Announcements

- Solution to Assignment 5 posted
- Assignment 6 due next Monday

"The mind is a neural computer, fitted by natural selection with combinatorial algorithms for causal and probabilistic reasoning about plants, animals, objects, and people."

. . .

"In a universe with any regularities at all, decisions informed about the past are better than decisions made at random. That has always been true, and we would expect organisms, especially informavores such as humans, to have evolved acute intuitions about probability. The founders of probability, like the founders of logic, assumed they were just formalizing common sense."

Steven Pinker, How the Mind Works, 1997, pp. 524, 343.

Review: So far. . .

- An agent acts in an environment, inputs: abilities, goals/preferences, prior knowledge, observations, past experiences

Review: So far. . .

- An agent acts in an environment, inputs: abilities, goals/preferences, prior knowledge, observations, past experiences
- Search is used to find paths in graphs. Multiple-path pruning and depth bounds reduce search. Depth-first methods can save space.

Review: So far. . .

- An agent acts in an environment, inputs: abilities, goals/preferences, prior knowledge, observations, past experiences
- Search is used to find paths in graphs. Multiple-path pruning and depth bounds reduce search. Depth-first methods can save space.
- Constraint satisfaction problems are defined in terms of variables, domains, constraints. Constraint satisfactions problems can be solved with: backtracking search, arc consistency + domain splitting, local search

Review: So far. . .

- An agent acts in an environment, inputs: abilities, goals/preferences, prior knowledge, observations, past experiences
- Search is used to find paths in graphs. Multiple-path pruning and depth bounds reduce search. Depth-first methods can save space.
- Constraint satisfaction problems are defined in terms of variables, domains, constraints. Constraint satisfactions problems can be solved with: backtracking search, arc consistency + domain splitting, local search
- Planning is finding a sequence of actions to achieve a goal. Planning is achieved by mapping to a search problem (forwards or regression) or a CSP.

Learning Objectives

At the end of the class you should be able to:

- justify the use and semantics of probability
- know how to compute marginals and apply Bayes' theorem
- identify conditional independence
- build a belief network for a domain

Review of Pre-class slides

- Probability is defined in terms of measures over possible worlds
- The probability of a proposition is the measure of the set of worlds in which the proposition is true.
- Conditioning on evidence: make the worlds incompatible with the evidence have measure 0 and multiply the others by a constant, to get a measure.
- A belief network is a representation of conditional independence:
  in a total ordering of the variables, each variable is independent of its predecessors given it's parents

Possible World Semantics

- A possible world specifies an assignment of one value to each random variable.
- A random variable is a function from possible worlds into the domain of the random variable.

Possible World Semantics

- A possible world specifies an assignment of one value to each random variable.
- A random variable is a function from possible worlds into the domain of the random variable.
- $\omega \models X = x$
  means variable X is assigned value x in world ω.

Possible World Semantics

- A possible world specifies an assignment of one value to each random variable.
- A random variable is a function from possible worlds into the domain of the random variable.
- $\omega \models X = x$
  means variable X is assigned value x in world ω.
- Logical connectives have their standard meaning:

  $\omega \models \alpha \land \beta$ if $\omega \models \alpha$ and $\omega \models \beta$

  $\omega \models \alpha \lor \beta$ if $\omega \models \alpha$ or $\omega \models \beta$

  $\omega \models \neg\alpha$ if $\omega \not\models \alpha$
- Let Ω be the set of all possible worlds.

Semantics of Probability

Probability defines a measure on sets of possible worlds.
A probability measure is a function μ from sets of worlds into the non-negative real numbers such that:

- $\mu(\Omega) = 1$
- $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$
  if $S_1 \cap S_2 = \{\}$.

Semantics of Probability

Probability defines a measure on sets of possible worlds.
A probability measure is a function μ from sets of worlds into the non-negative real numbers such that:

- $\mu(\Omega) = 1$
- $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$
  if $S_1 \cap S_2 = \{\}$.

