Remember: Next week is Fall vacation

CHAPTER 14:

GRAPHİCAL MODELS

Joint probabilites

- Recap from chapter 3

	Oranges	Apples	Peaches	$P(Y)$
Good	0.16	0.28	0.12	0.56
Bad	0.12	0.04	0.28	0.44
$P(X)$	0.28	0.32	0.40	

sum rule $$p(X) = \sum_Y p(X,Y)$$

product rule $$p(X,Y) = p(Y|X)p(X)$$

Joint probabilites

- Bayes' rule

$$P(Y|X) = \frac{P(X,Y)}{P(X)}$$

$$= \frac{P(X|Y)P(Y)}{P(X)}$$

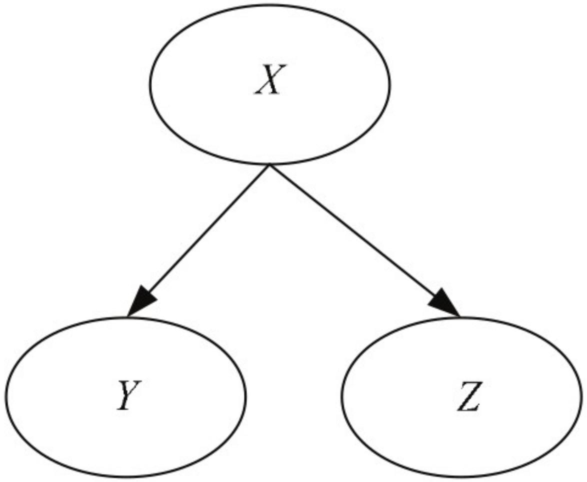
$$= \frac{P(X|Y)P(Y)}{\sum_Y P(X|Y)P(Y)}$$

Graphical Models

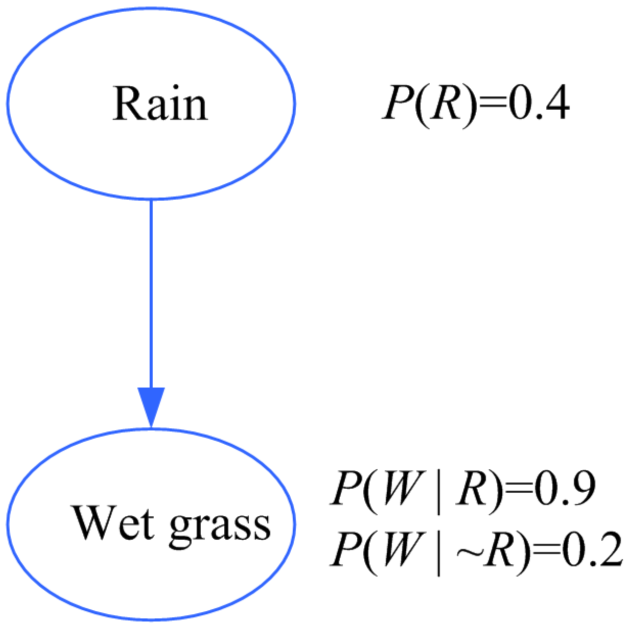
- Aka Bayesian networks, probabilistic graphical models, probabilistic networks
- A graphical representation of joint probability distributions
  - Influences
  - Independences

Graphical Models

- Nodes are random variables
- Arcs (edges) are direct influences between variables
- The structure is a directed acyclic graph (DAG)
- The parameters are the probability distributions



Simple example



Two random variables:

