Artificial Intelligence

Week 11: Reasoning under Uncertainty

COMP30024

May 17, 2021

Bayes' Theorem

- Conditioning on the known value of data x yields Bayes' theorem:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

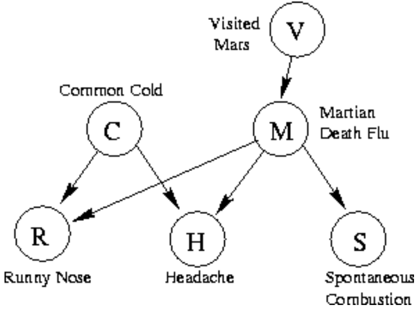
  - The likelihood $p(y|\theta)$ is the conditional probability of the data y given fixed θ.
  - The prior $p(\theta)$ represents information we have that is not part of the collected data y.
  - The evidence $p(y)$ is the average over all possible values of θ.

$$p(y) = \sum_{\theta} p(y|\theta)p(\theta)$$

- $p(\theta|y)$ is the posterior distribution, which represents our updated beliefs under our prior $p(\theta)$ now we have observed the data y.

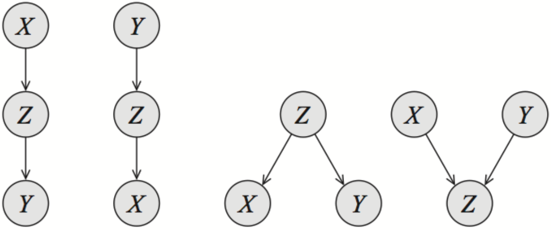
Bayesian Networks

- Graphical way of encoding conditional independence assumptions.



Bayesian Networks

- Vertices correspond to random variables.
- Edges, e.g. $X \to Y$ indicates X has a direct influence on Y. Causes usually parents of effects.
- Attach local conditional probability at each vertex explaining effects of parents: $P(X_i|\text{Parents}(X_i))$.



(a) (b) (c) (d)

Bayesian Networks

- Chain rule allows decomposition of joint into conditionals:

$$P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)\ldots P(X_n|X_{n-1}, \ldots, X_2, X_1)$$

- Conditional independence assumptions: each X_i only directly depends on a small number of variables: Parents(X_i).

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\text{Parents}(X_i))$$

- If each variable has d possible values and $\leq k$ parents, joint distribution has $\mathcal{O}(nd^k)$ entries that need to be learnt (versus $\mathcal{O}(d^n)$).
  - e.g. 30 random variables, 5 parents each, factorization uses ≈ 1000 random variables versus over 10^9 for the full joint.

Bayesian Inference

- 'Find probability of X, given we know e has occurred.' $\rightarrow P(X|e)$.

Bayesian Inference

- 'Find probability of X, given we know e has occurred.' $\rightarrow P(X|e)$.
- Let H denote all variables outside X, e (H for hidden). Bayes' Theorem says:

$$\begin{aligned} P(X|e) &= \frac{P(X,e)}{P(e)} \\ &= \frac{1}{P(e)} \sum_H P(X,e|H)p(H) \\ &= \frac{1}{P(e)} \sum_H P(X,e,H) \end{aligned}$$

Bayesian Inference

- 'Find probability of X, given we know e has occurred.' $\rightarrow P(X|e)$.
- Let H denote all variables outside X, e (H for hidden). Bayes' Theorem says:

$$\begin{aligned} P(X|e) &= \frac{P(X,e)}{P(e)} \\ &= \frac{1}{P(e)} \sum_H P(X,e|H)p(H) \\ &= \frac{1}{P(e)} \sum_H P(X,e,H) \end{aligned}$$

- To solve a Bayesian inference problem:

Bayesian Inference

- 'Find probability of X, given we know e has occurred.' $\rightarrow P(X|e)$.
- Let H denote all variables outside X, e (H for hidden). Bayes' Theorem says:

$$\begin{aligned} P(X|e) &= \frac{P(X,e)}{P(e)} \\ &= \frac{1}{P(e)} \sum_H P(X,e|H)p(H) \\ &= \frac{1}{P(e)} \sum_H P(X,e,H) \end{aligned}$$

- To solve a Bayesian inference problem:
    - Identify query (X), evidence (e) and hidden (H) variables.

Bayesian Inference

- 'Find probability of X, given we know e has occurred.' $\rightarrow P(X|e)$.
- Let H denote all variables outside X, e (H for hidden). Bayes' Theorem says:

$$
\begin{aligned}
P(X|e) &= \frac{P(X,e)}{P(e)} \\
&= \frac{1}{P(e)} \sum_{H} P(X,e|H)p(H) \\
&= \frac{1}{P(e)} \sum_{H} P(X,e,H)
\end{aligned}
$$

- To solve a Bayesian inference problem:
  - Identify query (X), evidence (e) and hidden (H) variables.
  - Use Bayesian network structure to dismantle joint into product of simpler conditional distributions.

