Fundamentals of AI and KR - Module 3

1. Introduction to uncertainty and probabilistic reasoning

Paolo Torroni

Fall 2024

Notice

Credits

The present slides are largely an adaptation of existing material, including:

- slides from Russel & Norvig
- slides by Daphne Koller on Probabilistic Graphical Models
- slides by Fabrizio Riguzzi on Data Mining and Analytics

I am especially grateful to these authors.

Downloading and ~~sharing~~

A copy of these slides can be downloaded from virtuale and stored for personal use only. Please do not redistribute.

Table of Contents

Introduction and logistics

Introduction

Problem-solving logical agents

- Consider 3 switches in a switchboard: A, B, and C. When A is on, B is also on. When C is on, B is off. If both A and C are on, the switchboard melts down. *Can that ever happen?*

Introduction

Problem-solving logical agents

- Consider 3 switches in a switchboard: A, B, and C. When A is on, B is also on. When C is on, B is off. If both A and C are on, the switchboard melts down. *Can that ever happen?*
- There are 12 coins distributed in piles of 4, 1, and 7 coins. We can move coins from a pile A to a pile B, but only by doubling the coins in B. *Can we distribute coins evenly among the three piles?*

Introduction

Problem-solving logical agents

- Consider 3 switches in a switchboard: A, B, and C. When A is on, B is also on. When C is on, B is off. If both A and C are on, the switchboard melts down. *Can that ever happen?*
- There are 12 coins distributed in piles of 4, 1, and 7 coins. We can move coins from a pile A to a pile B, but only by doubling the coins in B. *Can we distribute coins evenly among the three piles?*
- We're guests on a TV game show. We stand in front of three closed doors. A prize hides behind one of them. We choose the door on the left. At this point, the host, who knows where prize is, opens the middle door, to reveal it is empty. We are offered to modify our choice. *Should we?*

Handling uncertainty

Agents may need to handle uncertainty due to...

- partial observability
- nondeterminism
- a combination of both

Problem-solving and logical agents keep a belief state and generate a contingency plan. However...

- large and complex belief-state representations
- arbitrarily large contingency plan
- there may be no plan guaranteed to achieve the goal

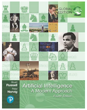
Application areas

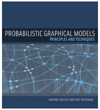
- Robotics
- Medical diagnosis
- Troubleshooting
- Decision-making
- Risk assessment
- Automated monitoring
- Predictions
- Image and speech synthesis/recognition
- Computational biology
- Economics
- ...

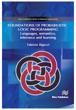
Topics

- Basic probability notation
- Inference using full joint distributions
- Independence
- Bayesian network representation
- Constructing Bayesian networks
- Exact and approximate inference
- Simple case studies

Learning resources







- Course slides
- Textbook
  - Artificial Intelligence. A Modern Approach, by Stuart Russel and Peter Norvig. Pearson Education. 2^{nd}, 3^{rd} Ed. Chapters 13 & 14, or 4^{th} Ed. Chapters 12 & 13.
- Additional reading
  - Probabilistic Graphical Models. Principles and Techniques, by Daphne Koller and Nir Friedman. MIT Press.
  - Foundations of Probabilistic Logic Programming. Languages, Semantics, Inference and Learning, by Fabrizio Riguzzi. River Publishers.
- Software
  - PGM Python library: pgmpy

Exam

- Two alternative possibilities: written exam or project
- Written exam:
  - Questions and/or exercises of the type seen in class
  - 4 dates per academic year
- Mini-project:
  - Implementation of simple case study in pgmpy or other library
  - Upload: ipython notebook and PDF report
  - Oral exam to present work done and answer questions
  - 3 discussion periods per academic year
- Grades for M3 on an 11-point scale (11 is A+)
  - Other modules' max points: 32 (M1) and 21 (M2)
- Final grade is:

$$\frac{\sum_{n=1}^{3} grade(Module\ n)}{2}$$

  - Round up between 18 and 30, but 30L only if $\frac{\sum_i (grade(M_i))}{2} \geq 31.0$

Contacts

- **Email**: paolo.torroni@unibo.it
- Office time this semester on Teams unless otherwise agreed
- Phone: 051 2093767

Acting under uncertainty

Uncertainty

We need to reach the airport on time. Let action A_t = leave for airport t minutes before flight. *Will A_t get me there on time?*

Problems:

1. partial observability (road state, other drivers' plans, etc.)
2. noisy sensors (traffic reports)
3. uncertainty in action outcomes (flat tire, etc.)
4. immense complexity of modelling and predicting traffic

Hence a purely logical approach either

1. risks falsehood: *"A_{25} will get me there on time,"* or
2. leads to conclusions that are too weak for decision making: *"A_{25} will get me there on time if there's no accident on the bridge, and it doesn't rain and my tires remain intact etc etc."*

(A_{1440} might reasonably be said to get me there on time
but I'd have to stay overnight in the airport ...)