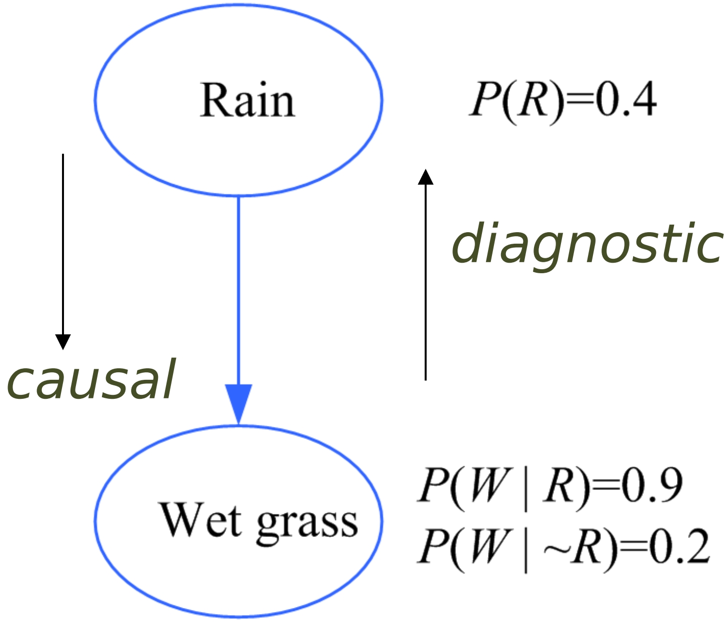
- *R: did it **Rain** last night?*
- *W: is the grass **Wet**?*

R “causes” W

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

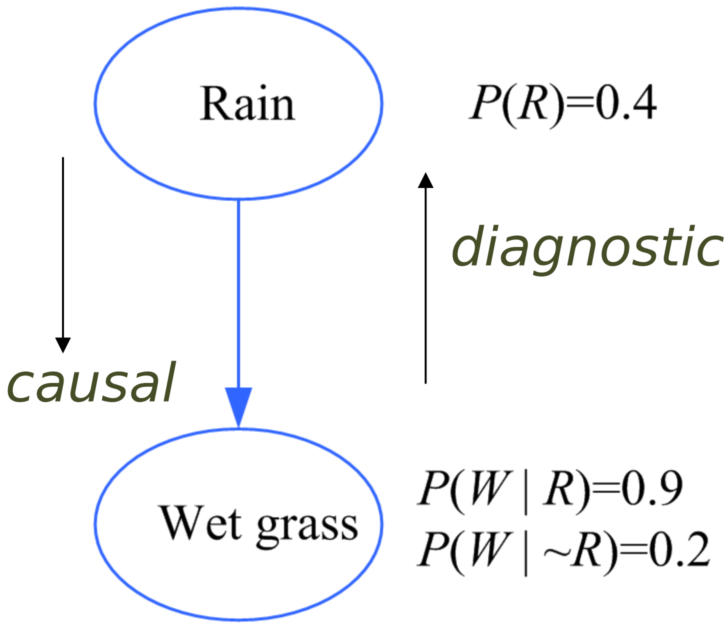
Note: the notation for binary variables in this chapter uses:
P(R) to mean P(Rain=True)
P(~R) to mean P(Rain=False)

Simple example



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

Causes and Bayes' Rule



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

$$
\begin{aligned}
P(R \mid W) &= \frac{P(W \mid R)P(R)}{P(W)} \\
&= \frac{P(W \mid R)P(R)}{P(W \mid R)P(R) + P(W \mid \sim R)P(\sim R)} \\
&= \frac{0.9 \times 0.4}{0.9 \times 0.4 + 0.2 \times 0.6} = 0.75
\end{aligned}
$$

