These slides are adapted from Poole & Mackworth, chap 8

From a Constraint Satisfaction Problem [Var,Dom,Con] to random variables with probabilities constrained by a graph

- The domain (range) of a variable X, written Dom(X), is the set of (possible) values X can take.

These slides are adapted from Poole & Mackworth, chap 8

From a Constraint Satisfaction Problem [Var,Dom,Con] to random variables with probabilities constrained by a graph

- The domain (range) of a variable X, written $\mathrm{Dom}(X)$, is the set of (possible) values X can take.
- A proposition α is an equation $X = x$ between a variable X and a value $x \in \mathrm{Dom}(X)$, or a Boolean combination of such.

These slides are adapted from Poole & Mackworth, chap 8

From a Constraint Satisfaction Problem [Var,Dom,Con] to random variables with probabilities constrained by a graph

- The domain (range) of a variable X, written Dom(X), is the set of (possible) values X can take.
- A proposition α is an equation $X = x$ between a variable X and a value $x \in$ Dom(X), or a Boolean combination of such.
- A proposition α is assigned a probability through
  - a notion $\models$ of a possible world ω satisfying α, and
  - a measure μ for weighing a set of possible worlds.

Satisfaction, measure and probability

Fix a set Ω of possible worlds ω that assign a value to each random variable, and interpret a proposition via $\models$

$$\begin{aligned}
\omega \models X = x &\iff \omega \text{ assigns } X \text{ the value } x\\
\omega \models \alpha \wedge \beta &\iff \omega \models \alpha \text{ and } \omega \models \beta\\
\omega \models \alpha \vee \beta &\iff \omega \models \alpha \text{ or } \omega \models \beta\\
\omega \models \neg\alpha &\iff \omega \not\models \alpha.
\end{aligned}$$

Satisfaction, measure and probability

Fix a set Ω of possible worlds ω that assign a value to each random variable, and interpret a proposition via $\models$

$$\begin{aligned}
\omega \models X = x &\iff \omega \text{ assigns } X \text{ the value } x \\
\omega \models \alpha \wedge \beta &\iff \omega \models \alpha \text{ and } \omega \models \beta \\
\omega \models \alpha \vee \beta &\iff \omega \models \alpha \text{ or } \omega \models \beta \\
\omega \models \neg\alpha &\iff \omega \not\models \alpha.
\end{aligned}$$

For finite Ω, a probability measure is a function

$$\mu : Pow(\Omega) \rightarrow [0, 1]$$

such that $\mu(\Omega) = 1$ and for any subset S of Ω,

$$\mu(S) = \sum_{\omega \in S} \mu(\{\omega\}).$$

Satisfaction, measure and probability

Fix a set Ω of possible worlds ω that assign a value to each random variable, and interpret a proposition via $\models$

$$
\begin{aligned}
\omega \models X = x \quad &\iff \quad \omega \text{ assigns } X \text{ the value } x \\
\omega \models \alpha \wedge \beta \quad &\iff \quad \omega \models \alpha \text{ and } \omega \models \beta \\
\omega \models \alpha \vee \beta \quad &\iff \quad \omega \models \alpha \text{ or } \omega \models \beta \\
\omega \models \neg\alpha \quad &\iff \quad \omega \not\models \alpha.
\end{aligned}
$$

For finite Ω, a probability measure is a function

$$\mu : Pow(\Omega) \to [0, 1]$$

such that $\mu(\Omega) = 1$ and for any subset S of Ω,

$$\mu(S) = \sum_{\omega \in S} \mu(\{\omega\}).$$

Given μ, a proposition α has probability

$$P(\alpha) = \mu(\{\omega \mid \omega \models \alpha\}).$$

Tuples, distributions and the sum rule

A tuple $X_1, \ldots, X_n$ of random variables is a random variable with domain

$$\mathrm{Dom}(X_1) \times \cdots \times \mathrm{Dom}(X_n).$$

Tuples, distributions and the sum rule

A tuple $X_1, \ldots, X_n$ of random variables is a random variable with domain

$$\mathrm{Dom}(X_1) \times \cdots \times \mathrm{Dom}(X_n).$$

A probability distribution on a random variable X is a function $P_X : \mathrm{Dom}(X) \to [0,1]$ s.t.

$$P_X(x) = P(X = x).$$

P_X is often written as $P(X)$, and $P_X(x)$ as $P(x)$.