Then $P(\alpha) = \mu(\{\omega \mid \omega \models \alpha\})$.
"The probability of α is the measure of the set of possible worlds in which α is true."

Conditioning

- Probabilistic conditioning specifies how to revise beliefs based on new information.

Conditioning

- Probabilistic conditioning specifies how to revise beliefs based on new information.
- An agent builds a probabilistic model taking all background information into account.
  This gives the prior probability.

Conditioning

- Probabilistic conditioning specifies how to revise beliefs based on new information.
- An agent builds a probabilistic model taking all background information into account.
  This gives the prior probability.
- All other information must be conditioned on.
- If evidence e is the all of the information obtained subsequently, the conditional probability $P(h \mid e)$ of h given e is the posterior probability of h.

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

$$\mu_e(S) = \left\{ \right.$$

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

$$\mu_e(S) = \begin{cases} & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

$$\mu_e(S) = \begin{cases} 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

$$\mu_e(S) = \begin{cases} & \text{if } \omega \models e \text{ for all } \omega \in S \\ 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

$$\mu_e(S) = \begin{cases} c \times \mu(S) & \text{if } \omega \models e \text{ for all } \omega \in S \\ 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

$$\mu_e(S) = \begin{cases} c \times \mu(S) & \text{if } \omega \models e \text{ for all } \omega \in S \\ 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

We can show $c =$

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

$$\mu_e(S) = \begin{cases} c \times \mu(S) & \text{if } \omega \models e \text{ for all } \omega \in S \\ 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

We can show $c = \frac{1}{P(e)}$.

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

$$\mu_e(S) = \begin{cases} c \times \mu(S) & \text{if } \omega \models e \text{ for all } \omega \in S \\ 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

  We can show $c = \frac{1}{P(e)}$.
- The conditional probability of formula h given evidence e is

$$P(h \mid e) \quad =$$

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

$$\mu_e(S) = \begin{cases} c \times \mu(S) & \text{if } \omega \models e \text{ for all } \omega \in S \\ 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

  We can show $c = \frac{1}{P(e)}$.
- The conditional probability of formula h given evidence e is

$$\begin{aligned} P(h \mid e) &= \mu_e(\{\omega : \omega \models h\}) \\ &= \end{aligned}$$

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds:

$$\mu_e(S) = \begin{cases} c \times \mu(S) & \text{if } \omega \models e \text{ for all } \omega \in S \\ 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

  We can show $c = \frac{1}{P(e)}$.
- The conditional probability of formula h given evidence e is

$$\begin{aligned} P(h \mid e) &= \mu_e(\{\omega : \omega \models h\}) \\ &= \frac{P(h \wedge e)}{P(e)} \end{aligned}$$

Conditioning

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \land sneeze)$ 0.16

(b) $P(flu \land \neg sneeze)$

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker)

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

(h) $P(sneeze \mid flu \wedge snore)$

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

(h) $P(sneeze \mid flu \wedge snore)$ 0.8

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

(h) $P(sneeze \mid flu \wedge snore)$ 0.8

(i) $P(flu \mid sneeze \wedge snore)$

Clicker Question

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16
(b) $P(flu \wedge \neg sneeze)$ 0.04
(c) $P(flu)$ (not clicker) 0.2
(d) $P(sneeze \mid flu)$ 0.8
(e) $P(\neg flu \wedge sneeze)$ 0.24
(f) $P(sneeze)$ 0.4
(g) $P(flu \mid sneeze)$ 0.4
(h) $P(sneeze \mid flu \wedge snore)$ 0.8
(i) $P(flu \mid sneeze \wedge snore)$ 0.4