Conditional Independence

- X and Y are independent if

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

Conditional Independence

- X and Y are independent if

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

- X and Y are <u>*conditionally*</u> independent given Z if

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

or

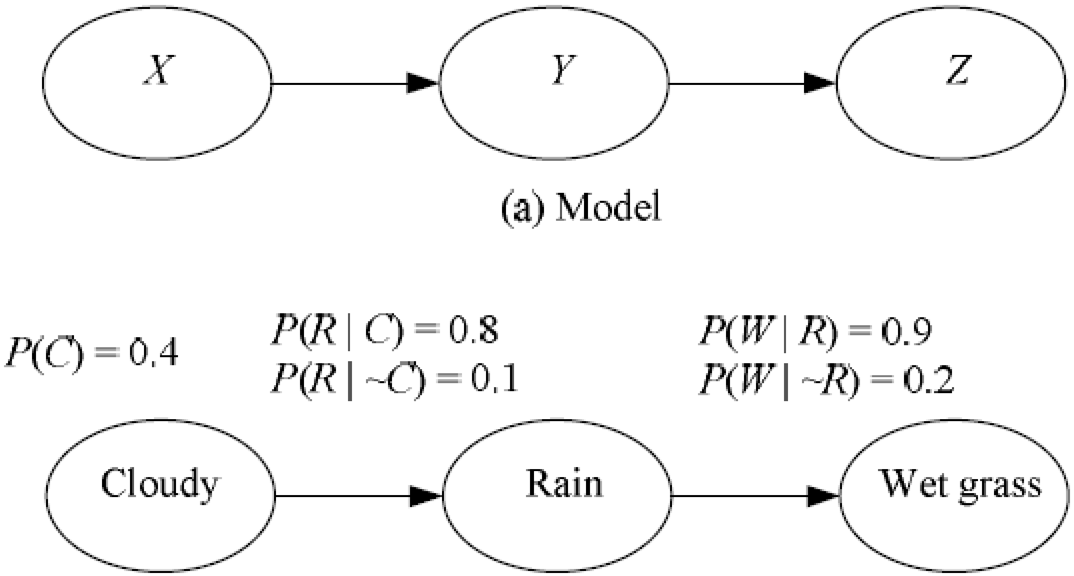
$$P(X|Y,Z)=P(X|Z)$$

Conditional Independence

- Graphical models show joint probability distributions
- Dependence and condititional independence can be seen directly from subgraphs of the model
- Three canonical cases: Head-to-tail, Tail-to-tail, head-to-head