Tuples, distributions and the sum rule

A tuple $X_1, \ldots, X_n$ of random variables is a random variable with domain

$$\mathrm{Dom}(X_1) \times \cdots \times \mathrm{Dom}(X_n).$$

A probability distribution on a random variable X is a function $P_X : \mathrm{Dom}(X) \to [0,1]$ s.t.

$$P_X(x) = P(X = x).$$

P_X is often written as $P(X)$, and $P_X(x)$ as $P(x)$.

sum rule
$$P(X) = \sum_Y P(X, Y)$$
$$P_X(x) = \sum_{y \in \mathrm{Dom}(Y)} P_{X,Y}(x, y) \quad \text{for } x \in \mathrm{Dom}(X)$$

Conditional probability

To incorporate a proposition α into the background assumptions, we restrict the set Ω of possible worlds to

$$\Omega \restriction \alpha \ := \ \{\omega \in \Omega \mid \omega \models \alpha\}$$

and assuming $\mu(\Omega \restriction \alpha) \neq 0$, map a subset $S \subseteq \Omega \restriction \alpha$ to

$$\mu^{\alpha}(S) \ := \ \frac{\mu(S)}{\mu(\Omega \restriction \alpha)}$$

for a probability measure $\mu^{\alpha} : Pow(\Omega \restriction \alpha) \to [0,1]$ on $\Omega \restriction \alpha$.

Conditional probability

To incorporate a proposition α into the background assumptions, we restrict the set Ω of possible worlds to

$$\Omega \restriction \alpha \; := \; \{\omega \in \Omega \mid \omega \models \alpha\}$$

and assuming $\mu(\Omega \restriction \alpha) \neq 0$, map a subset $S \subseteq \Omega \restriction \alpha$ to

$$\mu^{\alpha}(S) \; := \; \frac{\mu(S)}{\mu(\Omega \restriction \alpha)}$$

for a probability measure $\mu^{\alpha} : Pow(\Omega \restriction \alpha) \to [0, 1]$ on $\Omega \restriction \alpha$.

The conditional probability of α' given α is

$$P(\alpha' \mid \alpha) \; := \; \mu^{\alpha}(\Omega \restriction \alpha' \wedge \alpha) \; = \; \frac{P(\alpha' \wedge \alpha)}{P(\alpha)}$$

The product rule and Bayes' theorem

product rule

$$P(X, Y) = P(X|Y)P(Y)$$

$$P_{X,Y}(x, y) = P_X(x|Y = y)P_Y(y)$$

$$\text{for } x \in \mathrm{Dom}(X),\ \ y \in \mathrm{Dom}(Y)$$

The product rule and Bayes' theorem

product rule $$\begin{aligned} P(X, Y) &= P(X|Y)P(Y) \\ P_{X,Y}(x, y) &= P_X(x|Y = y)P_Y(y) \\ &\quad \text{for } x \in \text{Dom}(X),\ \ y \in \text{Dom}(Y) \end{aligned}$$

As conjunction is commutative $(\Omega \restriction \alpha' \wedge \alpha = \Omega \restriction \alpha \wedge \alpha')$,

$$P(X, Y) = P(Y, X)$$

and so the product rule yields

Bayes' theorem $$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \quad \text{if } P(Y) \neq 0$$

The product rule and Bayes' theorem

product rule
$$P(X,Y) = P(X|Y)P(Y)$$
$$P_{X,Y}(x,y) = P_X(x|Y=y)P_Y(y)$$
$$\text{for } x \in \mathrm{Dom}(X),\;\; y \in \mathrm{Dom}(Y)$$

As conjunction is commutative ($\Omega \restriction \alpha' \wedge \alpha = \Omega \restriction \alpha \wedge \alpha'$),

$$P(X,Y) = P(Y,X)$$

and so the product rule yields

Bayes' theorem
$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \quad \text{if } P(Y) \neq 0$$

The prior probability of α

$$P(\alpha) = \mu(\Omega \restriction \alpha)$$

is updated by $\alpha_\circ$ to the posterior probability given $\alpha_\circ$

$$P(\alpha \mid \alpha_\circ) = \mu^{\alpha_\circ}(\Omega \restriction (\alpha \wedge \alpha_\circ))$$

Why is Bayes' theorem interesting?