Chain Rule: probability of conjunctions

$$P(h \mid e) = \frac{P(h \wedge e)}{P(e)}$$

Therefore

$$P(h \wedge e) =$$

Chain Rule: probability of conjunctions

$$P(h \mid e) = \frac{P(h \wedge e)}{P(e)}$$

Therefore

$$P(h \wedge e) = P(h \mid e) \times P(e)$$

Chain Rule

Semantics of conditioning gives: $P(h \wedge e) = P(h \mid e) \times P(e)$

Chain Rule

Semantics of conditioning gives: $P(h \wedge e) = P(h \mid e) \times P(e)$

$$P(f_n \wedge f_{n-1} \wedge \ldots \wedge f_1)$$
$$=$$

Chain Rule

Semantics of conditioning gives: $P(h \wedge e) = P(h \mid e) \times P(e)$

$$
\begin{aligned}
& P(f_n \wedge f_{n-1} \wedge \ldots \wedge f_1) \\
& \quad = \quad P(f_n \mid f_{n-1} \wedge \cdots \wedge f_1) \times \\
& \qquad\quad P(f_{n-1} \wedge \cdots \wedge f_1) \\
& \quad =
\end{aligned}
$$

Chain Rule

Semantics of conditioning gives: $P(h \wedge e) = P(h \mid e) \times P(e)$

$$
\begin{aligned}
& P(f_n \wedge f_{n-1} \wedge \ldots \wedge f_1) \\
= \quad & P(f_n \mid f_{n-1} \wedge \cdots \wedge f_1) \times \\
& P(f_{n-1} \wedge \cdots \wedge f_1) \\
= \quad & P(f_n \mid f_{n-1} \wedge \cdots \wedge f_1) \times \\
& P(f_{n-1} \mid f_{n-2} \wedge \cdots \wedge f_1) \times \\
& P(f_{n-2} \wedge \cdots \wedge f_1) \\
= \quad & P(f_n \mid f_{n-1} \wedge \cdots \wedge f_1) \times \\
& P(f_{n-1} \mid f_{n-2} \wedge \cdots \wedge f_1) \\
& \times \cdots \times P(f_3 \mid f_2 \wedge f_1) \times P(f_2 \mid f_1) \times P(f_1) \\
= \quad & \prod_{i=1}^{n} P(f_i \mid f_1 \wedge \cdots \wedge f_{i-1})
\end{aligned}
$$

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$P(h \wedge e) =$$

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$P(h \wedge e) = P(h \mid e) \times P(e)$$

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$\begin{aligned} P(h \wedge e) &= P(h \mid e) \times P(e) \\ &= P(e \mid h) \times P(h). \end{aligned}$$

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$\begin{aligned} P(h \wedge e) &= P(h \mid e) \times P(e) \\ &= P(e \mid h) \times P(h). \end{aligned}$$

If $P(e) \neq 0$, divide the right hand sides by $P(e)$:

$$P(h \mid e) =$$

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$\begin{aligned} P(h \wedge e) &= P(h \mid e) \times P(e) \\ &= P(e \mid h) \times P(h). \end{aligned}$$

If $P(e) \neq 0$, divide the right hand sides by $P(e)$:

$$P(h \mid e) = \frac{P(e \mid h) \times P(h)}{P(e)}.$$

This is Bayes' theorem.

Conditional independence

Random variable X is independent of random variable Y given random variable(s) Z if,

$$P(X \mid Y, Z) = P(X \mid Z)$$

Conditional independence

Random variable X is independent of random variable Y given random variable(s) Z if,

$$P(X \mid Y, Z) = P(X \mid Z)$$

i.e. for all $x_i \in domain(X)$, $y_j \in domain(Y)$, $y_k \in domain(Y)$ and $z_m \in domain(Z)$,

$$\begin{aligned} & P(X = x_i \mid Y = y_j \wedge Z = z_m) \\ = \ & P(X = x_i \mid Y = y_k \wedge Z = z_m) \\ = \ & P(X = x_i \mid Z = z_m). \end{aligned}$$

That is, knowledge of Y's value doesn't affect the belief in the value of X, given a value of Z.