Case 1: Head-to-Head

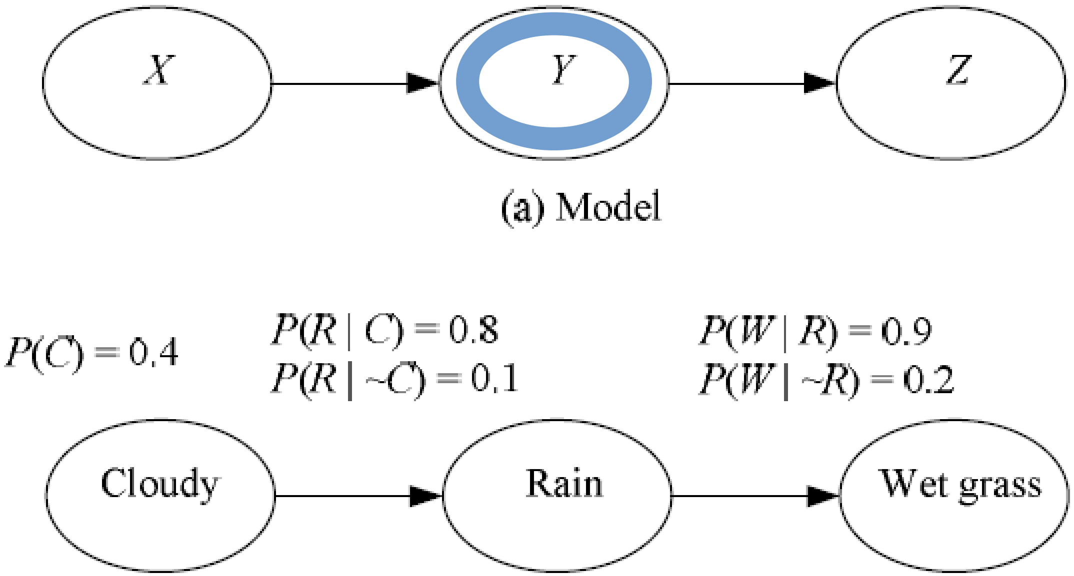
- $P(X,Y,Z)=P(X)P(Y|X)P(Z|Y)$



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

Case 1: Head-to-Head

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

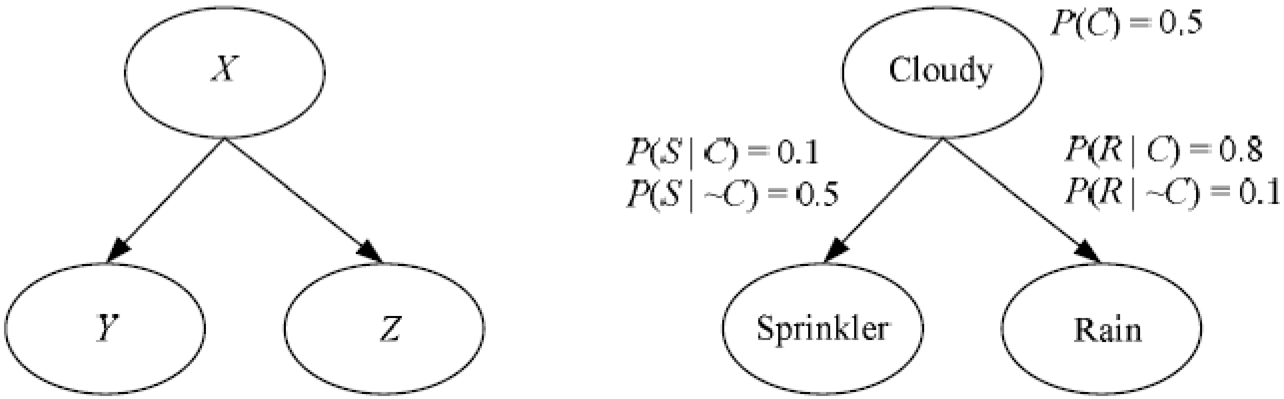


Observing Y *blocks* the path between X and Z: Makes X and Z independent

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

