COMPSCI 688: Probabilistic Graphical Models

Lecture 4: Directed Graphical Models: D-Separation, Queries

- Quiz 1, due Fri 11:59pm

– HW 1, Canvas, due 2 weeks on Wed

Dan Sheldon

Manning College of Information and Computer Sciences
University of Massachusetts Amherst

Partially based on materials by Benjamin M. Marlin (marlin@cs.umass.edu) and Justin Domke (domke@cs.umass.edu)

Review

Review

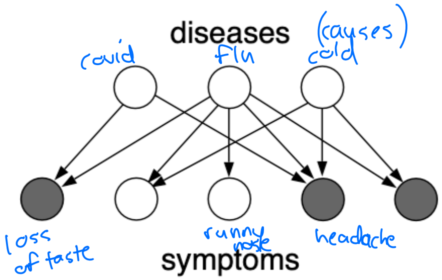
- Bayes net: $p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid \mathbf{x}_{\mathsf{pa}(i)})$
- Factorization $\iff$ conditional independence (one statement per node)

$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid \mathbf{x}_{\mathsf{pa}(i)}) \quad \iff \quad X_i \perp \mathbf{X}_{\mathsf{nd}(i)} \mid \mathbf{X}_{\mathsf{pa}(i)} \text{ for all } i$$

- We would like to chain together conditional independence properties using the graph structure to derive new ones → D-separation

Examples

Example 1: Medical Diagnosis

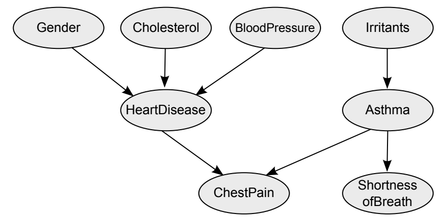


p(covid | loss, headache, ...)

observable (effects)

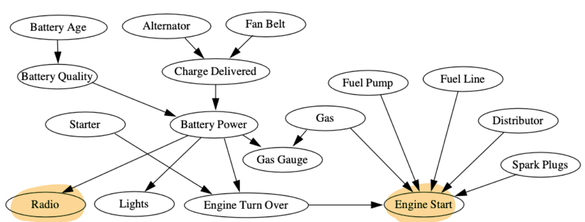
shading = conditioning "observed"

Example 2: Medical Diagnosis

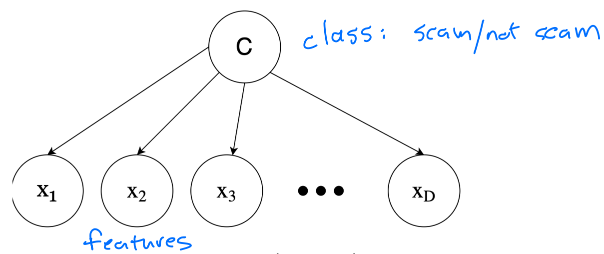


Example 3: Equipment Diagnostics

Suppose engine does not start but radio plays? What do you believe about gas in tank?



Example 4: Naive Bayes



- given class variable C (discrete), the observed features $X_1, X_2, X_3, \dots X_D$ are conditionally independent: $\forall i \neq j, X_i \perp X_j | C$
- joint distribution is

$$p(c, x) = p(c) \prod_{i=1}^{D} p(x_i | c).$$

Warning: Causality

Bayes nets are not causal

- Many of our examples are motivated by a causal model, but Bayes net arrows could just as easily point from "effect" to "cause" (or have no causal semantics at all)
- Can be given causal semantics (beyond our scope)

breakfast hungry argue

$A \rightarrow B \rightarrow C \quad \Rightarrow \quad A \leftarrow B \leftarrow C$

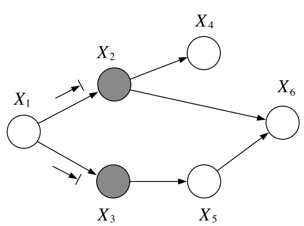
$p(a,b,c) = p(a)p(b|a)p(c|b) \qquad p(a,b,c) = p(c)p(b|c)p(a|b)$

D-Separation

Independence Properties

$X_a \perp X_b \mid X_c \Rightarrow X_b \perp X_a \mid X_c$

So far, we know $X_i \perp X_{\text{nd}(i)} | X_{\text{pa}(i)}$ for all i



$X_2 \perp X_4, X_5 \mid X_1$

$X_6 \perp X_1, X_2, X_4, X_3, X_5 \mid X_2, X_5$

$X_1 \perp X_6 \mid X_2, X_3$? yes, but not known yet

However, this also implies other conditional independence properties. E.g., it's true that $X_1 \perp X_6 | X_2, X_3$ in this network. How can we determine this?

