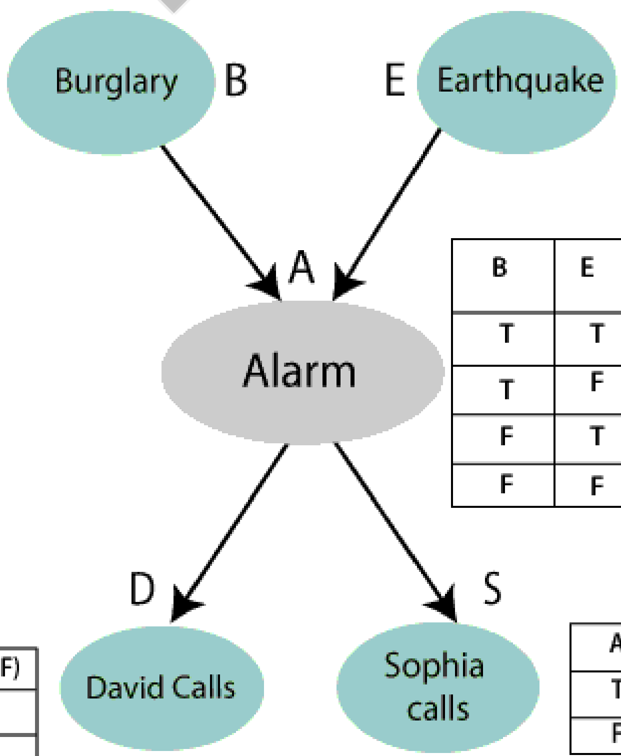
Medical Diagnosis: To determine the probability of diseases given symptoms.

Decision Support Systems: To model complex decisions with uncertain outcomes.

Risk Assessment: For identifying risks in complex systems like finance or engineering.

For example:

T	0.002
F	0.998



T	0.001
F	0.999

B	E	P(A=T)	P(A=F)
T	T	0.94	0.06
T	F	0.95	0.04
F	T	0.69	0.69
F	F	0.999	0.999

A	P (D=T)	P (D=F)
T	0.91	0.09
F	0.05	0.95

A	P (S=T)	P (S=F)
T	0.75	0.25
F	0.02	0.98

Belief Propagation in Bayesian Networks:

Belief propagation is a technique used to calculate the probability of a certain outcome by updating beliefs throughout the network based on evidence.

- **Evidence Setting**: Evidence is provided for one or more nodes (observed variables).
- **Message Passing:** Beliefs (probabilities) are passed from one node to its neighboring nodes. This is done in two ways:
- **Upward Pass (Bottom-Up)**: Starting from observed nodes, probabilities are propagated upward toward root nodes.

Downward Pass (Top-Down): The belief is propagated back from root nodes to other connected nodes.
Updating Beliefs: Each node updates its belief by recalculating probabilities based on incoming messages (using Bayes' Theorem or CPTs).

For example, in a medical Bayesian Network, if a symptom (evidence) is observed, the network updates the belief in potential diseases (causes) and propagates this belief to adjust the likelihood of other related symptoms or diseases accordingly.

V. Fuzzy Logic and Fuzzy Sets & Operations on Fuzzy Sets:

__**Fuzzy Logic**__ is an extension of traditional Boolean logic designed to handle the concept of partial truth. In Boolean logic, statements are either completely true or completely false (1 or 0). However, in real life, many situations are not black and white; they involve varying degrees of truth. Fuzzy logic allows truth values between 0 and 1, accommodating situations that are ambiguous or uncertain. This makes it very useful in areas like control systems, decision-making, and pattern recognition.

Fuzzy Sets are a fundamental concept in fuzzy logic. Unlike classical sets, where elements either belong to the set (1) or do not belong (0), fuzzy sets allow for partial membership. Each element in a fuzzy set has a membership degree between 0 and 1, representing how much it belongs to that set.

For example, let's say we have a fuzzy set "Tall" representing people's heights:

- A person with height 5.6 feet might belong to the set "Tall" with a membership of 0.6.
- A person with height 6.2 feet might belong with a membership of 0.9.

This way, fuzzy sets capture the gradual transition rather than a strict cutoff, making them ideal for real-world applications with vagueness.

Operations on Fuzzy Sets:

Just like classical sets, fuzzy sets support various operations such as union, intersection, and complement, but they are adapted to handle partial memberships.

1. **Union (OR Operation)**
   The union of two fuzzy sets A and B results in a new fuzzy set C, where the membership of each element in C is the maximum of its memberships in A and B.

   Formula:
   $\mu_C(x)=\max(\mu_A(x),\mu_B(x))$

   Example:

Suppose we have two fuzzy sets representing two different criteria:

Set A: “Moderately Tall” heights, with memberships:
5.5 feet: 0.5
6.0 feet: 0.8
Set B: “Very Tall” heights, with memberships:
5.5 feet: 0.3
6.0 feet: 0.9

The **union $C=A\cup B$** will have:
5.5 feet: max(0.5,0.3)=0.5
6.0 feet: max(0.8,0.9)=0.9

2. **Intersection (AND Operation)**
The intersection of two fuzzy sets A and B results in a new fuzzy set D, where the membership of each element in D is the minimum of its memberships in A and B.

Formula:
$\mu_D(x)=\min(\mu_A(x),\mu_B(x))$

Example:

Using the same sets as in the union example:

Set A: “Moderately Tall” with memberships:
5.5 feet: 0.5
6.0 feet: 0.8
Set B: “Very Tall” with memberships:
5.5 feet: 0.3
6.0 feet: 0.9

The **intersection $D=A\cap B$** will have:
5.5 feet: min(0.5,0.3)=0.3min(0.5,0.3)=0.3

6.0 feet: min(0.8,0.9)=0.8min(0.8,0.9)=0.8

3. **Complement (NOT Operation)**
   The complement of a fuzzy set A results in a new fuzzy set E, where the membership of each element in E is 1 minus its membership in A.

   Formula:
   $\mu_E(x)=1-\mu_A(x)$

   Example:
   Suppose Set A is “Moderately Tall” with memberships:
   5.5 feet: 0.5
   6.0 feet: 0.8

   The complement E=A’ will have:
   5.5 feet: 1−0.5=0.51−0.5=0.56.0 feet: 1−0.8=0.21−0.8=0.2