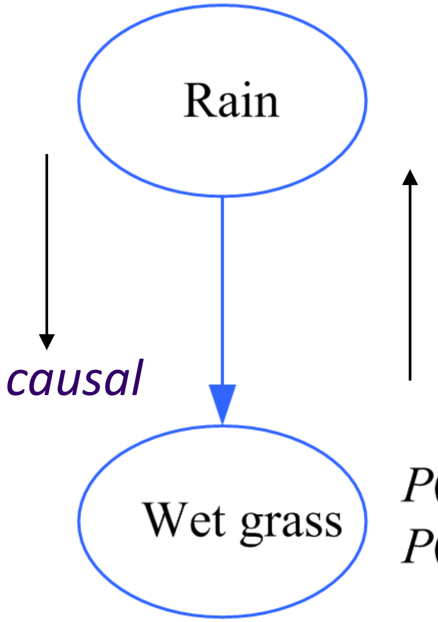
CHAPTER 14:

Graphical Models

Graphical Models

- Aka Bayesian networks, probabilistic networks
- Nodes are hypotheses (random vars) and the probabilities corresponds to our belief in the truth of the hypothesis
- Arcs are direct influences between hypotheses
- The structure is represented as a directed acyclic graph (DAG)
- The parameters are the conditional probabilities in the arcs (Pearl, 1988, 2000; Jensen, 1996; Lauritzen, 1996)

Causes and Bayes' Rule



$P(R)=0.4$

diagnostic

$P(W \mid R)=0.9$
$P(W \mid \sim R)=0.2$

Diagnostic inference:
Knowing that the grass is wet, what is the probability that rain is the cause?

$$P(R \mid W) = \frac{P(W \mid R)P(R)}{P(W)}$$

$$= \frac{P(W \mid R)P(R)}{P(W \mid R)P(R) + P(W \mid \sim R)P(\sim R)}$$

$$= \frac{0.9 \times 0.4}{0.9 \times 0.4 + 0.2 \times 0.6} = 0.75$$

Conditional Independence

- X and Y are independent if

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

- X and Y are conditionally independent given Z if

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

  or

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

- Three canonical cases: Head-to-tail, Tail-to-tail, head-to-head

Naïve Bayes

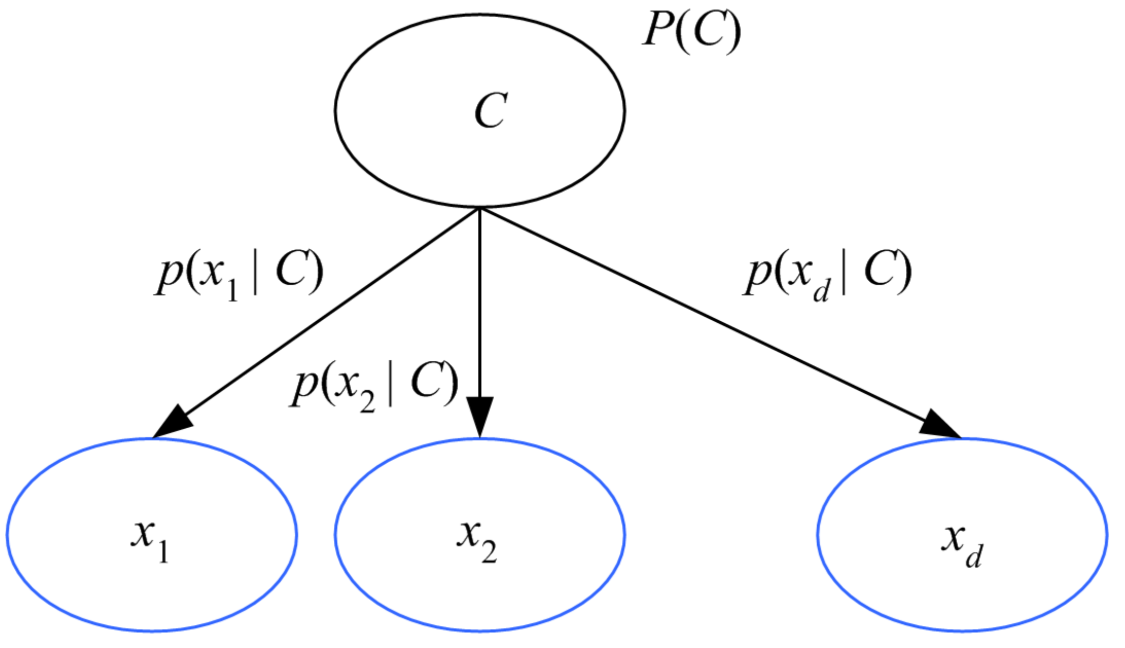
- Assumption: Individual attributes are independent given the class.