The core principles can be understood by examining three-node networks, then "chaining" ideas together...

Three-Node Bayes Nets: Common Parent, Chains

$A - B - C$

Networks $A \leftarrow B \rightarrow C$, $A \rightarrow B \rightarrow C$ and $C \rightarrow B \rightarrow A$

"common cause" "causal chain"

$A \leftarrow B \rightarrow C \qquad A \rightarrow B \rightarrow C \qquad A \leftarrow B \leftarrow C$

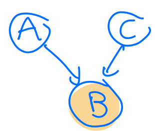
$A \perp C \mid B \qquad A \perp C \mid B \qquad A \perp C \mid B$

$A \not\perp C \qquad A \not\perp C \qquad A \not\perp C$

$A \not\perp C$ but $A \perp C | B$. Observing B *blocks* dependence of A and C

Three-Node Bayes Nets: V-Structure

Network $A \to B \leftarrow C$



causes

"common effect"

$A \perp C$

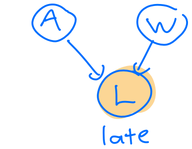
$A \not\perp C \mid B$

$A \perp C$ but $A \not\perp C|B$. Observing B *induces* dependence of A and C

Explaining Away

Judea Pearl Turing Award 2012

- "Explaining away" via V-structures is a distinguishing property of Bayes nets:
- **Example**: Alice is late. Was she abducted by aliens or did she forget her watch?



$P(A=yes) = 0.00001$

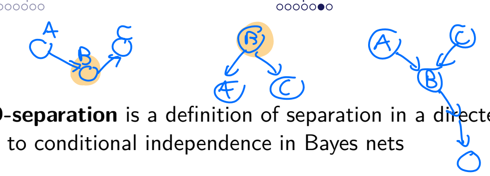
$P(A=yes \mid L=yes) = 0.00002$

$P(A=yes \mid L=yes, W=yes) = 0.000011$

$L = A$ or W

In words: if there are two possible causes for the observed evidence, knowing about one of the causes provides information about the other

D-Separation



Directed separation or **D-separation** is a definition of separation in a directed graph that corresponds exactly to conditional independence in Bayes nets

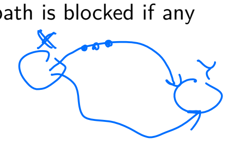
A three-node path is blocked iff has one of the following types:

1) $A \to B \to C$ or $C \to B \to A$ and B is observed
2) $A \leftarrow B \to C$ and B is observed
3) $A \to B \leftarrow C$ and neither B nor any descendent of B is observed

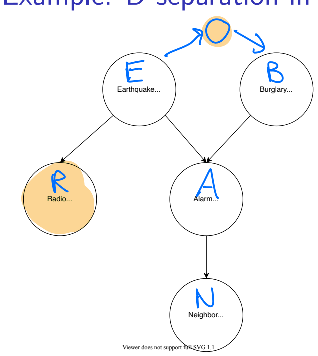
Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be three sets of nodes. $\mathbf{X}$ and $\mathbf{Y}$ are d-separated given observed nodes $\mathbf{Z}$ iff every path from $\mathbf{X}$ to $\mathbf{Y}$ is blocked, where a path is blocked if any three-node sequence in the path is blocked.

claim:

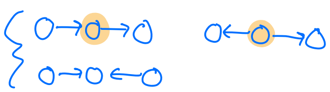
$\mathbf{X}$ and $\mathbf{Y}$ are d-separated given $\mathbf{Z}$ $\iff$ $\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}$



Example: D-separation in the Alarm Model



blocked

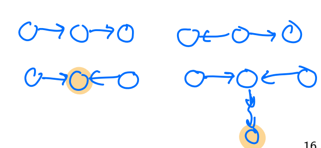


E ⊥ B|A, R? no

E ⊥ B|R, N? no

E ⊥ B|R? yes

open

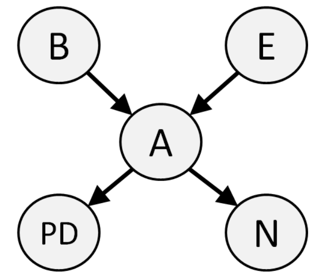


Queries

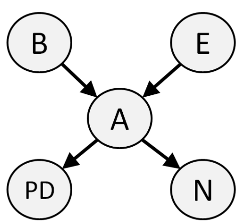
"inference"

The Alarm Network (II)

- You live in the suburbs of LA. Your home alarm may go off because of a break-in or earthquake. If your alarm goes off you might get a call from the police or your neighbor.
- **Random Variables:** Break-in (B), Earthquake (E), Alarm (A), Police Department calls (PD), Neighbor calls (N).