Belief networks

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Theorem of probability theory (chain rule):
  $P(X_1, \ldots, X_n) =$

Belief networks

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Theorem of probability theory (chain rule):
  $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$

Belief networks

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Theorem of probability theory (chain rule):
  $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$
- The parents $parents(X_i)$ of X_i are those predecessors of X_i that render X_i independent of the other predecessors. That is,

Belief networks

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Theorem of probability theory (chain rule): $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$
- The parents $parents(X_i)$ of X_i are those predecessors of X_i that render X_i independent of the other predecessors. That is, $parents(X_i) \subseteq X_1, \ldots, X_{i-1}$ and $P(X_i \mid parents(X_i)) = P(X_i \mid X_1, \ldots, X_{i-1})$

Belief networks

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Theorem of probability theory (chain rule): $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$
- The parents $parents(X_i)$ of X_i are those predecessors of X_i that render X_i independent of the other predecessors. That is, $parents(X_i) \subseteq X_1, \ldots, X_{i-1}$ and $P(X_i \mid parents(X_i)) = P(X_i \mid X_1, \ldots, X_{i-1})$
- So $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$

Belief networks

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Theorem of probability theory (chain rule):
  $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$
- The parents $parents(X_i)$ of X_i are those predecessors of X_i that render X_i independent of the other predecessors. That is, $parents(X_i) \subseteq X_1, \ldots, X_{i-1}$ and
  $P(X_i \mid parents(X_i)) = P(X_i \mid X_1, \ldots, X_{i-1})$
- So $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$
- A belief network is a graph: the nodes are random variables; there is an arc from the parents of each node into that node.

Student Writing an Exam Example

Give a belief network for the variables in order:

- *WorksHard*: Whether the student works hard
- *Intelligent*: Whether the student is intelligent
- *Answers*: The student's answers on the exam
- *Mark*: The student's mark on an exam

Student Writing an Exam Example

Give a belief network for the variables in order:

- *WorksHard*: Whether the student works hard
- *Intelligent*: Whether the student is intelligent
- *Answers*: The student's answers on the exam
- *Mark*: The student's mark on an exam

What if the variables were in the opposite order?

Example: fire alarm belief network

Variables:

- Fire: there is a fire in the building
- Tampering: someone has been tampering with the fire alarm
- Smoke: what appears to be smoke is coming from an upstairs window
- Alarm: the fire alarm goes off
- Leaving: people are leaving the building *en masse*.
- Report: a colleague says that people are leaving the building *en masse*. (A noisy sensor for leaving.)

Example: fire alarm belief network

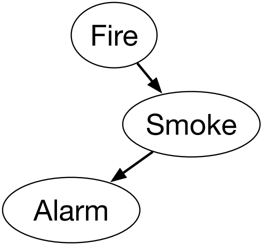
Variables:

- Fire: there is a fire in the building
- Tampering: someone has been tampering with the fire alarm
- Smoke: what appears to be smoke is coming from an upstairs window
- Alarm: the fire alarm goes off
- Leaving: people are leaving the building *en masse*.
- Report: a colleague says that people are leaving the building *en masse*. (A noisy sensor for leaving.)

See "Fire Alarm Belief Network" in AIspace.org Belief and Decision Networks App

Clicker Question

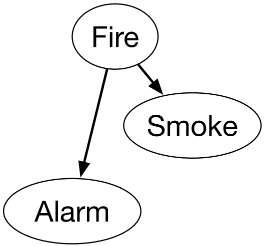
For the belief network, and the ordering *Fire*, *Smoke*, *Alarm*



- A *Alarm* is independent of *Smoke* given *Fire*
- B *Alarm* is independent of *Fire* given *Smoke*
- C *Alarm* is independent of *Fire* given $\{\}$
- D All of the above independencies hold
- E There are no independencies

Clicker Question

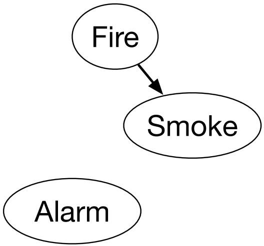
For the belief network, and the ordering *Fire*, *Smoke*, *Alarm*



