Announcements

- Midterm 10/21!
- One page note sheet
- Review sessions Friday and Sunday (similar)
- OHs on various topics TBD on web page
- See webpage for details and practice exams
- Warning: watch web page for reporting instructions – may have different room(s)
- Section this week but not next
- Next reading needs login/password

Probabilistic Models

- A probabilistic model is a joint distribution over a set of variables
  \[ P(X_1, X_2, \ldots, X_n) \]
- Given a joint distribution, we can reason about unobserved variables given observations (evidence)
  \[ P(X_i | x_1, \ldots, x_k) \]

This kind of posterior distribution is also called the belief function of an agent which uses this model.

Independence

- Two variables are independent if:
  \[ \forall x, y : P(x, y) = P(x)P(y) \]

  - This says that their joint distribution factors into a product of two simpler distributions
  - Another form:
    \[ \forall x, : P(x|y) = P(x) \]
  - We write: \( X \perp Y \)

  Independence is a simplifying modeling assumption
  - Empirical joint distributions: at best “close” to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?

Example: Independence

- N fair, independent coin flips:

  \[
  \begin{array}{c|c|c}
  X_1 & T & 0.5 \\
  H & 0.5 \\
  \end{array}
  \]

  \[
  \begin{array}{c|c|c}
  X_2 & T & 0.5 \\
  H & 0.5 \\
  \end{array}
  \]

  \[
  \ldots
  \]

  \[
  \begin{array}{c|c|c}
  X_n & T & 0.5 \\
  H & 0.5 \\
  \end{array}
  \]

  \[
  \begin{array}{c|c|c|c}
  & T & P & 0.5 \\
  & H & 0.5 \\
  & \cdots & \cdots & \cdots \\
  & \cdots & \cdots & \cdots \\
  \end{array}
  \]

  \[
  \begin{array}{c|c|c|c}
  & T & P & 0.5 \\
  & H & 0.5 \\
  & \cdots & \cdots & \cdots \\
  & \cdots & \cdots & \cdots \\
  \end{array}
  \]

Example: Independence?

- \( P(T) \)

  \[
  \begin{array}{c|c|c}
  & T & P & 0.5 \\
  & H & 0.5 \\
  & \cdots & \cdots & \cdots \\
  & \cdots & \cdots & \cdots \\
  \end{array}
  \]

  \[
  \begin{array}{c|c|c|c}
  & T & W & P & 0.5 \\
  & \text{warm} & \text{sun} & 0.4 \\
  & \text{warm} & \text{rain} & 0.1 \\
  & \text{cold} & \text{sun} & 0.2 \\
  & \text{cold} & \text{rain} & 0.3 \\
  \end{array}
  \]

- \( P(W) \)

  \[
  \begin{array}{c|c|c|c}
  & W & P & 0.5 \\
  & \text{sun} & 0.6 \\
  & \text{rain} & 0.4 \\
  \end{array}
  \]
Conditional Independence

- P(Toothache, Cavity, Catch)
- If I have a cavity, the probability that the probe catches in it doesn’t depend on whether I have a toothache:
  \[ P(\text{catch} | \text{toothache}, \text{cavity}) = P(\text{catch} | \text{cavity}) \]
- The same independence holds if I don’t have a cavity:
  \[ P(\text{catch} | \text{toothache}, \neg \text{cavity}) = P(\text{catch} | \neg \text{cavity}) \]
- Catch is conditionally independent of Toothache given Cavity:
  \[ P(\text{Catch} | \text{Toothache}, \text{Cavity}) = P(\text{Catch} | \text{Cavity}) \]
- Equivalent statements:
  - \[ P(\text{Toothache} | \text{Catch}, \text{Cavity}) = P(\text{Toothache} | \text{Cavity}) \]
  - \[ P(\text{Toothache}, \text{Catch} | \text{Cavity}) = P(\text{Toothache} | \text{Cavity})P(\text{Catch} | \text{Cavity}) \]

Unconditional (absolute) independence is very rare (why?)

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

What about this domain:
- Traffic
- Umbrella
- Raining
- What about fire, smoke, alarm?

