Elective Course
Course Code: CS4103
Autumn 2025-26

Lecture #29

Artificial Intelligence for Data Science

Week-8:

Probabilistic Reasoning [Part-I]

Reasoning Under Uncertainty, Bayesian Network

Course Instructor:

Dr. Monidipa Das

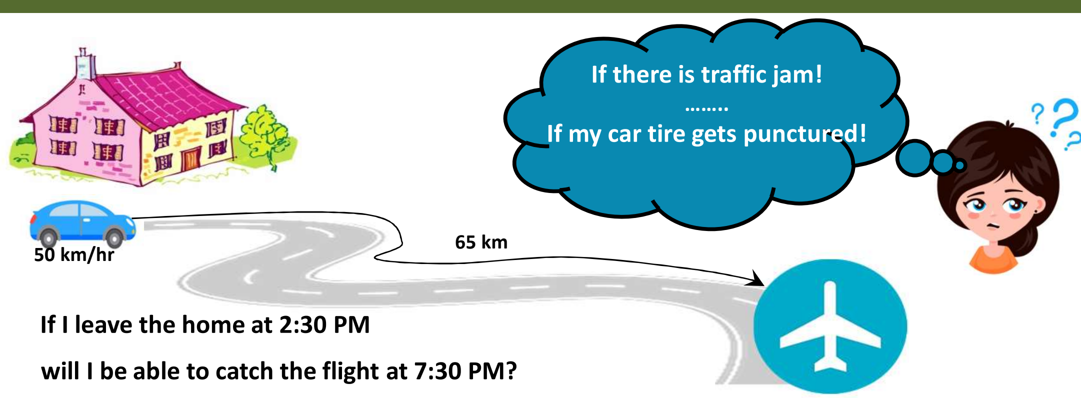
Assistant Professor

Department of Computational and Data Sciences

Indian Institute of Science Education and Research Kolkata, India 741246

Probabilistic Reasoning?

2



- Reasoning under ***uncertainty***
- Probabilistic reasoning only gives probabilistic results (summarizes uncertainty from various sources)

Sources of Uncertainty

3

- Uncertain **inputs**
  - Missing data, Noisy data
- Uncertain **knowledge**
  - Multiple causes lead to multiple effects
  - Incomplete enumeration of conditions or effects
  - Incomplete knowledge of causality in the domain
  - Probabilistic/stochastic effects
- Uncertain **outputs**

Reasoning Under Uncertainty

4

- Some major formalisms for representing and reasoning about uncertainty
  - **Probability theory (e.g. Bayesian belief networks)**
  - Dempster-Shafer theory
  - Fuzzy logic
  - Non-monotonic reasoning

Uncertainty Tradeoffs

5

- **Bayesian networks:** Nice theoretical properties combined with efficient reasoning make BNs very popular; limited expressiveness, knowledge engineering challenges may limit uses
- **Nonmonotonic logic:** Represent commonsense reasoning, but can be computationally very expensive
- **Dempster-Shafer theory:** Has nice formal properties, but can be computationally expensive, and intervals tend to grow towards [0,1] (not a very useful conclusion)
- **Fuzzy reasoning:** Semantics are unclear (fuzzy!), but has proved very useful for commercial applications

Probability Theory

6

- **Random variables**
  - Domain
- Alarm, Burglary, Earthquake
  - Boolean (like these), discrete, continuous
- **Atomic event**: complete specification of state
- (Alarm=True ∧ Burglary=True ∧ Earthquake=False) or equivalently (alarm ∧ burglary ∧ ¬earthquake)
- **Prior probability**: degree of belief without any other evidence
- P(burglary) = 0.1   P(alarm) = 0.19
- **Joint probability**: matrix of combined probabilities of a set of variables
- P(Alarm, Burglary) =

	alarm	¬alarm
burglary	0.09	0.01
¬burglary	0.1	0.8

Probability Theory

- **Conditional probability**: probability of effect given causes
- **Computing conditional probs**:
  – $P(a \mid b) = P(a \wedge b) / P(b)$
  – $1/P(b)$: **normalizing** constant
- **Product rule**:
  – $P(a \wedge b) = P(a \mid b) P(b)$
- **Marginalizing**:
  – $P(B) = \Sigma_a P(B, a)$
  – $P(B) = \Sigma_a P(B \mid a) P(a)$ (**conditioning**)