- A *Alarm* is independent of *Smoke* given *Fire*
- B *Alarm* is independent of *Fire* given *Smoke*
- C *Alarm* is independent of *Fire* given $\{\}$
- D All of the above independencies hold
- E There are no independencies

Clicker Question

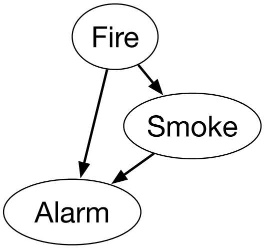
For the belief network, and the ordering *Fire*, *Smoke*, *Alarm*



- A *Alarm* is independent of *Smoke* given *Fire*
- B *Alarm* is independent of *Fire* given *Smoke*
- C *Alarm* is independent of *Fire* given {}
- D All of the above independencies hold
- E There are no independencies

Clicker Question

For the belief network, and the ordering *Fire*, *Smoke*, *Alarm*



A *Alarm* is independent of *Smoke* given *Fire*

B *Alarm* is independent of *Fire* given *Smoke*

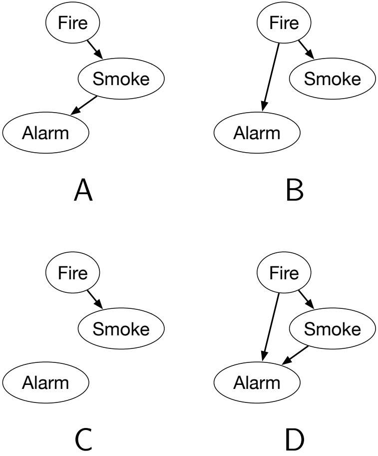
C *Alarm* is independent of *Fire* given $\{\}$

D All of the above independencies hold

E There are no independencies

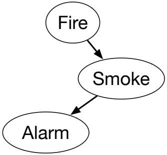
Clicker Question

Which network best fits a fire alarm that only detects the heat of the fire?

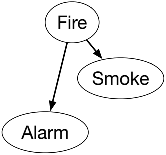


Clicker Question

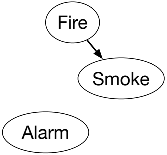
Which network best fits a smoke alarm (that only detects smoke)?



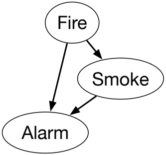
A



B



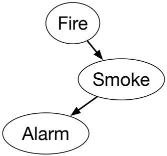
C



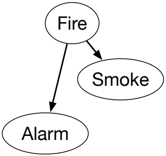
D

Clicker Question

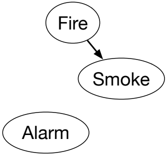
Which network best fits a fire alarm that detects both smoke and the heat of the fire?



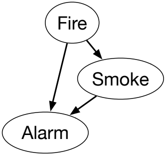
A



B



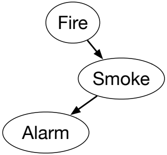
C



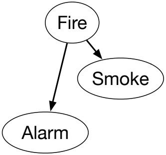
D

Clicker Question

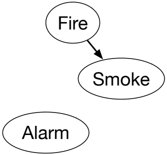
Which network best fits a burglary alarm that doesn't detect heat or smoke?



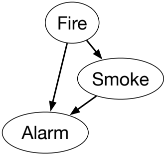
A



B



C



D

Components of a belief network

A belief network consists of:

- a directed acyclic graph with nodes labeled with random variables
- a domain for each random variable
- a set of conditional probabilities, one for each variable given its parents (including prior probabilities for nodes with no parents).

Belief network summary

- A belief network is a directed acyclic graph (DAG) where nodes are random variables.

Belief network summary

- A belief network is a directed acyclic graph (DAG) where nodes are random variables.
- The parents of a node n are those variables on which n directly depends.

Belief network summary

- A belief network is a directed acyclic graph (DAG) where nodes are random variables.
- The parents of a node n are those variables on which n directly depends.
- A belief network is automatically acyclic by construction.

