Bayes & Uncertainty I

PROF LIM KWAN HUI

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

50.021 Artificial Intelligence

Outline & Objectives

- Recap on statistical concepts such as product rule, chain rule, conditional independence, Bayes rules
- Able to represent a problem in terms of a Bayesian network and its corresponding conditional probability table
- Learn about how Bayes net can be used in various scenarios
- Learn about the Naïve Bayes Classifier and its application to text

Uncertainty

- 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 (radio traffic reports)
  - 3) uncertainty in action outcomes (flat tyre, 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)

Probability

- Probabilistic assertions summarize effects of
  - 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., $P(A_{25} \mid \text{no reported accidents}) = 0.06$
- These are not claims of some 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., $P(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15$

Prior probability

- Prior or unconditional probabilities of propositions
  - e.g., P(Cavity =true) = 0.1 and P(Weather =sunny) = 0.72
  - correspond to belief prior to arrival of any (new) evidence
- Probability distribution gives values for all possible assignments:
  - P(Weather) = <0.72, 0.1, 0.08, 0.1> (normalized, i.e., sums to 1)
- 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)
  - P(Weather,Cavity) = a 4 x 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

Conditional probability

- Conditional or posterior probabilities
  - e.g., P(cavity | toothache) = 0.8
  - i.e., given that toothache is all I know
- Notation for conditional distributions:
  - P(Cavity | Toothache) = 2-element vector of 2-element vectors
- If we know more, e.g., cavity is also given, then we have
  - P(cavity | toothache, cavity) = 1
- New evidence may be irrelevant, allowing simplification, e.g.,
  - P(cavity | toothache, chelseaWins) = P(cavity | toothache) = 0.8
- This kind of inference, sanctioned by domain knowledge, is crucial

Conditional probability

- Definition of conditional probability:

$$P(A \mid B) = \frac{P(A \land B)}{P(B)}$$

- Product rule gives an alternative formulation:

$$\begin{aligned} P(A \land B) &= P(A \mid B)\, P(B) \\ &= P(B \mid A)\, P(A) \end{aligned}$$

- A general version holds for whole distributions, e.g.,
  - **P**(*Weather,Cavity*) = **P**(*Weather* | *Cavity*) **P**(*Cavity*)
  - (View as a set of 4 × 2 equations, not matrix multiplication)

Conditional probability

- Definition of conditional probability:

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

- Product rule gives an alternative formulation: $P(A \wedge B) = P(A \mid B)\, P(B)$
  $= P(B \mid A)\, P(A)$

- A general version holds for whole distributions, e.g.,
  - **P**(*Weather,Cavity*) = **P**(*Weather | Cavity*) **P**(*Cavity*)
  - (View as a set of 4 × 2 equations, not matrix multiplication)

- Chain rule is derived by successive application of product rule:

- $\mathbf{P}(X_1, \ldots, X_n)$

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

Inference by enumeration

- Start with the joint distribution:

	toothache		¬toothache	
	catch	**¬catch**	**catch**	**¬catch**
cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

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

Inference by enumeration

- Start with the joint distribution:

	toothache		¬toothache	
	catch	**¬catch**	**catch**	**¬catch**
cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

- For any proposition , sum the atomic events where it is true:
  - P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

Inference by enumeration

- Start with the joint distribution:

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

- For any proposition , sum the atomic events where it is true:
  - P(cavity ∨ toothache) = 0.108+0.012+0.072+0.008+0.016+0.064 = 0.28

Inference by enumeration

- Start with the joint distribution:

	toothache		¬toothache	
	catch	**¬catch**	**catch**	**¬catch**
cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

- Can also compute conditional probabilities:
  - $$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \land toothache)}{P(toothache)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Exercise: Inference by enum.

- Start with the joint distribution:

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

- Compute the probabilities for the following:
  - P(*Cavity*)
  - P(*Toothache* | *cavity*)

Exercise: Inference by enum.

- Start with the joint distribution:

	toothache		¬toothache	
	catch	**¬catch**	**catch**	**¬catch**
cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

- Compute the probabilities for the following:
- P(*Cavity*) = <0.108 + 0.012 + 0.072 + 0.008, 0.016 + 0.064 + 0.144 + 0.576>

  = < 0.2, 0.8 >

Exercise: Inference by enum.