Methods for handling uncertainty (M2)

Default or nonmonotonic logic

- Assume my car does not have a flat tire
- Assume A_{25} works unless contradicted by evidence

What assumptions are reasonable?

Rule-based systems with fudge factors

- $A_{25} \mapsto_{0.3} AtAirportOnTime$
- rules for causal reasoning: $FaultyPowerCord \mapsto_{0.99} DisplayOff$
- rules for diagnostic reasoning: $DisplayOff \mapsto_{0.7} SleepMode$

Issues with locality, e.g., how can 0.3 account for "all" the evidence?
Issues with combination, e.g., $FaultyPowerCord$ causes $SleepMode$?

Methods for handling uncertainty

Probability

Given the available evidence, A_{25} will get me there on time with probability 0.04

Remark. Fuzzy logic handles **degree of truth** NOT uncertainty e.g.,

- *TrafficCongested* is true to degree 0.8

Probability

Probabilistic assertions **summarize uncertainty** due to:

- laziness: failure to enumerate exceptions, qualifications, etc.
- ignorance: lack of relevant facts, initial conditions, etc.

Subjective or Bayesian probability:
Probabilities relate propositions to one's own state of knowledge

- e.g., $P(A_{25}|\text{no reported accidents}) = 0.06$

These are **not** claims of a "probabilistic tendency" in the current situation

(but might be learned from past experience of similar situations)

Probabilities of propositions change with new evidence:

- e.g., $P(A_{25}|\text{no reported accidents, 5 a.m.}) = 0.15$

Making decisions under uncertainty

Suppose I believe the following:

$$
\begin{aligned}
P(A_{25} \text{ gets me there on time} | \ldots) &= 0.04 \\
P(A_{90} \text{ gets me there on time} | \ldots) &= 0.70 \\
P(A_{120} \text{ gets me there on time} | \ldots) &= 0.95 \\
P(A_{1440} \text{ gets me there on time} | \ldots) &= 0.9999
\end{aligned}
$$

Which action to choose?

- Depends on my preferences for missing flight vs. airport cuisine, etc.
- Being Rational means following Maximum Expected Utility principle
- Utility theory is used to represent and infer preferences
- Decision theory = utility theory + probability theory

Basic probability notation

Probability basics

Consider the assertions about possible worlds

- Logical assertions say which worlds are ruled out
- Probabilistic assertions say how probable they are

Sample space and events

The set of all possible worlds is called the sample space, denoted Ω. Any subset $A \subseteq \Omega$ is an event. Any element $\omega \in \Omega$ is called a sample point/possible world/atomic event

e.g., 6 possible rolls of a die; die roll < 4; die roll $= 3$

Probability basics

Probability space

A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.

- $0 \leq P(\omega) \leq 1$
- $\sum_\omega P(\omega) = 1$

Accordingly,

$$P(A) = \sum_{\omega \in A} P(\omega)$$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.
$P(\text{die roll} < 4) = P(1) + P(2) + P(3) = 1/6 + 1/6 + 1/6 = 1/2$

Random variables

Random variables

A random variable is **a function** from sample points to some range, e.g., the reals or Booleans.

e.g., $Odd(1) = true$.

Probability distribution

P induces a probability distribution for any r.v. X:

$$P(X = x_i) = \sum_{\omega : X(\omega) = x_i} P(\omega)$$

e.g., $P(Odd = true) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2$

Propositions

Think of a proposition as the event where the proposition is true.
e.g., given Boolean random variables A and B:

- event a = set of sample points where $A(\omega) = true$
- event $\neg a$ = set of sample points where $A(\omega) = false$
- event $a \wedge b$ = points where $A(\omega) = true$ and $B(\omega) = true$

With Boolean variables, sample point = propositional logic model
e.g., $A = true$, $B = false$, or $a \wedge \neg b$.