Form a hypothesis h given evidence e with $P(e) \neq 0$ via Bayes

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

Why is Bayes' theorem interesting?

Form a hypothesis h given evidence e with $P(e) \neq 0$ via Bayes

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

We often have causal knowledge

$$P(\text{symptom} \mid \text{disease}), \quad P(\text{alarm} \mid \text{fire})$$

$P(\text{image} =$  $\mid \text{a tree is in front of a car})$

but want to do evidential reasoning

$$P(\text{disease} \mid \text{symptom}), \quad P(\text{fire} \mid \text{alarm})$$

$P(\text{a tree is in front of a car} \mid \text{image} =$ $)$

Tuples and the chain rule

Recall: a tuple $X_1, \ldots, X_n$ of random variables is a random variable.

Let us write

$$X_{1:n} \quad \text{for} \quad X_1, \ldots, X_n$$

Tuples and the chain rule

Recall: a tuple $X_1, \ldots, X_n$ of random variables is a random variable.

Let us write

$$X_{1:n} \text{ for } X_1, \ldots, X_n$$

and apply the product rule repeatedly for

$$\begin{aligned} P(X_{1:n}) &= P(X_n \,|\, X_{1:n-1}) P(X_{1:n-1}) \\ &= P(X_n \,|\, X_{1:n-1}) P(X_{n-1} \,|\, X_{1:n-2}) P(X_{1:n-2}) \\ &= \cdots \\ &= \prod_{i=1}^{n} P(X_i \,|\, X_{1:i-1}) \quad \text{chain rule} \end{aligned}$$

with $X_{1:0}$ as the empty tuple and $P(X_1 \,|\, X_{1:0}) = P(X_1)$.

Simplifying the chain rule via conditional independence

Choose a sub-tuple $parents(X_i)$ of $X_{1:i-1}$ such that

$$P(X_i \mid X_{1:i-1}) \;=\; P(X_i \mid parents(X_i))\ .$$

Simplifying the chain rule via conditional independence

Choose a sub-tuple $\mathit{parents}(X_i)$ of $X_{1:i-1}$ such that

$$P(X_i \mid X_{1:i-1}) \;=\; P(X_i \mid \mathit{parents}(X_i)) \; .$$

X is independent of Y given Z, written $X \perp\!\!\!\perp Y \mid Z$,

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

Simplifying the chain rule via conditional independence

Choose a sub-tuple $parents(X_i)$ of $X_{1:i-1}$ such that

$$P(X_i \mid X_{1:i-1}) \;=\; P(X_i \mid parents(X_i)) \;.$$

X is independent of Y given Z, written $X \perp\!\!\!\perp Y \mid Z$,

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

i.e. for all $x \in \mathrm{Dom}(X)$, $y \in \mathrm{Dom}(Y)$, and $z \in \mathrm{Dom}(Z)$,

$$P(X = x \mid Y = y \wedge Z = z) \;=\; P(X = x \mid Z = z)$$

— knowing Y's value says nothing about X's value, given Z's value.

Simplifying the chain rule via conditional independence

Choose a sub-tuple $\mathit{parents}(X_i)$ of $X_{1:i-1}$ such that

$$P(X_i \mid X_{1:i-1}) \;=\; P(X_i \mid \mathit{parents}(X_i))\ .$$

X is independent of Y given Z, written $X \perp\!\!\!\perp Y \mid Z$,

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

i.e. for all $x \in \mathrm{Dom}(X)$, $y \in \mathrm{Dom}(Y)$, and $z \in \mathrm{Dom}(Z)$,

$$P(X = x \mid Y = y \wedge Z = z) \;=\; P(X = x \mid Z = z)$$

— knowing Y's value says nothing about X's value, given Z's value.