  More in the class

Example: Weather data

Outlook	Temp	Humidity	Windy	Play?
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Naive Bayes' Classifier



Given C, x_j are independent:

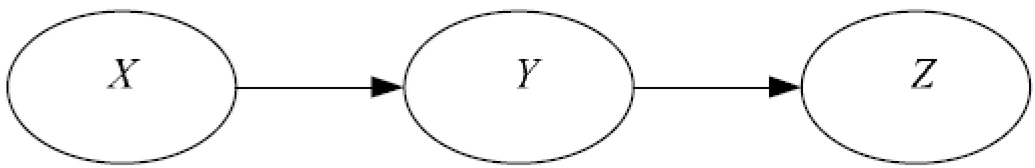
$$p(\boldsymbol{x}|C) = p(x_1|C)\, p(x_2|C) \ldots p(x_d|C)$$

Bayesian Network

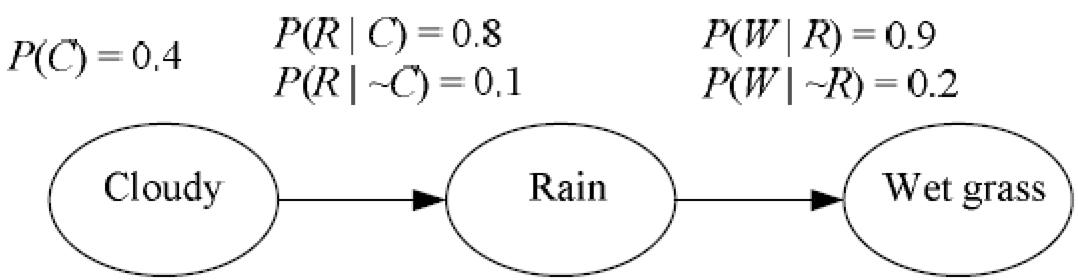
- Captures causality
- Structure represents conditional independent relationship
  - Nodes: Variables
  - Arc: causality
- Product of conditional independpence gives a joint distribution.
  - How it is better ?