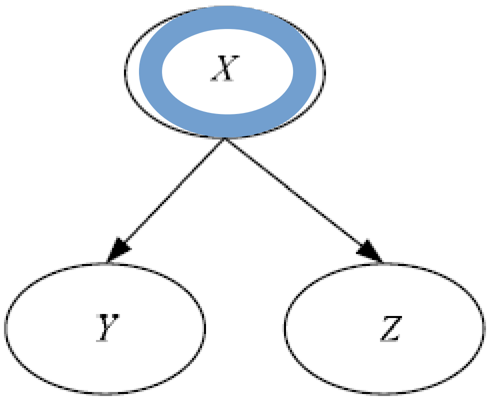
Case 2: Tail-to-Tail

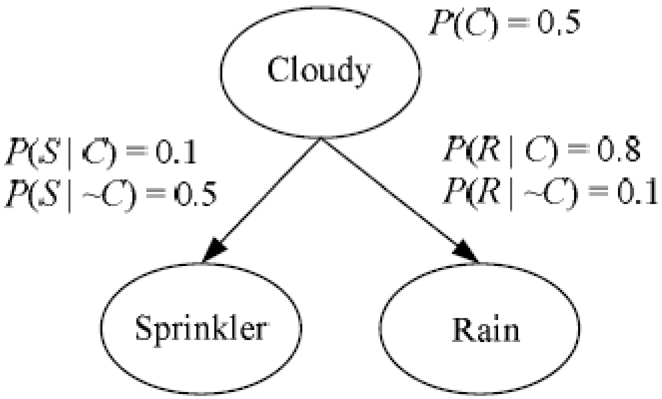
- $P(X,Y,Z)=P(X)P(Y|X)P(Z|X)$



Case 2: Tail-to-Tail

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

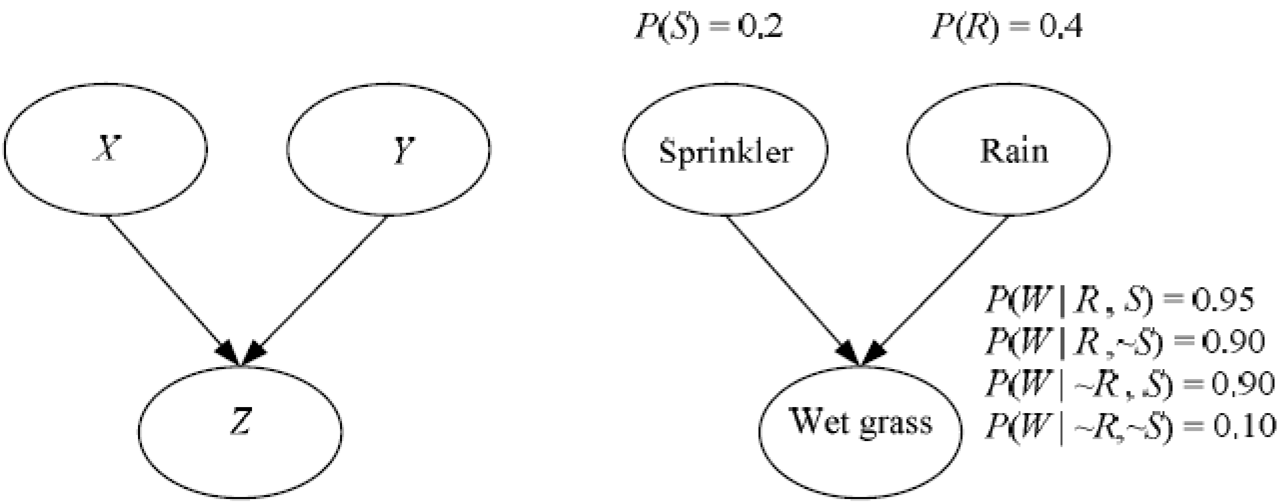




Observing X makes Y and Z independent

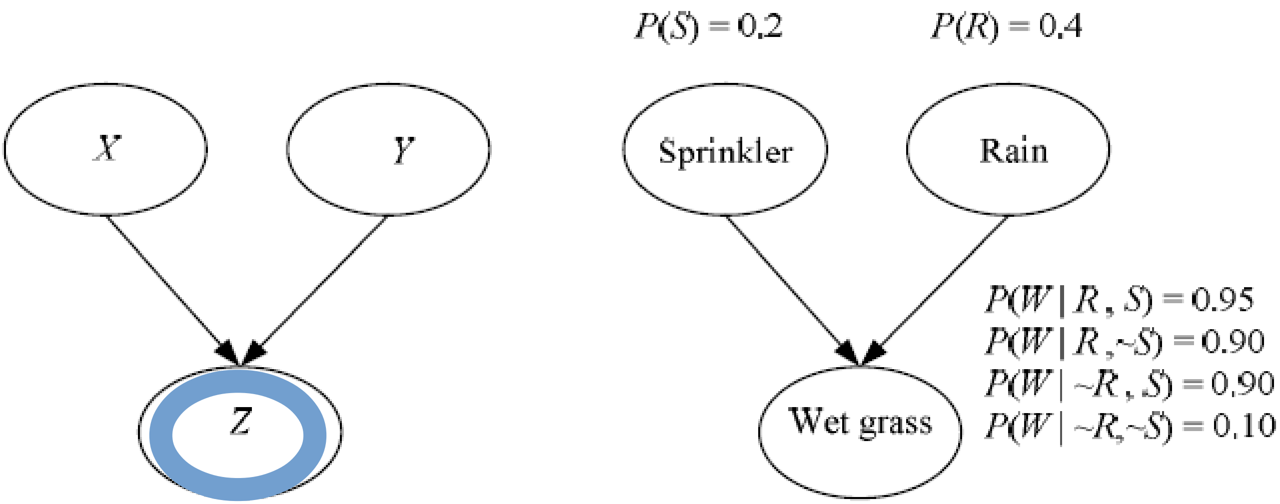
Case 3: Head-to-Head