Note

$$\begin{aligned} X \perp\!\!\!\perp Y \mid Z \quad &\Longleftrightarrow \quad P(X, Y \mid Z) \;=\; P(X \mid Z)P(Y \mid Z) \\ &\Longleftrightarrow \quad Y \perp\!\!\!\perp X \mid Z \end{aligned}$$

Totally order the variables of interest

$$X_1 < X_2 < \cdots < X_n$$

and for each i from 1 to n, choose $parents(X_i)$ from $X_{1:i-1}$ s.t.

$$P(X_i \mid X_{1:i-1}) \;=\; P(X_i \mid parents(X_i)) \qquad (\dagger)$$

Belief networks

Totally order the variables of interest

$$X_1 < X_2 < \cdots < X_n$$

and for each i from 1 to n, choose $parents(X_i)$ from $X_{1:i-1}$ s.t.

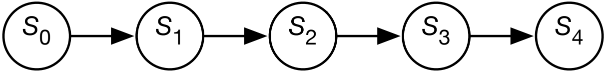
$$P(X_i \mid X_{1:i-1}) \;=\; P(X_i \mid parents(X_i)) \qquad (\dagger)$$

A belief network consists of:

- a directed acyclic graph with nodes = random variables, and an arc from the parents of each node into that node
- a domain for each random variable
- conditional probability tables for each variable given its parents respecting ($\dagger$)

Example: Markov chain

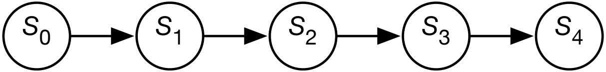
A Markov chain is a special sort of belief network:



What probabilities need to be specified?

Example: Markov chain

A Markov chain is a special sort of belief network:

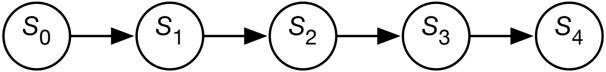


What probabilities need to be specified?

- $P(S_0)$ specifies initial conditions
- $P(S_{t+1}|S_t)$ specifies the dynamics

Example: Markov chain

A Markov chain is a special sort of belief network:



What probabilities need to be specified?

- $P(S_0)$ specifies initial conditions
- $P(S_{t+1}|S_t)$ specifies the dynamics

What independence assumptions are made?

$$P(S_{t+1}|S_{0:t}) = P(S_{t+1}|S_t)$$

S_t represents the state at time t, capturing everything about the past ($< t$) that can affect the future ($> t$)

The future is independent of the past given the present.

Two elaborations

In a stationary Markov chain,

$$\mathrm{Dom}(S_i) = \mathrm{Dom}(S_0) \;\text{ and }\; P(S_{i+1}|S_i) = P(S_1|S_0) \;\text{ for all } i \geq 0$$

so it is enough to specify $P(S_0)$ and $P(S_1|S_0)$.

- Simple model, easy to specify
- The network can extend indefinitely

Two elaborations

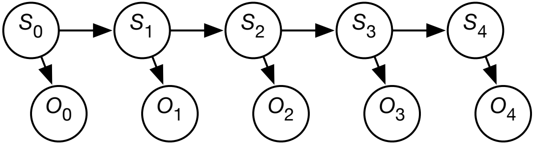
In a stationary Markov chain,

$$\mathrm{Dom}(S_i) = \mathrm{Dom}(S_0) \quad \text{and} \quad P(S_{i+1}|S_i) = P(S_1|S_0) \quad \text{for all } i \geq 0$$

so it is enough to specify $P(S_0)$ and $P(S_1|S_0)$.

- Simple model, easy to specify
- The network can extend indefinitely

A Hidden Markov Model (HMM) is a belief network of the form



- $P(S_0)$ specifies initial conditions
- $P(S_{i+1}|S_i)$ specifies the dynamics
- $P(O_i|S_i)$ specifies the sensor model