Proposition = disjunction of atomic events in which it is true
e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$
$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Syntax for propositions

- Propositional or Boolean random variables
  e.g., *Cavity* (do I have a cavity?)
  $Cavity = true$ is a proposition, also written *cavity*
- Discrete random variables (finite or infinite)
  e.g., *Weather* is one of $\langle sunny, rain, cloudy, snow \rangle$
  $Weather = rain$ is a proposition
  Values must be exhaustive and mutually exclusive
- Continuous random variables (bounded or unbounded)
  e.g., $Temp = 21.6$; $Temp < 22.0$
- Arbitrary Boolean combinations of basic propositions

Prior probability

Prior probability

Prior or unconditional probabilities of propositions correspond to belief prior to arrival of any (new) evidence

e.g., $P(Cavity = true) = 0.1$ and $P(Weather = sunny) = 0.72$

Probability distribution

A probability distribution gives values for all possible assignments.

e.g. $\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalized, i.e., sums to 1)

Joint Probability Distribution

Joint Probability Distribution

The Joint Probability Distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)

e.g. $\mathbf{P}(Weather, Cavity)$ = a 4×2 matrix of values:

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Probability for continuous variables

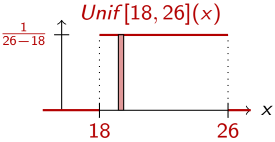
Probability density function

A function $p : \mathbb{R} \to \mathbb{R}$ is a probability density function (pdf) for X if it is a nonnegative integrable function s.t.

$$\int_{Val(X)} p(x)dx = 1.$$

A common pdf is the Uniform distribution

$$p(x) = Unif[a, b](x) = \begin{cases} \frac{1}{b-a} & b \geq x \geq a \\ 0 & \text{otherwise} \end{cases}$$



What $P(X = 20.5) = 0.125$ really means is:

$$\lim_{dx \to 0} \frac{P(20.5 \leq X \leq 20.5 + dx)}{dx} = 0.125$$

Probability for continuous variables

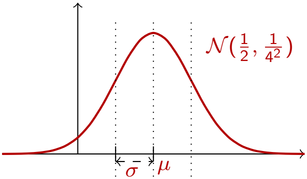
Probability density function

A function $p : \mathbb{R} \to \mathbb{R}$ is a probability density function (pdf) for X if it is a nonnegative integrable function s.t.

$$\int_{Val(X)} p(x)dx = 1.$$

Another common pdf is the Gaussian (Normal) distribution

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



Standard Gaussian: $\mathcal{N}(0, 1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$

Conditional probability

With respect to prior probabilities $P(X)$, conditional or posterior probabilities $P(X|Evidence)$ represent a more informed distribution in the light of the (new) *Evidence*.

e.g., $P(cavity|toothache) = 0.8$:
i.e., **given that** *toothache* **is all I know**
NOT "if *toothache* then 80% chance of *cavity*"

(Notation for sets of conditional distributions:
$\mathbf{P}(Cavity|Toothache)$ is a 2-element vector of 2-element vectors)

Conditional probability

$P(cavity|toothache) = 0.8$

If we know more, e.g., *cavity* is also given, then we have

$P(cavity|toothache, cavity) = 1$

Note: the less specific belief **remains valid** after more evidence arrives, but is not always **useful**

New evidence may be irrelevant, allowing simplification, e.g.,

$P(cavity|toothache, 49ersWin) = P(cavity|toothache) = 0.8$

In this case tootache is indpendent by 49ersWin

This kind of inference, sanctioned by domain knowledge, is crucial

Conditional probability

the event where A e B are true / the event where B is true

Definition of conditional probability: $P(a|b) = \frac{P(a \wedge b)}{P(b)}$ if $P(b) \neq 0$

Product rule gives an alternative formulation:
$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

A general version holds for whole distributions, e.g.,
$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$
(View as a 4×2 set of equations, **not** matrix mult.)