- P(burglary | alarm) = 0.47
  P(alarm | burglary) = 0.9
- P(burglary | alarm) =
  P(burglary $\wedge$ alarm) / P(alarm)
  = 0.09 / 0.19 = 0.47
- P(burglary $\wedge$ alarm) =
  P(burglary | alarm) P(alarm) =
  0.47 * 0.19 = 0.09
- P(alarm) =
  P(alarm $\wedge$ burglary) +
  P(alarm $\wedge$ ¬burglary) =
  0.09 + 0.1 = 0.19

Probability Theory

- **Law of Total Probability:** Also called *marginalization*:

$$P(a) = \sum_b P(a, b)$$
$$= \sum_b P(a \mid b)\, P(b)$$ where B is any random variable

- Why is it useful?
  - Given a joint distribution we can obtain any "marginal" probability

$$P(b) = \sum_a \sum_c \sum_d P(a, b, c, d)$$

  - We can also compute any "conditional" probability of interest given a joint distribution

$$P(c \mid b) = \sum_a \sum_d P(a, c, d \mid b) = \gamma \sum_a \sum_d P(a, c, d, b)$$

where γ is just a normalization constant

Probability Theory

- **The Chain Rule or Factoring**
  - By the definition of joint probability
    $P(a, b, c, \ldots z) \;=\; P(a \mid b, c, \ldots . z)\, P(b, c, \ldots z)$
  - Repeatedly applying this idea, we can write

Chain Rule

$$P(a, b, c, \ldots z) \;=\; P(a \mid b, c, \ldots . z)\, P(b \mid c, .. z)\, P(c \mid .. z) .. P(z)$$

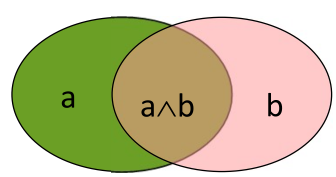
- **Conditional Independence**
  - Two random variables A and B are conditionally independent given C iff
    $P(a, b \mid c) \;=\; P(a \mid c)\, P(b \mid c)$ for all values a, b, c
  - Equivalent conditional formulation: A and B are conditionally independent given C iff $P(a \mid b, c) \;=\; P(a \mid c)$

i.e., ***B* contains no information about *A*, beyond what *C* provides**

Axioms of Probability Theory

- Kolmogorov showed that three simple axioms lead to the rules of probability theory

1. All probabilities are between 0 and 1:
   - $0 \leq P(a) \leq 1$
2. Valid propositions (tautologies) have probability 1, and unsatisfiable propositions have probability 0:
   - P(true) = 1 ; P(false) = 0
3. The probability of a disjunction is given by:
   - $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



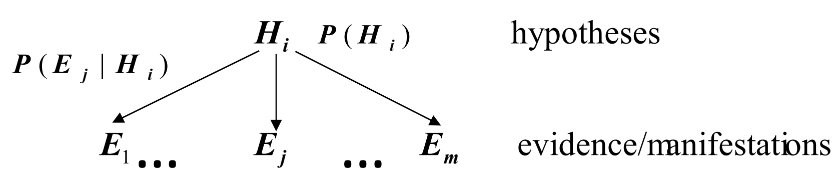
Bayes' rule

- Bayes' rule is derived from the product rule:
  - P(Y | X) = P(X | Y) P(Y) / P(X)

- Often useful for diagnosis:
  - If X are (observed) effects and Y are (hidden) causes,
  - We may have a model for how causes lead to effects (P(X | Y))
  - We may also have prior beliefs (based on experience) about the frequency of occurrence of cause (P(Y))
  - Which allows us to reason abductively from effects to causes (P(Y | X)).