Case 1: Head-to-Head

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



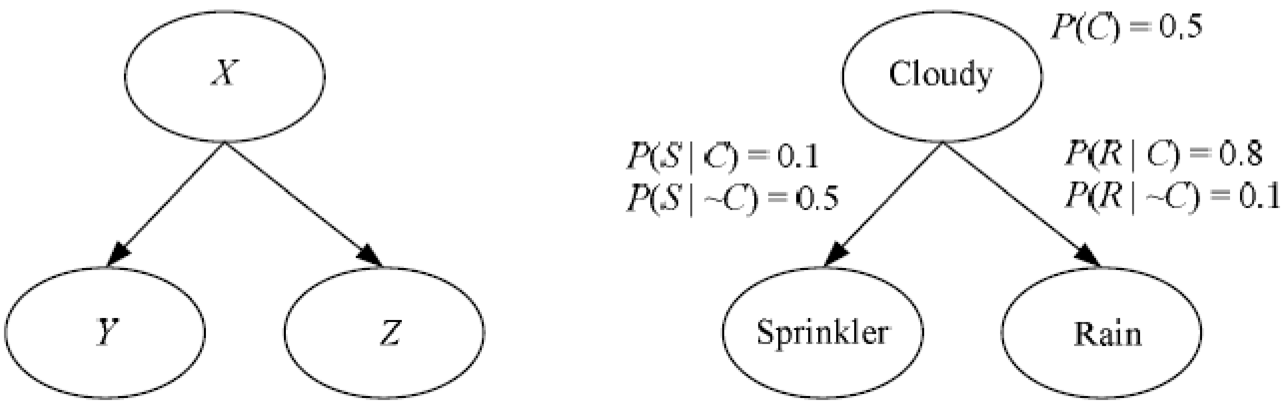
(a) Model



- $P(W|C)=P(W|R)P(R|C)+P(W|\sim R)P(\sim R|C)$

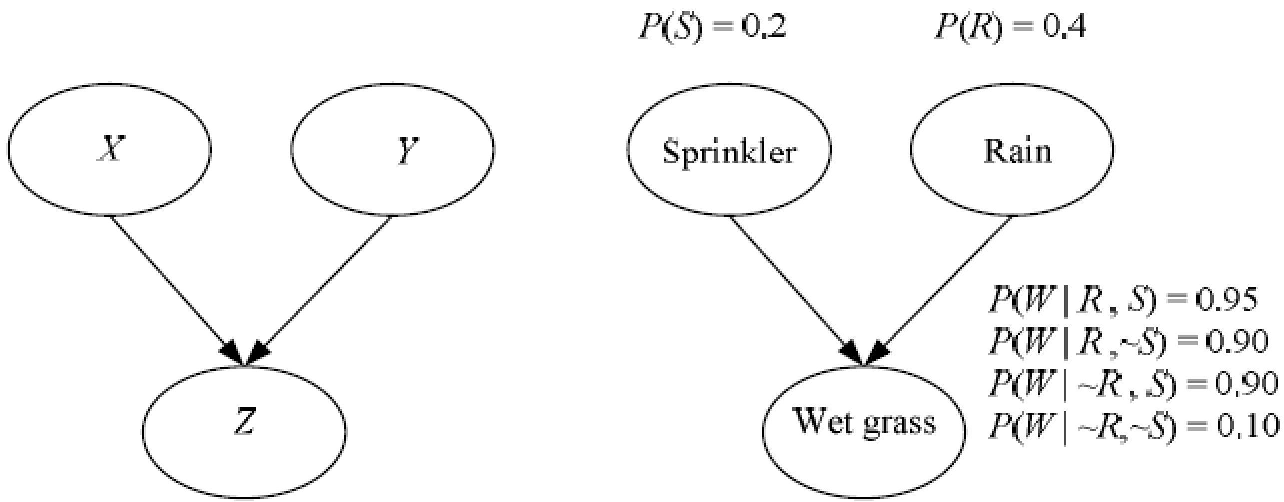
Case 2: Tail-to-Tail

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



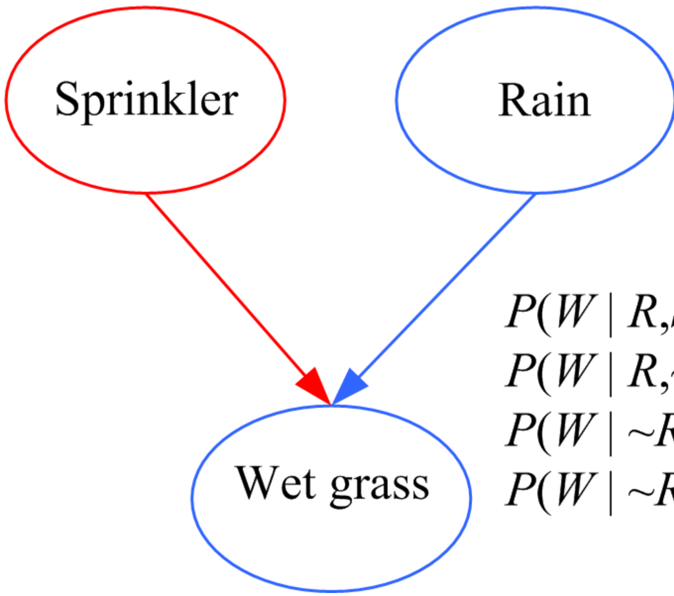
Case 3: Head-to-Head

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



Causal vs Diagnostic Inference

$P(S)=0.2$ $\quad$ $P(R)=0.4$



$P(W \mid R,S)=0.95$
$P(W \mid R,\sim S)=0.90$
$P(W \mid \sim R,S)=0.90$
$P(W \mid \sim R,\sim S)=0.10$

Causal inference: If the sprinkler is on, what is the probability that the grass is wet?