Naive Bayes Classifier

Problem: classify on the basis of features F_i

$$P(Class|F_{1:n}) = \frac{P(F_{1:n}|Class)P(Class)}{P(F_{1:n})}$$

Naive Bayes Classifier

Problem: classify on the basis of features F_i

$$P(Class|F_{1:n}) = \frac{P(F_{1:n}|Class)P(Class)}{P(F_{1:n})}$$

Assume F_i are independent of each other given $Class$

$$P(F_{1:n}|Class) = \prod_i P(F_i|Class)$$

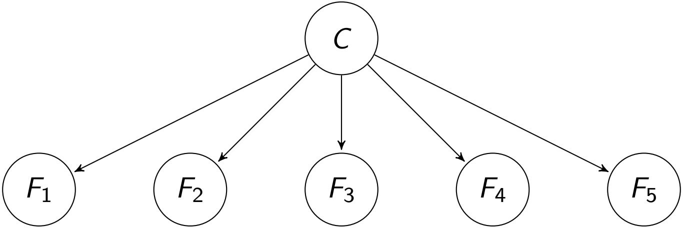
Naive Bayes Classifier

Problem: classify on the basis of features F_i

$$P(Class|F_{1:n}) = \frac{P(F_{1:n}|Class)P(Class)}{P(F_{1:n})}$$

Assume F_i are independent of each other given $Class$

$$P(F_{1:n}|Class) = \prod_i P(F_i|Class)$$



Assume the values of features F_i are predictable given a class.

Requires $P(Class)$ and $P(F_i|Class)$ for each F_i

Learning Probabilities

F_1	F_2	F_3	F_4	C	$Count$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
t	f	t	t	1	40
t	f	t	t	2	10
t	f	t	t	3	50
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Learning Probabilities

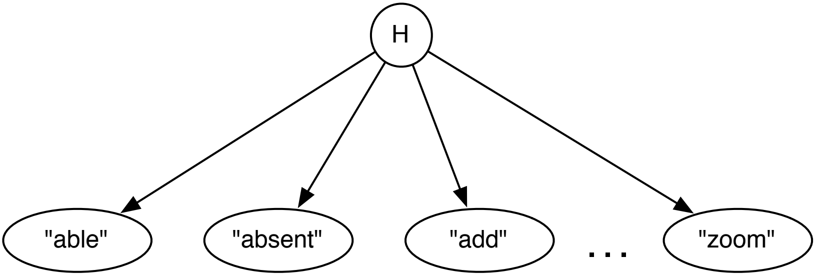
F_1	F_2	F_3	F_4	C	$Count$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
t	f	t	t	1	40
t	f	t	t	2	10
t	f	t	t	3	50
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$P(C{=}c) = \frac{\sum_{\omega \models C=c} Count(\omega)}{\sum_{\omega} Count(\omega)}$$

$$P(F_k = b | C{=}c) = \frac{\sum_{\omega \models C=c \wedge F_k=b} Count(\omega)}{\sum_{\omega \models C=c} Count(\omega)}$$

with pseudo-counts (Cromwell's rule)

Help System



- The domain of H is the set of all help pages.
  The observations are the words in the query.
- What probabilities are needed?
  What pseudo-counts and counts are used?
  What data can be used to learn from?

Constructing a belief network

To represent a domain in a belief network, we 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?
  Express this in terms of a directed graph, representing how each variable X_i is generated from its predecessors $X_{1:i-1}$.

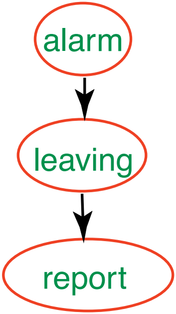
  The parents of X are variables on which X directly depends
  - X is independent of its non-descendants given its parents.
- How does the value of each variable depend on its parents?
  This is expressed in terms of the conditional probabilities.

Example: fire alarm belief network

Variables:

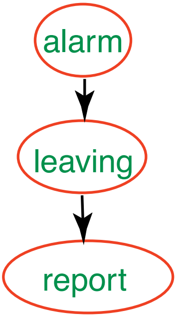
- Fire: there is a fire in the building
- Tampering: someone has been tampering with the fire alarm
- Smoke: what appears to be smoke is coming from an upstairs window
- Alarm: the fire alarm goes off
- Leaving: people are leaving the building *en masse*.
- Report: a colleague says that people are leaving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