- Start with the joint distribution:

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

- Compute the probabilities for the following:
- **P**(Toothache | cavity) = <*P*(toothache,cavity) / *P*(cavity), *P*(!toothache,cavity) / *P*(cavity)>

  = < (0.108+0.012)/0.2, (0.072 + 0.008)/0.2 >

  = < 0.12/0.2, 0.08/0.2 >

  = < 0.6, 0.4 >

Normalization

- Denominator can be viewed as a normalization constant α
  - $P(Cavity \mid toothache) = \alpha\, P(Cavity, toothache)$

$$= \alpha\, [P(Cavity, toothache, catch) + P(Cavity, toothache, \neg catch)]$$

$$= \alpha\, [\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$

$$= \alpha\, \langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$$

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

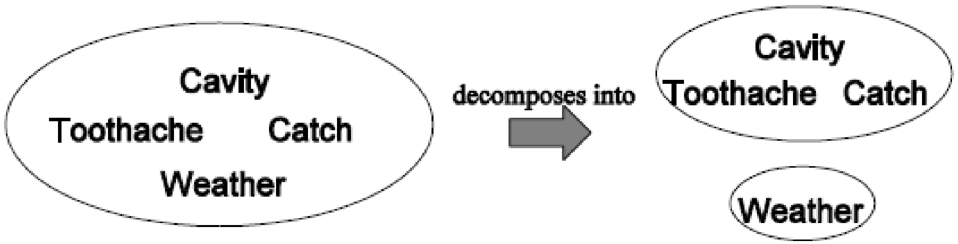
	toothache		**¬toothache**	
	catch	**¬catch**	**catch**	**¬catch**
cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

Inference by enumeration

- Typically, we are interested in
  - the posterior joint distribution of the query variables Y
  - given specific values e for the evidence variables E
- Let the hidden variables be H = X – Y – E
- Then the required summation of joint entries is done by summing out the hidden variables:
  - $P(Y \mid E=e) = \alpha\, P(Y, E=e) = \alpha \sum_h P(Y, E=e, H=h)$
- The terms in the summation are joint entries because Y, E, and H together exhaust the set of random variables
- 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

Independence

- A and B are independent iff
  - P(A|B)=P(A) or
  - P(B|A)=P(B) or
  - P(A,B)=P(A)P(B)



- P(Toothache,Catch,Cavity,Weather)

  = P(Toothache,Catch,Cavity)P(Weather)

- 32 entries reduced to 12
- Absolute independence powerful but rare
- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

Conditional independence

- P(Toothache,Cavity,Catch) has 2^3 independent entries
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - $P(catch | toothache, cavity) = P(catch | cavity)$
- The same independence holds if I haven't got a cavity:
  - $P(catch | toothache, \neg cavity) = P(catch | \neg cavity)$
- Catch is conditionally independent of Toothache given Cavity:
  - $P(Catch | Toothache, Cavity) = P(Catch | Cavity)$
- Equivalent statements:
  - $P(Toothache | Catch, Cavity) = P(Toothache | Cavity)$
  - $P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)$

Conditional independence

- Write out full joint distribution using chain rule:
  - P(Toothache,Catch,Cavity)

    = P(Toothache|Catch,Cavity)P(Catch,Cavity)

    = P(Toothache|Catch,Cavity)P(Catch|Cavity)P(Cavity)

    = P(Toothache|Cavity)P(Catch|Cavity)P(Cavity)

- I.e., 2 + 2 + 1 = 5 independent numbers

- In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in n to linear in n.

- Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Bayes Rule

- Product rule: $P(a \wedge b) = P(a|b)\, P(b) = P(b|a)\, P(a)$

  $\Rightarrow$ Bayes rule: $P(A \mid B) = \dfrac{P(B \mid A)\, P(A)}{P(B)}$

- Or in distribution form

  $$P(Y \mid X) = \frac{P(X \mid Y)\, P(Y)}{P(X)} = \alpha\, P(X \mid Y)\, P(Y)$$

Next

- Learn how to represent a problem in terms of a Bayesian network and its corresponding conditional probability table