$P(W|S) = P(W|R,S)\, P(R|S) + P(W|\sim R,S)\, P(\sim R|S)$

$= P(W|R,S)\, P(R) + P(W|\sim R,S)\, P(\sim R)$

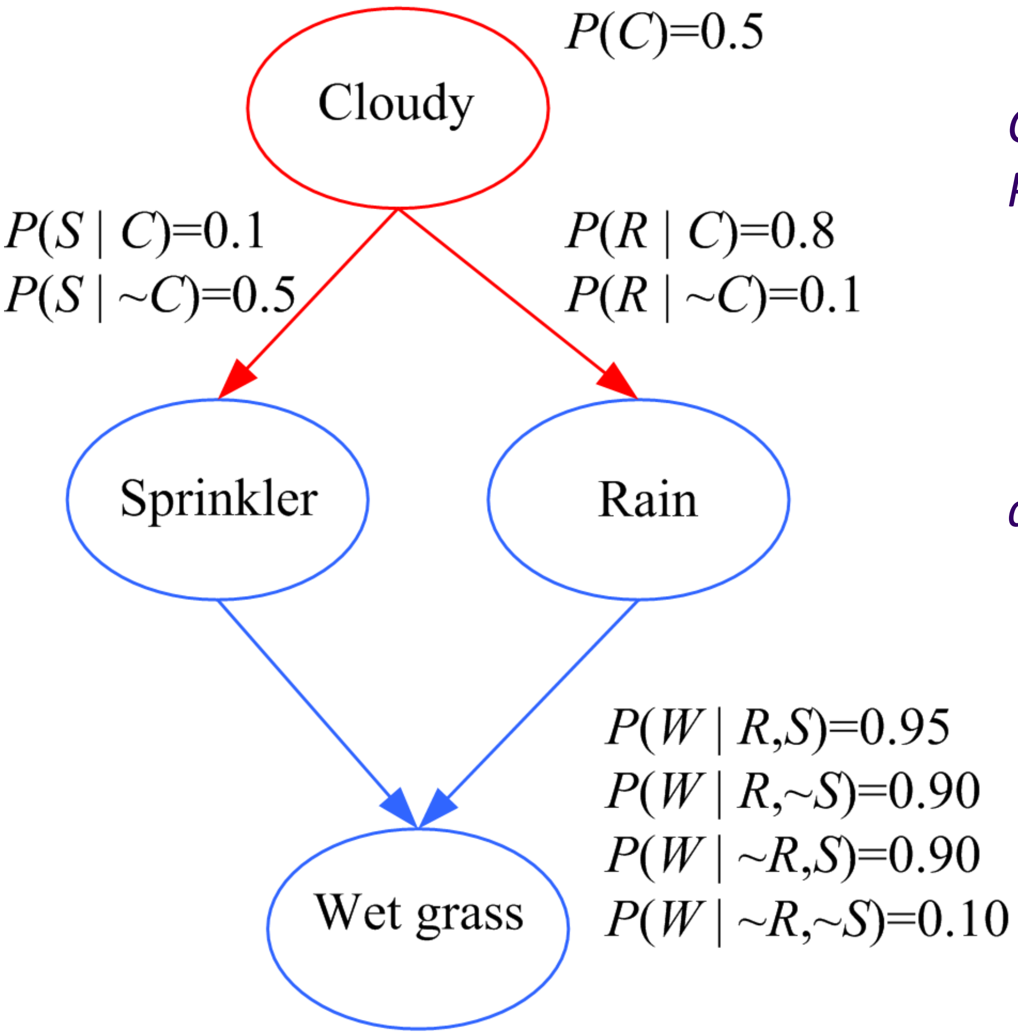
$= 0.95\ 0.4 + 0.9\ 0.6 = 0.92$

Diagnostic inference: If the grass is wet, what is the probability that the sprinkler is on? $P(S|W) = 0.35 > 0.2\ P(S)$

$P(S|R,W) = 0.21$

Explaining away: Knowing that it has rained decreases the probability that the sprinkler is on.