Head-to-tail: Chain



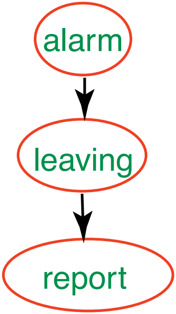
- *alarm* and *report* are

Head-to-tail: Chain



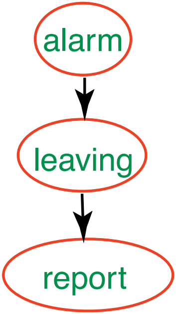
- *alarm* and *report* are dependent

Head-to-tail: Chain



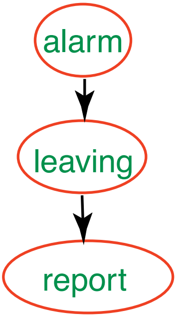
- *alarm* and *report* are dependent
- *alarm* and *report* are given *leaving*

Head-to-tail: Chain



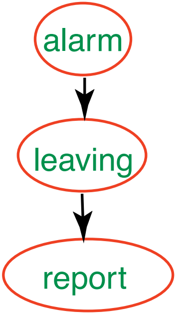
- *alarm* and *report* are dependent
- *alarm* and *report* are independent given *leaving*

Head-to-tail: 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*.

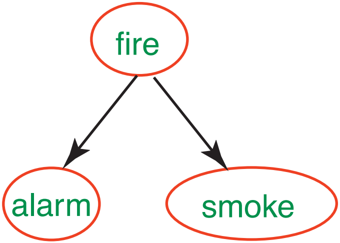
Head-to-tail: Chain



- *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*.

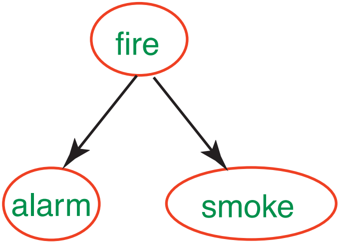
$$
\begin{aligned}
P(\text{report}, \text{alarm} \mid \text{leaving}) &= \frac{P(\text{report}, \text{alarm}, \text{leaving})}{P(\text{leaving})} \\
&= \frac{P(\text{alarm})P(\text{leaving} \mid \text{alarm})P(\text{report} \mid \text{leaving})}{P(\text{leaving})} && \text{net} \\
&= \frac{P(\text{alarm}, \text{leaving})}{P(\text{leaving})} P(\text{report} \mid \text{leaving}) && \text{product} \\
&= P(\text{alarm} \mid \text{leaving}) P(\text{report} \mid \text{leaving}) && \text{for } \perp\!\!\!\perp
\end{aligned}
$$

Tail-to-tail: Common ancestors



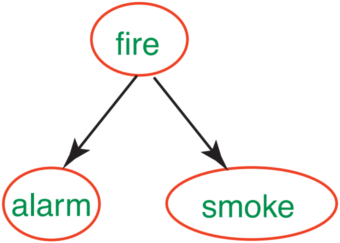
- *alarm* and *smoke* are

Tail-to-tail: Common ancestors



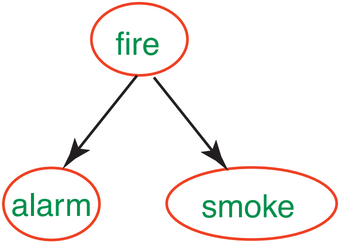
- *alarm* and *smoke* are dependent

Tail-to-tail: Common ancestors



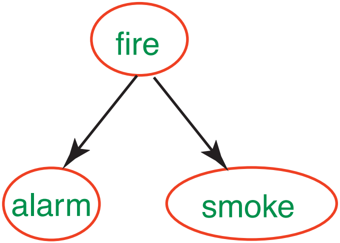
- *alarm* and *smoke* are dependent
- *alarm* and *smoke* are given *fire*

Tail-to-tail: Common ancestors



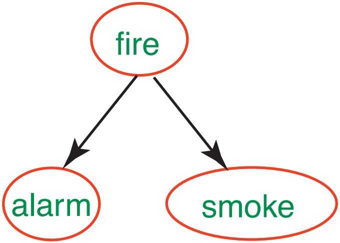
- *alarm* and *smoke* are dependent
- *alarm* and *smoke* are independent given *fire*

Tail-to-tail: 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*.