Bayesian inference

- In the setting of diagnostic/evidential reasoning



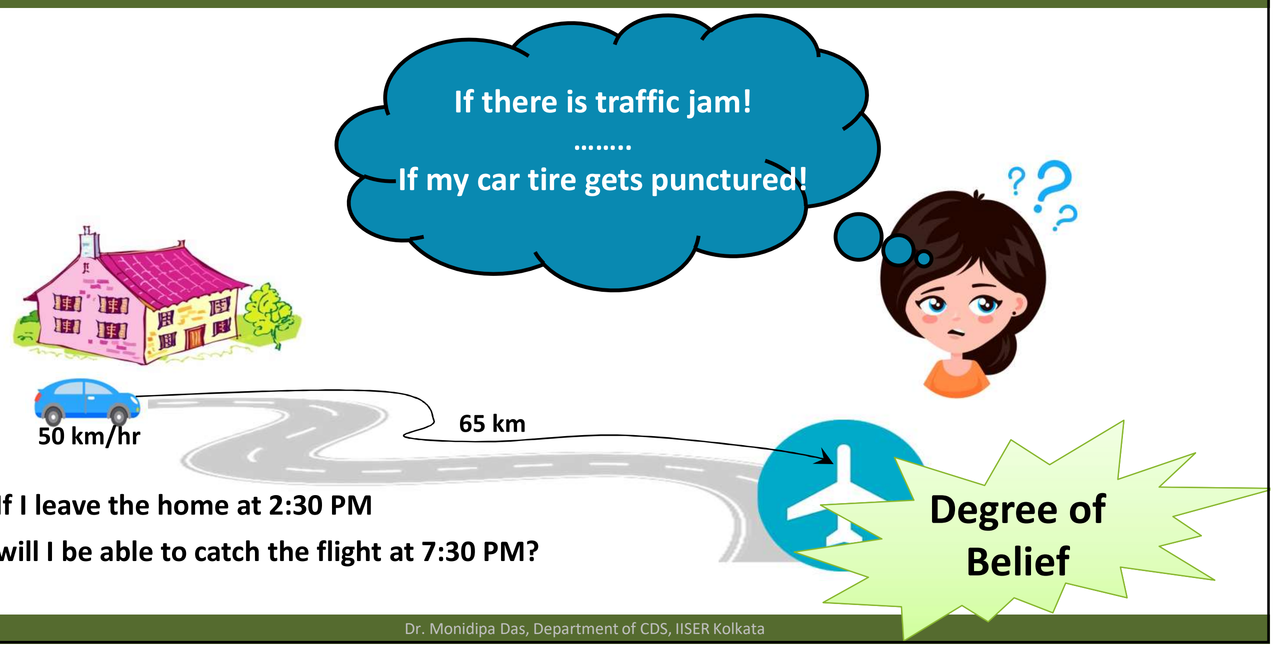
  - Know prior probability of hypothesis — $P(H_i)$
  - conditional probability — $P(E_j \mid H_i)$
  - Want to compute the *posterior probability* — $P(H_i \mid E_j)$

- Bayes' theorem: $P(H_i \mid E_j) = P(H_i) P(E_j \mid H_i) / P(E_j)$

Bayesian Network

- A **graphical model for probabilistic relationships** among a set of variables
- Useful tool in AI for **probabilistic reasoning**
- **Bayes network** (BN), **belief network**
- **Judea Pearl** (UCLA)--- ACM Turing Award 2011

Probabilistic Reasoning



If I leave the home at 2:30 PM
will I be able to catch the flight at 7:30 PM?