Bayesian Inference

- 'Find probability of X, given we know e has occurred.' $\rightarrow P(X|e)$.
- Let H denote all variables outside X, e (H for hidden). Bayes' Theorem says:

$$\begin{aligned} P(X|e) &= \frac{P(X,e)}{P(e)} \\ &= \frac{1}{P(e)} \sum_H P(X,e|H)p(H) \\ &= \frac{1}{P(e)} \sum_H P(X,e,H) \end{aligned}$$

- To solve a Bayesian inference problem:
  - Identify query (X), evidence (e) and hidden (H) variables.
  - Use Bayesian network structure to dismantle joint into product of simpler conditional distributions.
  - Fix evidence variables to observed values e.

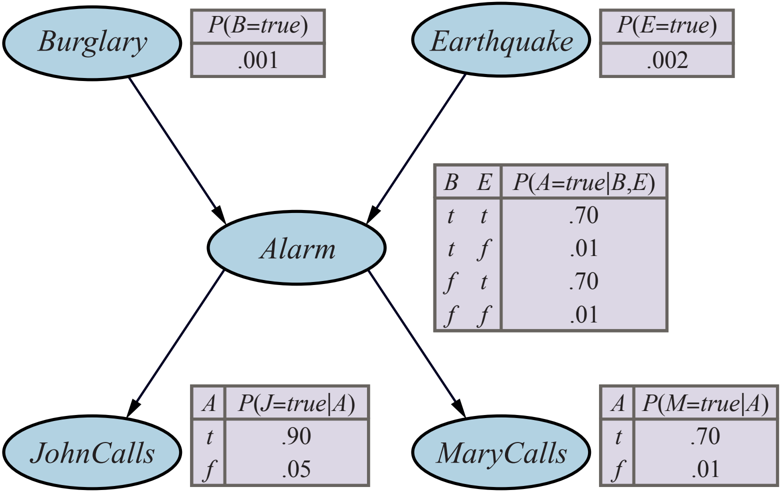
Bayesian Inference

- 'Find probability of X, given we know e has occurred.' $\rightarrow P(X|e)$.
- Let H denote all variables outside X, e (H for hidden). Bayes' Theorem says:

$$
\begin{aligned}
P(X|e) &= \frac{P(X,e)}{P(e)} \\
&= \frac{1}{P(e)} \sum_H P(X,e|H)p(H) \\
&= \frac{1}{P(e)} \sum_H P(X,e,H)
\end{aligned}
$$

- To solve a Bayesian inference problem:
  - Identify query (X), evidence (e) and hidden (H) variables.
  - Use Bayesian network structure to dismantle joint into product of simpler conditional distributions.
  - Fix evidence variables to observed values e.
  - Sum (marginalize) joint over remaining hidden variables H.

Bayesian Inference



$P(B{=}true)$
.001

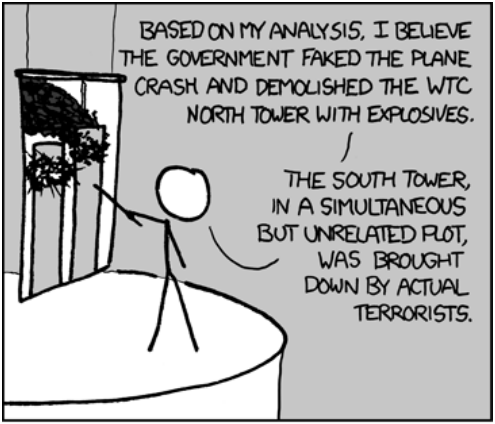
$P(E{=}true)$
.002

B	E	$P(A{=}true \mid B,E)$
t	t	.70
t	f	.01
f	t	.70
f	f	.01

A	$P(J{=}true \mid A)$
t	.90
f	.05

A	$P(M{=}true \mid A)$
t	.70
f	.01

V-structure / Explaining Away



THE 9/11 TRUTHERS RESPONDED POORLY TO MY COMPROMISE THEORY.

$$G \rightarrow 9/11 \leftarrow T$$

V-structure / Explaining Away

- Consider the graph $B \rightarrow A \leftarrow E$ - are B and E independent?
  - Without A: B gives us no information about E and vice-versa $\rightarrow B \perp E|C$ if A unknown.
  - With A: couples parents $\rightarrow B \not\perp E|C$ if A known.
- Parents conditionally independent if child is unobserved, but dependent when child is observed.

$$R \rightarrow W \leftarrow S$$

- Suppose your lawn is wet in the morning ($W =$ True). R (rain) and S (sprinkler) are the only causes of wetness.
- If W true, R is false, S must be true. Observing child and one parent gives you information about the other parent.