Tail-to-tail: Common ancestors

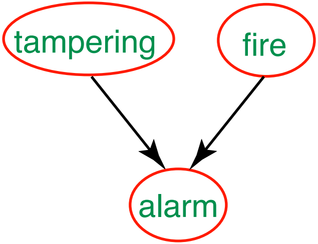


- *alarm* and *smoke* are dependent
- *alarm* and *smoke* are independent given *fire*
- Intuitively, *fire* can explain *alarm* and *smoke*; learning one can affect the other by changing your belief in *fire*.

$$\text{smoke} \perp\!\!\!\perp \text{alarm} \mid \text{fire}$$

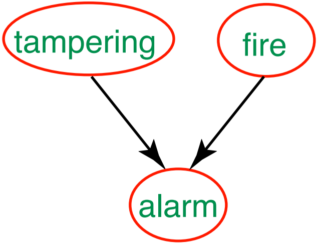
$$\begin{aligned}
P(\text{smoke}, \text{alarm} \mid \text{fire}) &= \frac{P(\text{smoke}, \text{alarm}, \text{fire})}{P(\text{fire})} \\
&= \frac{P(\text{fire})P(\text{alarm} \mid \text{fire})P(\text{smoke} \mid \text{fire})}{P(\text{fire})} \quad \text{net} \\
&= P(\text{alarm} \mid \text{fire})P(\text{smoke} \mid \text{fire}) \qquad \text{for } \perp\!\!\!\perp
\end{aligned}$$

Head-to-head: Common descendants



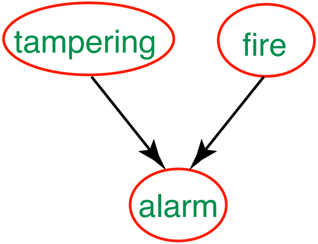
- *tampering* and *fire* are

Head-to-head: Common descendants



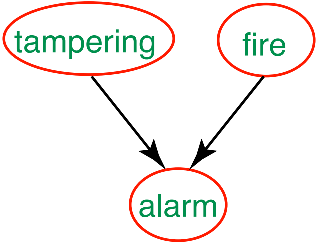
- *tampering* and *fire* are independent

Head-to-head: Common descendants



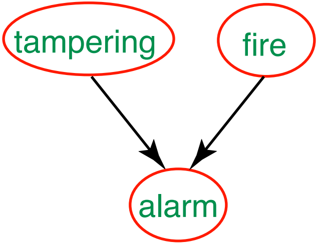
- *tampering* and *fire* are independent
- *tampering* and *fire* are given *alarm*

Head-to-head: Common descendants



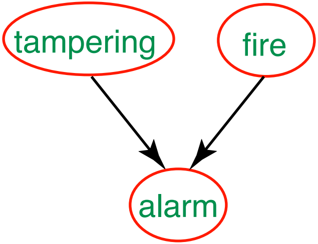
- *tampering* and *fire* are independent
- *tampering* and *fire* are dependent given *alarm*

Head-to-head: Common descendants



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

Head-to-head: Common descendants



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

$$P(\text{fi} = 1 \mid \text{am} = 1) > P(\text{fi} = 1 \mid \text{am} = 1 \wedge \text{tg} = 1)$$

$$\begin{aligned}
\text{for} \quad & P(\text{tg} = 0) = 0.9 \qquad P(\text{fi} = 0) = 0.9 \\
& P(\text{am} = 1 \mid \text{tg} = 1 \wedge \text{fi} = 1) = 0.95 \\
& P(\text{am} = 1 \mid \text{tg} = 1 \wedge \text{fi} = 0) = 0.5 \\
& P(\text{am} = 1 \mid \text{tg} = 0 \wedge \text{fi} = 1) = 0.9 \\
& P(\text{am} = 1 \mid \text{tg} = 0 \wedge \text{fi} = 0) = 0.1
\end{aligned}$$