The Chain Rule

- Can always factor any joint distribution as an incremental product of conditional distributions:
  \[ P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \ldots \]
  \[ P(X_1, X_2, \ldots, X_n) = \prod_{i} P(X_i|X_1, \ldots, X_{i-1}) \]
- Why?

  - This actually claims nothing…
  - What are the sizes of the tables we supply?

Why?

Why?

Simple, local distributions

Bayes’ nets / graphical models help manage independence

Models are always simplifications
- May not account for every variable
- May not account for all interactions between variables

Probabilistic Models

Models are descriptions of how (a portion of) the world works

Models are always simplifications
- May not account for every variable
- May not account for all interactions between variables

What do we do with probabilistic models?
- We (or our agents) need to reason about unknown variables, given evidence
- Example: explanation (diagnostic reasoning)
- Example: prediction (causal reasoning)
- Example: value of information

Bayes’ Nets: Big Picture

- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
- For about 10 min, we’ll be vague about how these interactions are specified
Example Bayes’ Net

Graphical Model Notation

- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
  - Similar to CSP constraints
  - Indicate “direct influence” between variables
  - For now: imagine that arrows mean direct causation

Example: Coin Flips

- N independent coin flips
- $X_1, X_2, \ldots, X_n$
- No interactions between variables: absolute independence

Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic
- Model 1: independence
- Model 2: rain causes traffic
- Why is an agent using model 2 better?

Example: Traffic II

- Let’s build a causal graphical model
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Example: Alarm Network

- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!
Bayes’ Net Semantics

- Let’s formalize the semantics of a Bayes’ net
- A set of nodes, one per variable $X$
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over $X$, one for each combination of parents’ values
  - $P(X|A_1 \ldots A_n)$
- CPT: conditional probability table
- Description of a noisy “causal” process

$A$ Bayes net = Topology (graph) + Local Conditional Probabilities

Probabilities in BNs

- Bayes’ nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
    
    $P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(x_i))$

- Example:
  
  $P(\text{cavity}, \text{catch}, \neg\text{toothache})$

- This lets us reconstruct any entry of the full joint
- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies

Example: Coin Flips

- $X_1$, $X_2$, ..., $X_n$
- $P(X_1)$, $P(X_2)$, ..., $P(X_n)$
- $P(h, h, t, h) =$

Only distributions whose variables are absolutely independent can be represented by a Bayes’ net with no arcs.

Example: Traffic

- $P(R)$, $P(r, \neg t) =$
- $P(T|R)$

Example: Alarm Network

- $P(B|\neg E)$, $P(E)$
- $P(\text{Alarm})$

- $P(\text{JohnCalls}, \neg A|T)$, $P(\text{MaryCalls}, \neg A|T)$

- $P(h, c, \neg a, j, m) =$

Example: Traffic II

- Variables
  - $T$: Traffic
  - $R$: It rains
  - $L$: Low pressure
  - $D$: Roof drips
  - $B$: Ballgame
Size of a Bayes’ Net

- How big is a joint distribution over N Boolean variables?
- How big is an N-node net if nodes have k parents?
- Both give you the power to calculate \( P(X_1, X_2, \ldots, X_n) \)
- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also turns out to be faster to answer queries (coming)

Building the (Entire) Joint

- We can take a Bayes’ net and build the full joint distribution it encodes
  \[
  P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(x_i))
  \]
- Typically, there’s no reason to build ALL of it
- But it’s important to know you could!
- To emphasize: every BN over a domain implicitly represents some joint distribution over that domain

Example: Traffic

- Basic traffic net
- Let’s multiply out the joint

Example: Reverse Traffic

- Reverse causality?

Causality?

- When Bayes’ nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g., consider the variables Traffic and Drips
  - End up with arrows that reflect correlation, not causation
- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology really encodes conditional independence

Bayes’ Nets

- So far: how a Bayes’ net encodes a joint distribution
- Next: how to answer queries about that distribution
  - Key idea: conditional independence
  - Last class: assembled BNs using an intuitive notion of conditional independence as causality
  - Today: formalize these ideas
  - Main goal: answer queries about conditional independence and influence
- After that: how to answer numerical queries (inference)