The Alarm Network: Factorization

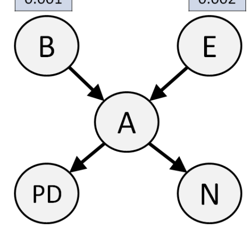


- **Factorization:** $P(B, E, A, PD, N) = P(B)P(E)P(A|B, E)P(PD|A)P(N|A)$

The Alarm Network: Parameters

P(B=1)
0.001

P(E=1)
0.002



B	E	P(A=1|B,E)
1	1	0.950
1	0	0.940
0	1	0.290
0	0	0.001

A	P(PD=1|A)
1	0.900
0	0.005

A	P(N=1|A)
1	0.750
0	0.100

The Alarm Network: Joint Query

- **Question:** What is the probability that there is a break-in, but no earthquake, the alarm goes off, the police call, but your neighbor does not call?

$P(B\!=\!1, E\!=\!0, A\!=\!1, PD\!=\!1, N\!=\!0)$ factorization

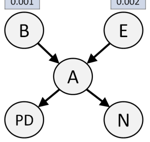
$= P(B\!=\!1)P(E\!=\!0)P(A\!=\!1|B\!=\!1, E\!=\!0)P(PD\!=\!1|A\!=\!1)P(N\!=\!0|A\!=\!1)$

$= 0.001 \cdot (1 - 0.002) \cdot 0.94 \cdot 0.9 \cdot (1 - 0.75)$

P(B=1)
0.001

P(E=1)
0.002

B	E	P(A=1|B,E)
1	1	0.950
1	0	0.940
0	1	0.290
0	0	0.001



A	P(PD=1|A)
1	0.900
0	0.005

A	P(N=1|A)
1	0.750
0	0.100

The Alarm Network: Marginal Query

- **Question:** What is the probability that there was a break-in, but no earthquake, the police call, but your neighbor does not call?

$P(B\!=\!1, E\!=\!0, PD\!=\!1, N\!=\!0)$

$$= \sum_{a=0}^{1} P(B\!=\!1, E\!=\!0, A\!=\!a, PD\!=\!1, N\!=\!0)$$

$= P(B\!=\!1)P(E\!=\!0)P(A\!=\!1|B\!=\!1, E\!=\!0)P(PD\!=\!1|A\!=\!1)P(N\!=\!0|A\!=\!1)$
$+ P(B\!=\!1)P(E\!=\!0)P(A\!=\!0|B\!=\!1, E\!=\!0)P(PD\!=\!1|A\!=\!0)P(N\!=\!0|A\!=\!0)$

$= 0.001 \cdot (1 - 0.002) \cdot 0.94 \cdot 0.9 \cdot (1 - 0.75)$
$+ 0.001 \cdot (1 - 0.002) \cdot (1 - 0.94) \cdot 0.005 \cdot (1 - 0.1)$

The Alarm Network: Conditional Query

- **Question:** What is the probability that the alarm went off given that there was a break-in, but no earthquake, the police call, but your neighbor does not call?

$P(A = 1|B\!=\!1, E\!=\!0, PD\!=\!1, N\!=\!0)$

$$= \frac{P(B\!=\!1, E\!=\!0, A\!=\!1, PD\!=\!1, N\!=\!0)}{\sum_{a=0}^{1} P(B\!=\!1, E\!=\!0, A\!=\!a, PD\!=\!1, N\!=\!0)}$$

$$= \frac{P(B\!=\!1)P(E\!=\!0)P(A\!=\!1|B\!=\!1, E\!=\!0)P(PD\!=\!1|A\!=\!1)P(N\!=\!0|A\!=\!1)}{\sum_{a=0}^{1} P(B\!=\!1)P(E\!=\!0)P(A\!=\!a|B\!=\!1, E\!=\!0)P(PD\!=\!1|A\!=\!a)P(N\!=\!0|A\!=\!a)}$$

The Alarm Network: More Queries

- What is the probability that there is a break-in given that there is an earthquake?

  What is the probability that your neighbor calls given that the alarm goes off and there is an earthquake?

  What is the probability that the police call given that the alarm goes off and your neighbor calls?

  What is the probability of a break-in given that the alarm goes off and the police call?

  What is the probability that your neighbor calls given that there is an earthquake?

  What is the probability that there is a break-in given that there is an earthquake and the alarm goes off?

  What is the probability that your neighbor calls given that the police call?