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



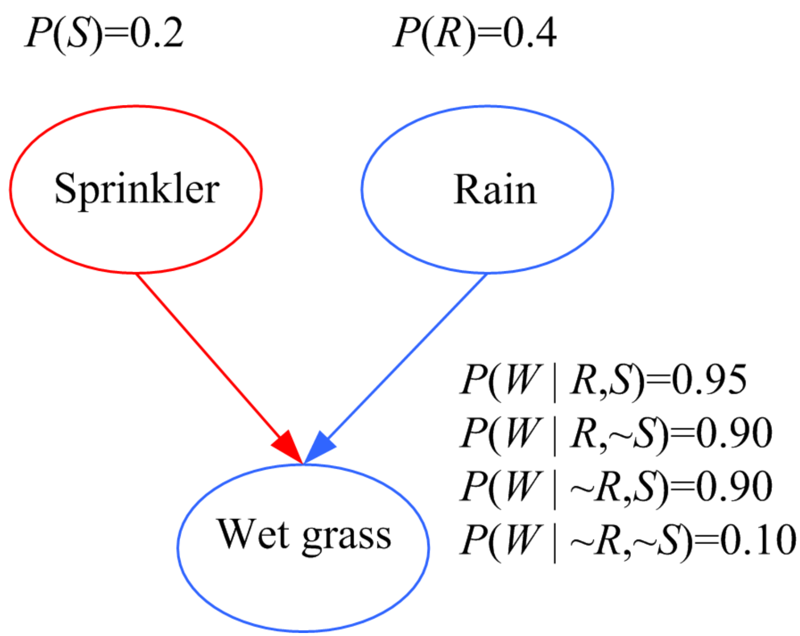
Case 3: Head-to-Head

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



X and Y are independent if Z is NOT observed!

Causal vs Diagnostic Inference



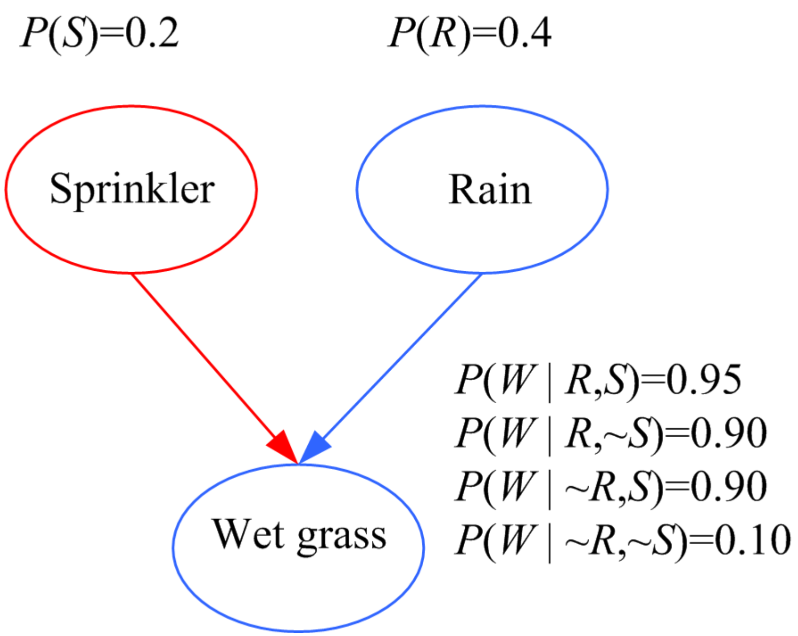
Causal inference: If the sprinkler is on, what is the probability that the grass is wet?