Causes



Causal inference:

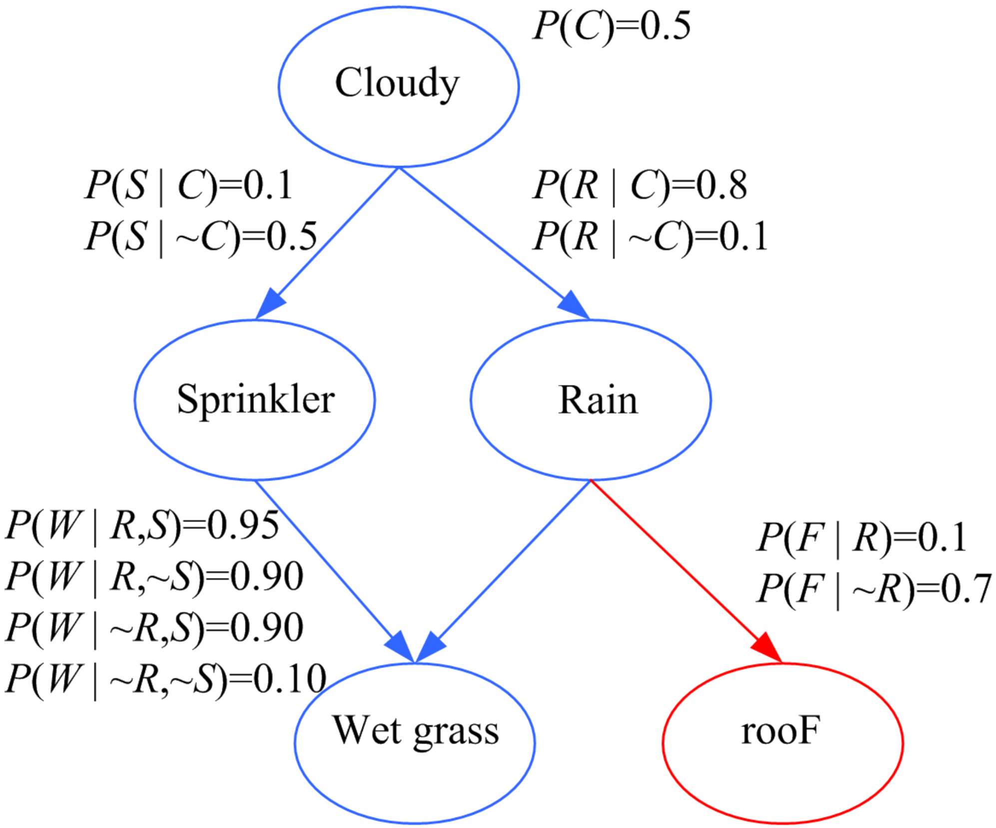
$$P(W|C) = P(W|R,S)\,P(R,S|C) + P(W|\sim R,S)\,P(\sim R,S|C) + P(W|R,\sim S)\,P(R,\sim S|C) + P(W|\sim R,\sim S)\,P(\sim R,\sim S|C)$$