Belief network summary

- A belief network is a directed acyclic graph (DAG) where nodes are random variables.
- The parents of a node n are those variables on which n directly depends.
- A belief network is automatically acyclic by construction.
- A belief network is a graphical representation of dependence and independence:
  - A variable is independent of its non-descendants given its parents.

Constructing belief networks

To represent a domain in a belief network, you need to consider:

- What are the relevant variables?

Constructing belief networks

To represent a domain in a belief network, you need to consider:

- What are the relevant variables?
  - What will you observe?

Constructing belief networks

To represent a domain in a belief network, you need to consider:

- What are the relevant variables?
  - What will you observe?
  - What would you like to find out (query)?

Constructing belief networks

To represent a domain in a belief network, you need to consider:

- What are the relevant variables?
  - What will you observe?
  - What would you like to find out (query)?
  - What other features make the model simpler?

Constructing belief networks

To represent a domain in a belief network, you need to consider:

- What are the relevant variables?
  - What will you observe?
  - What would you like to find out (query)?
  - What other features make the model simpler?
- What values should these variables take?

Constructing belief networks

To represent a domain in a belief network, you need to consider:

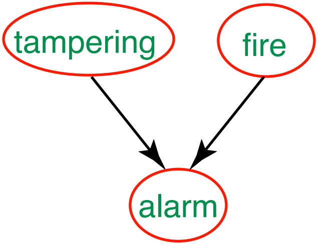
- What are the relevant variables?
  - What will you observe?
  - What would you like to find out (query)?
  - What other features make the model simpler?
- What values should these variables take?
- What is the relationship between them? This should be expressed in terms of a directed graph, representing how each variable is generated from its predecessors.

Constructing belief networks

To represent a domain in a belief network, you need to consider:

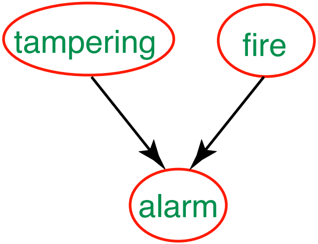
- What are the relevant variables?
  - What will you observe?
  - What would you like to find out (query)?
  - What other features make the model simpler?
- What values should these variables take?
- What is the relationship between them? This should be expressed in terms of a directed graph, representing how each variable is generated from its predecessors.
- How does the value of each variable depend on its parents? This is expressed in terms of the conditional probabilities.

Understanding Independence: Common descendants



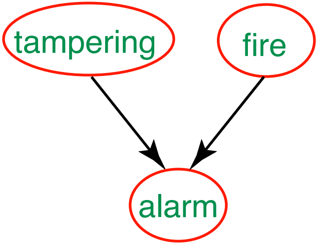
- *tampering* and *fire* are

Understanding Independence: Common descendants



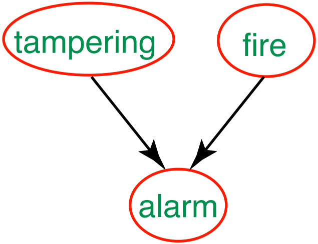
- *tampering* and *fire* are independent

Understanding Independence: Common descendants



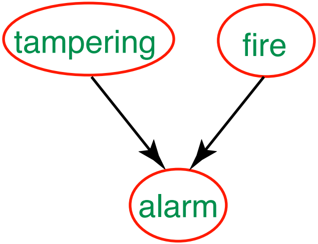
- *tampering* and *fire* are independent
- *tampering* and *fire* are given *alarm*

Understanding Independence: Common descendants



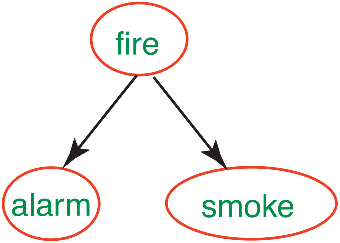
- *tampering* and *fire* are independent
- *tampering* and *fire* are dependent given *alarm*