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

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

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

Causal vs Diagnostic Inference



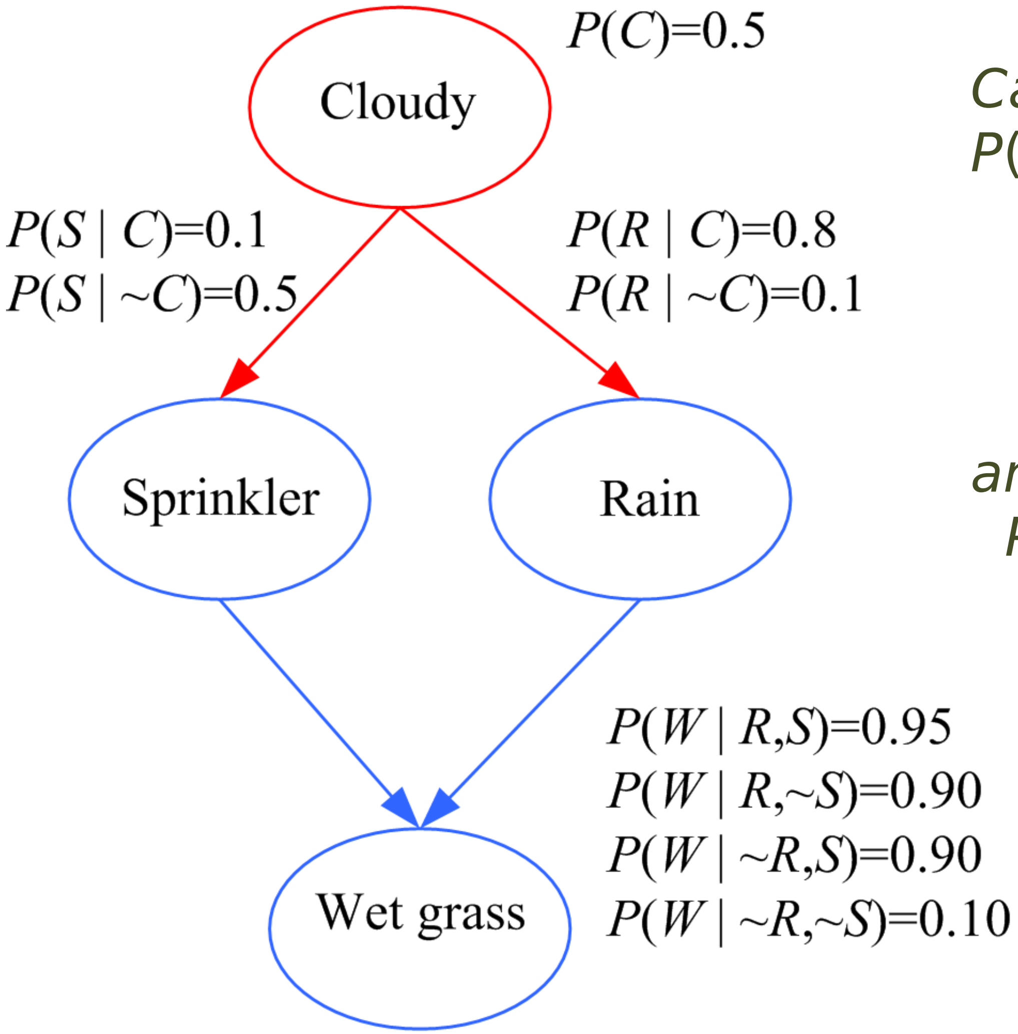
Causal inference: If the sprinkler is on, what is the probability that the grass is wet?

$P(W|S) = P(W|R,S)\ P(R|S) +$
$\quad P(W|\sim R,S)\ P(\sim R|S)$
$= P(W|R,S)\ P(R) +$
$\quad P(W|\sim R,S)\ P(\sim R)$
$= 0.95\ 0.4 + 0.9\ 0.6 = 0.92$