Why BN?--- Motivating Example

15

Variables in the study:

1. **Rush-hour (yes/no)**
2. **Bad-weather (yes/no)**
3. **Accident (yes/no)**
4. **Traffic-jam (heavy/light)**
5. **Miss-Flight (yes/no)**

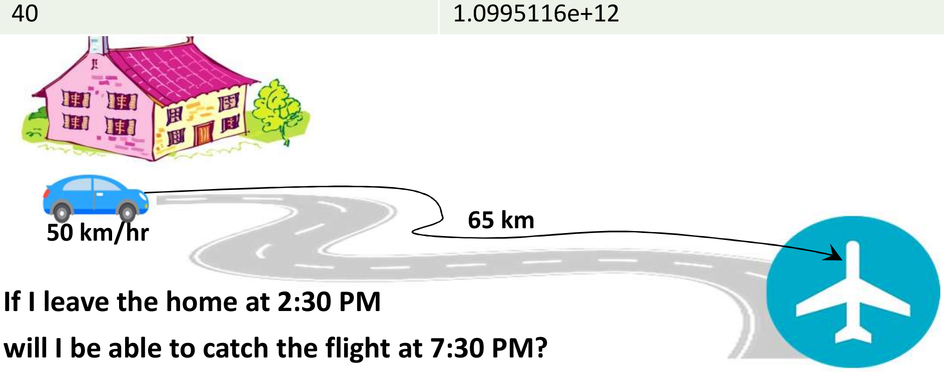
If I leave the home at 2:30 PM

will I be able to catch the flight at 7:30 PM?

Why BN?--- Motivating Example

16

Count of Variables	Table size
5	32
10	1024
20	1,048,576
40	1.0995116e+12



If I leave the home at 2:30 PM

will I be able to catch the flight at 7:30 PM?

Belief Networks are successful examples of probabilistic systems that exploit conditional independence to reason effectively under uncertainty.

Dr. Monidipa Das, Department of CDS, IISER Kolkata

Bayesian Network

- Probabilistic graphical model
- Represents dependence/independence via *directed acyclic graph* (DAG)
- Structure of the graph ⇔ Conditional independence relations

In general,

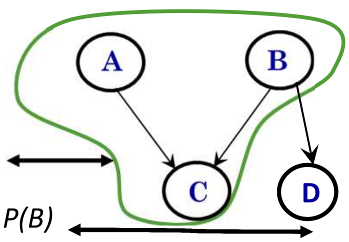
$$P(X1, X2, \ldots .XN) = \prod P(Xi \mid parents(Xi))$$

The full joint distribution *The graph-structured approximation*

- Components of Bayesian network
  - *Qualitative: Graph structure*
  - *Quantitative: Numerical probabilities*

$P(A,B,C) = P(C \mid A,B)P(A)P(B)$

$P(A,B,C,D)=P(D|B).P(C|A,B).P(A).P(B)$



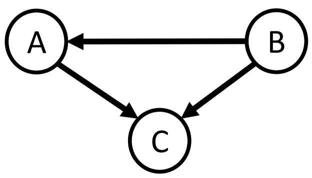
Homework

- **Given the full joint distribution shown below, calculate the following:**
  - **a. P(toothache).**
  - **b. P(Cavity).**
  - **c. P(Cavity | toothache ∨ catch).**

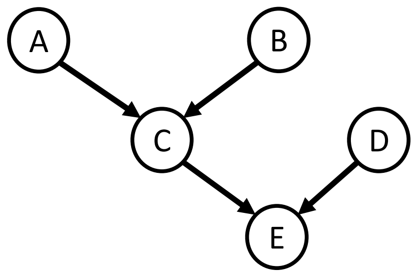
	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Homework

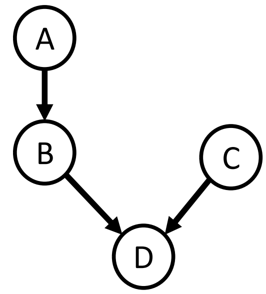
Given the following networks, write the graph structured approximation of the full joint distributions



(a)



(b)



(c)

Questions?

Slide Content taken from
Prof. Cesare Tinelli, Prof. Stuart Russell, and
Prof. Jim Martin