$Weather =$	sunny	rain	cloudy	snow
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

Conditional probability

Chain rule is derived by successive application of product rule:

$$\begin{aligned}
\mathbf{P}(X_1,\ldots,X_n) &= \mathbf{P}(X_1,\ldots,X_{n-1})\,\mathbf{P}(X_n|X_1,\ldots,X_{n-1})\\
&= \mathbf{P}(X_1,\ldots,X_{n-2})\,\mathbf{P}(X_{n-1}|X_1,\ldots,X_{n-2})\,\mathbf{P}(X_n|X_1,\ldots,X_{n-1})\\
&= \ldots\\
&= \textstyle\prod_{i=1}^{n}\mathbf{P}(X_i|X_1,\ldots,X_{i-1})
\end{aligned}$$

P(a,b) = P(b)Pp(a|b)
P(a,b,c) = P(b,c) P(a | b,c)

Inference using full joint distributions

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

For any proposition ϕ, sum the atomic events where it is true:
$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$

e.g. $\phi = toothache$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

For any proposition ϕ, sum the atomic events where it is true:
$P(\phi) = \sum_{\omega:\omega \models \phi} P(\omega)$

e.g. $\phi = toothache$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

For any proposition ϕ, sum the atomic events where it is true:
$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

For any proposition ϕ, sum the atomic events where it is true:
$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$
e.g. $\phi = cavity \vee toothache$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

For any proposition ϕ, sum the atomic events where it is true:
$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$
$P(cavity \lor toothache) = 0.108+0.012+0.072+0.008+0.016+0.064 = 0.28$

Inference by enumeration

Start with the joint distribution:

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Can also compute conditional probabilities:

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \land toothache)}{P(toothache)}$$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

Can also compute conditional probabilities:
probability of not cavity given toothache

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Normalization

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

Denominator can be viewed as a normalization constant α

$$\mathbf{P}(Cavity|toothache) = \alpha\,\mathbf{P}(Cavity, toothache)$$

a = 1/ P(toothache)

We do this normalization at the end

Normalization

Remember than in this case toothache is given so the red part we don't even watch it, it is not possibile that we are in that situaion

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

Denominator can be viewed as a normalization constant α

$$\mathbf{P}(Cavity|toothache) = \alpha\,\mathbf{P}(Cavity, toothache)$$
$$= \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$$
$$= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$
$$= \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$$

Normalization

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

Denominator can be viewed as a normalization constant α

$$\mathbf{P}(Cavity|toothache) = \alpha\,\mathbf{P}(Cavity, toothache)$$

$$\begin{aligned} &= \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\ &= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle] \\ &= \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle \end{aligned}$$

General idea: compute distribution on query variable by fixing evidence variables and summing over hidden variables

Common terminology for operations on CPDs

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

- Marginalization or Summing Out
  e.g., $P(Weather = sunny)$

Common terminology for operations on CPDs

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

- Marginalization or Summing Out
  e.g., $P(Weather = sunny)$
- Conditioning
  e.g., condition on $Weather = sunny$: $P(Cavity|Weather = sunny)$

  $\Rightarrow$ reduction and renormalization

Probability queries

Y is a query, e is an evidence

Probability query

A probability query $\mathbf{P}(\mathbf{Y}|\mathbf{e})$ defines the posterior joint distribution of a set of query variables $\mathbf{Y}$ given specific values $\mathbf{e}$ for some evidence variables.

We thus have three sets of r.v.s: query variables $\mathbf{Y}$, evidence variables $\mathbf{E}$, and hidden variables $\mathbf{H}$ (all else).

In principle, one could answer the query by summing out.

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha\mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}) = \ldots$$

Probability queries

Probability query

A probability query $\mathbf{P}(\mathbf{Y}|\mathbf{e})$ defines the posterior joint distribution of a set of query variables $\mathbf{Y}$ given specific values $\mathbf{e}$ for some evidence variables.

In principle, one could answer the query by summing out.

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e}) = \alpha\sum_{\mathbf{h}}\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e},\mathbf{H}=\mathbf{h})$$

Probability of a query given Evidence is a constant

Obvious problems:

1. Worst-case time complexity
2. Space complexity
3.

Probability queries

Probability query

A probability query $\mathbf{P}(\mathbf{Y}|\mathbf{e})$ defines the posterior joint distribution of a set of query variables $\mathbf{Y}$ given specific values $\mathbf{e}$ for some evidence variables.

In principle, one could answer the query by summing out.

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}) = \alpha \sum_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}, \mathbf{H}=\mathbf{h})$$

Obvious problems:

1. Worst-case time complexity $O(d^n)$ where d is the largest arity
2. Space complexity $O(d^n)$ to store the joint distribution
3. How to find the numbers for $O(d^n)$ entries???

Questions?