$P(\text{fi} = 1|\text{am} = 1) \approx 0.418$

$$P(\text{fi} = 1|\text{am} = 1) = \frac{P(\text{am} = 1|\text{fi} = 1)P(\text{fi} = 1)}{P(\text{am} = 1)} \qquad \text{Bayes}$$

$$P(\text{am} = 1|\text{fi} = 1) = \sum_{tg} \underbrace{P(\text{am} = 1, tg|\text{fi} = 1)}_{P(\text{am} = 1|tg, \text{fi} = 1)\,\underbrace{P(tg|\text{fi} = 1)}_{P(tg)}} \qquad \text{sum}$$

$$\text{product}$$

$$\text{net}$$

$$P(\text{am} = 1) = \sum_{tg} \sum_{fi} \underbrace{P(\text{am} = 1, tg, fi)}_{P(tg)P(fi)P(\text{am} = 1|tg, fi)} \qquad \text{sum}$$

$$\text{net}$$

$P(\mathrm{fi}=1|\mathrm{am}=1,\mathrm{tg}=1) \approx 0.174$

$$P(\mathrm{fi}=1|\mathrm{am}=1,\mathrm{tg}=1) \;=\; \frac{P(\mathrm{am}=1|\mathrm{fi}=1,\mathrm{tg}=1)\,\overbrace{P(\mathrm{fi}=1|\mathrm{tg}=1)}^{P(\mathrm{fi}=1)\quad \text{net}}}{P(\mathrm{am}=1|\mathrm{tg}=1)}$$

Bayes

$$P(\mathrm{am}=1|\mathrm{tg}=1) \;=\; \sum_{fi} \underbrace{P(\mathrm{am}=1, fi|\mathrm{tg}=1)}_{P(\mathrm{am}=1|fi,\mathrm{tg}=1)\,\underbrace{P(fi|\mathrm{tg}=1)}_{P(fi)\quad \text{net}}\quad \text{product}} \qquad \text{sum}$$

Conditional independence via d-separation

Given disjoint sets A, B, C of nodes (variables).

When are the variables in A independent of those in B given C?

When A is d-separated from B by C
— i.e., all paths from A to B are C-blocked, where

a path from A to B is *C-blocked* if it has a node Z s.t.

(i) Z is in C, and the arrows on the path meet head-to-tail or tail-to-tail at Z

or

(ii) neither Z nor any of its descendants are in C, and the arrows on the path meet head-to-head at Z.

Conditional independence via d-separation

Given disjoint sets A, B, C of nodes (variables).

When are the variables in A independent of those in B given C?

When A is d-separated from B by C
— i.e., all paths from A to B are C-blocked, where

a path from A to B is *C-blocked* if it has a node Z s.t.

(i) Z is in C, and the arrows on the path meet head-to-tail or tail-to-tail at Z

or

(ii) neither Z nor any of its descendants are in C, and the arrows on the path meet head-to-head at Z.

Fact. If A is d-separated from B by C, then the variables in A are independent of those in B given C (for all network probabilities).

We can make the net undirected with disconnected = d-separate.

Understanding conditional independence

From non-implications

$$
\begin{array}{rcll}
A \perp\!\!\!\perp B & \not\Longrightarrow & A \perp\!\!\!\perp B \mid C & \text{(head-to-head)} \\
A \perp\!\!\!\perp B \mid C & \not\Longrightarrow & A \perp\!\!\!\perp B & \text{(head-to-tail, tail-to-tail)}
\end{array}
$$

to

Graphoid axioms (Pearl & Paz)

$A \perp\!\!\!\perp B \mid C$ as "C intercepts all paths from A to B"

(i) $A \perp\!\!\!\perp B \mid C$ implies $B \perp\!\!\!\perp A \mid C$

(ii) $A \perp\!\!\!\perp B, B' \mid C$ implies $A \perp\!\!\!\perp B \mid C$

(iii) $A \perp\!\!\!\perp B, B' \mid C$ implies $A \perp\!\!\!\perp B \mid B', C$

(iv) $A \perp\!\!\!\perp B \mid B', C$ and $A \perp\!\!\!\perp B' \mid C$ implies $A \perp\!\!\!\perp B, B' \mid C$

(v) $A \perp\!\!\!\perp B \mid B', C$ and $A \perp\!\!\!\perp B' \mid B, C$ implies $A \perp\!\!\!\perp B, B' \mid C$