Understanding Independence: Common descendants



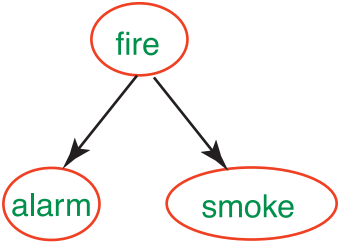
- *tampering* and *fire* are independent
- *tampering* and *fire* are dependent given *alarm*
- Intuitively, *tampering* can explain away *fire*

Understanding Independence: Common ancestors



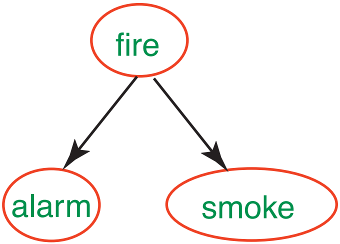
- *alarm* and *smoke* are

Understanding Independence: Common ancestors



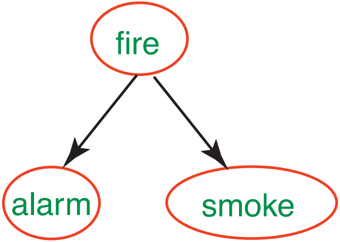
- *alarm* and *smoke* are dependent

Understanding Independence: Common ancestors



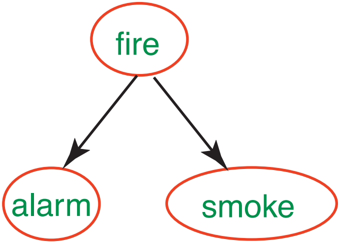
- *alarm* and *smoke* are dependent
- *alarm* and *smoke* are given *fire*

Understanding Independence: Common ancestors



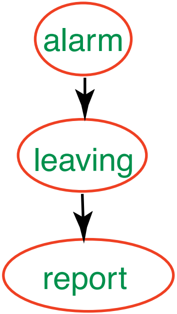
- *alarm* and *smoke* are dependent
- *alarm* and *smoke* are independent given *fire*

Understanding Independence: Common ancestors



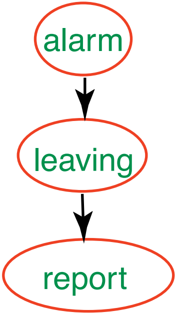
- *alarm* and *smoke* are dependent
- *alarm* and *smoke* are independent given *fire*
- Intuitively, *fire* can explain *alarm* and *smoke*; learning one can affect the other by changing your belief in *fire*.

Understanding Independence: Chain



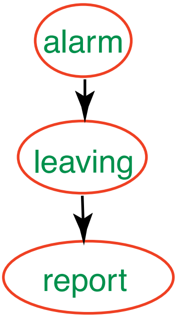
- *alarm* and *report* are

Understanding Independence: Chain



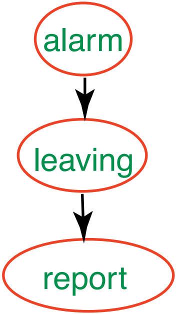
- *alarm* and *report* are dependent

Understanding Independence: Chain



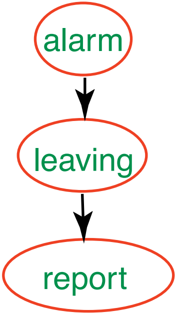
- *alarm* and *report* are dependent
- *alarm* and *report* are given *leaving*

Understanding Independence: Chain



- *alarm* and *report* are dependent
- *alarm* and *report* are independent given *leaving*

Understanding Independence: Chain



- *alarm* and *report* are dependent
- *alarm* and *report* are independent given *leaving*
- Intuitively, the only way that the *alarm* affects *report* is by affecting *leaving*.

Pruning Irrelevant Variables

Suppose you want to compute $P(X \mid e_1 \ldots e_k)$:

- Prune any variables that have no observed or queried descendents.
- Connect the parents of any observed variable.
- Remove arc directions.
- Remove observed variables.
- Remove any variables not connected to X in the resulting (undirected) graph.