and use the fact that

$$P(R,S|C) = P(R|C)\,P(S|C)$$

Diagnostic: $P(C|W) = ?$

Exploiting the Local Structure

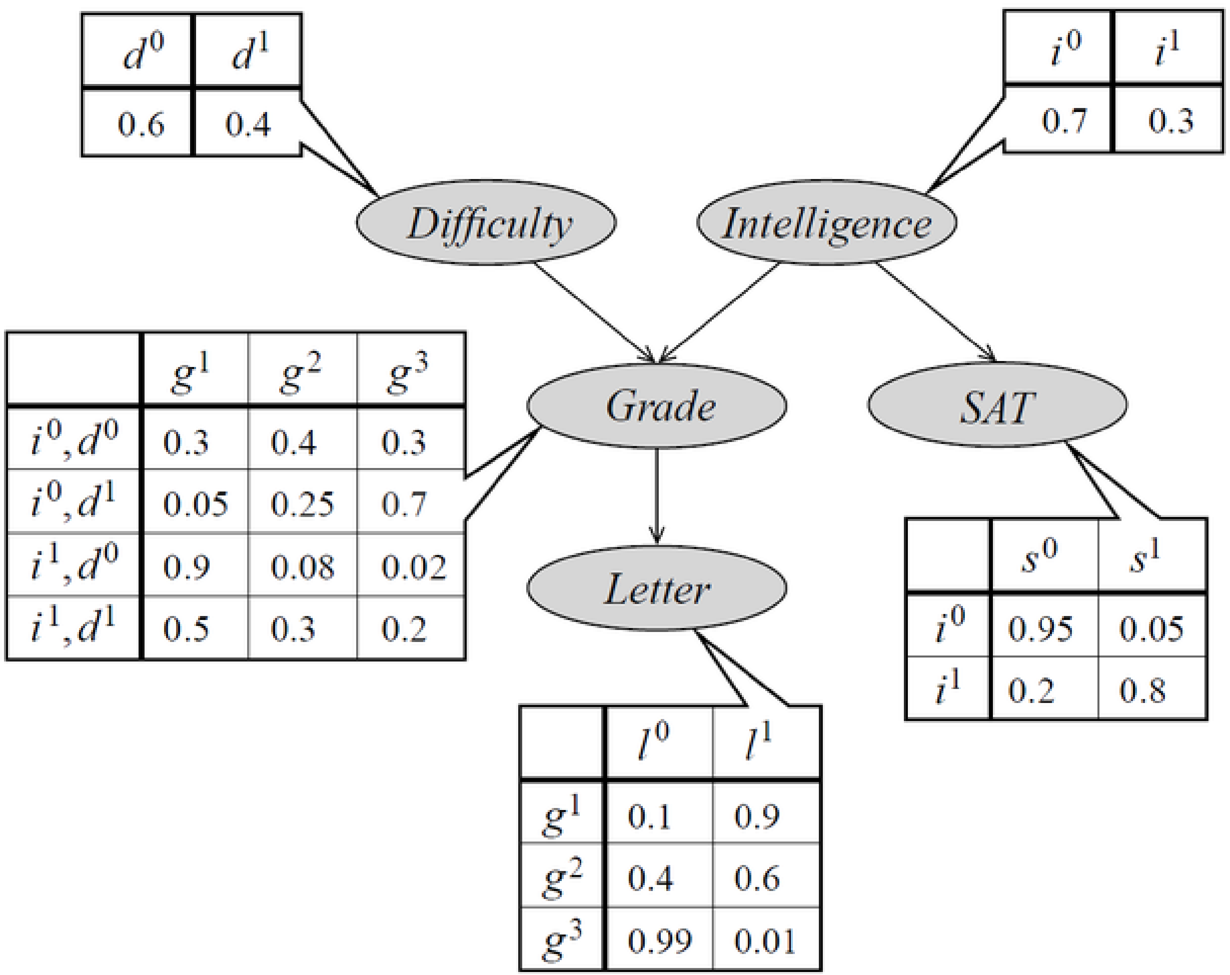


$P(F \mid C) = ?$

$$P(C,S,R,W,F) = P(C)P(S \mid C)P(R \mid C)P(W \mid S,R)P(F \mid R)$$

$$P(X_1, \ldots X_d) = \prod_{i=1}^{d} P(X_i \mid \text{parents}(X_i))$$

Bayesian Network: Another example



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

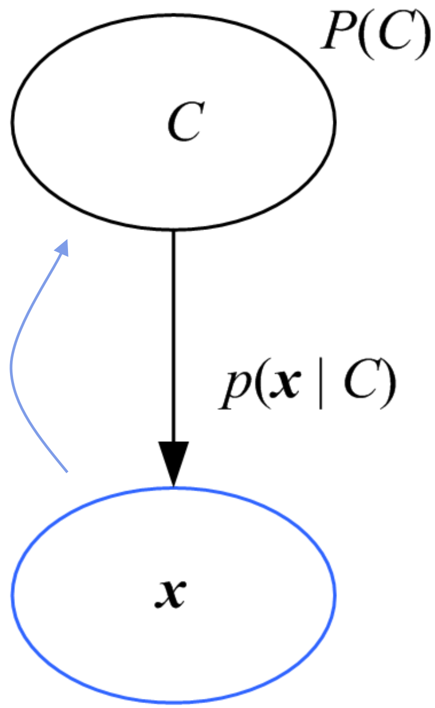
	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Bayesian Network: Classification

diagnostic

$P\,(C \mid x\,)$



Bayes' rule inverts the arc:

$$P(C \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid C)P(C)}{p(\mathbf{x})}$$

How to learn a Graphical Model

- Learning the conditional probabilities, either as tables (for discrete case with small number of parents), or as parametric functions

- Learning the structure of the graph: Doing a state-space search over a score function that uses both goodness of fit to data and some measure of complexity