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.*

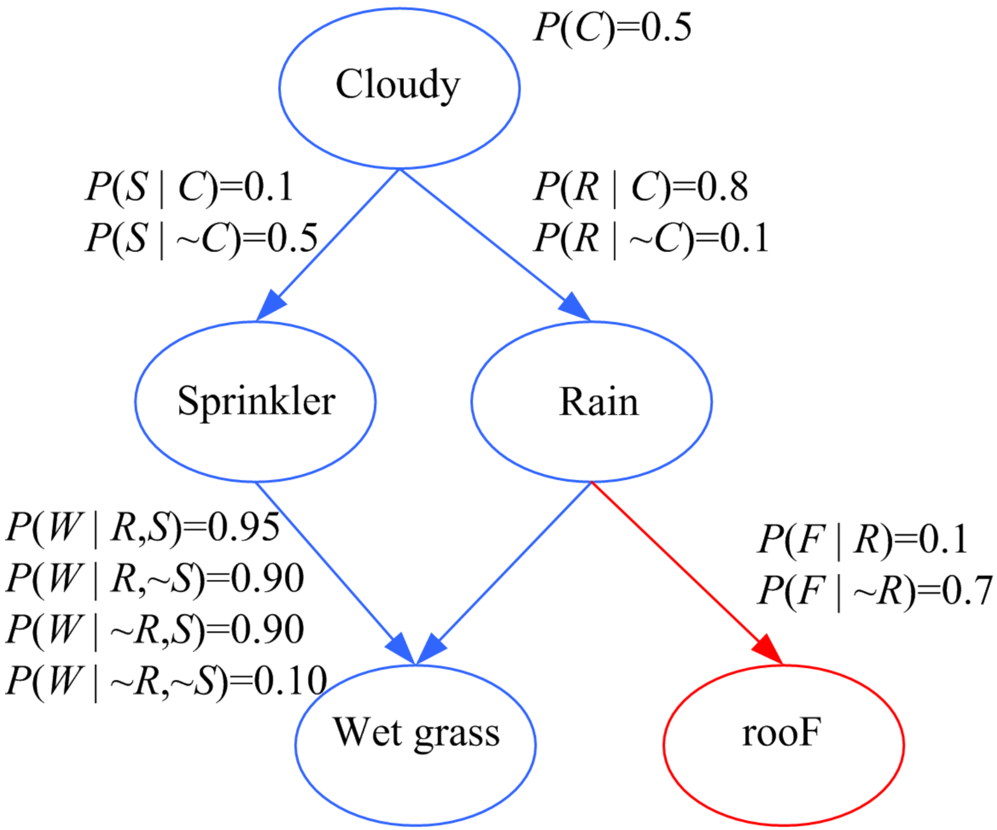
Combining graphs



Causal inference:
P(W|C) = P(W|R,S) P(R,S|C) +
P(W|~R,S) P(~R,S|C) +
P(W|R,~S) P(R,~S|C) +
P(W|~R,~S) P(~R,~S|C)

and use the fact that
P(R,S|C) = P(R|C) P(S|C)

Exploiting the Local Structure

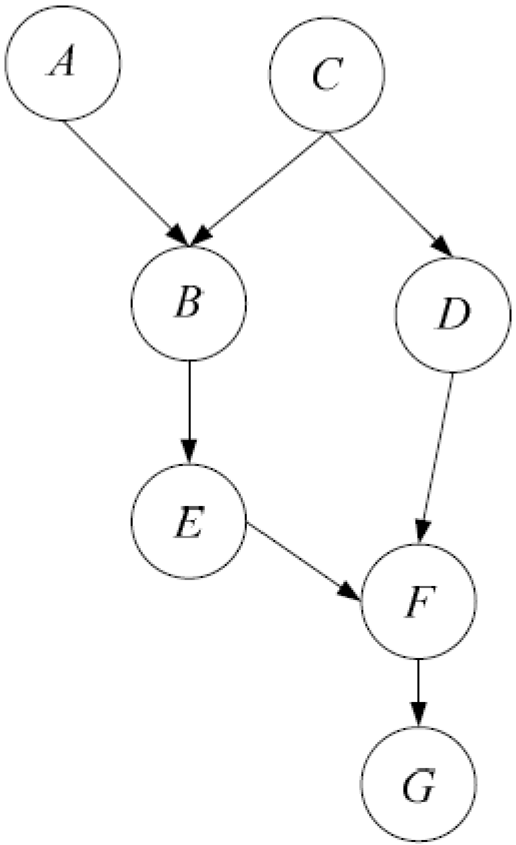


$$P(C,S,R,W,F) = P(C)P(S \mid C)P(R \mid C)P(W \mid S,R)P(F \mid R)$$

$$P(x_1,\ldots x_d) = \prod_{i=1}^{d} P(x_i \mid \text{parents}(x_i))$$

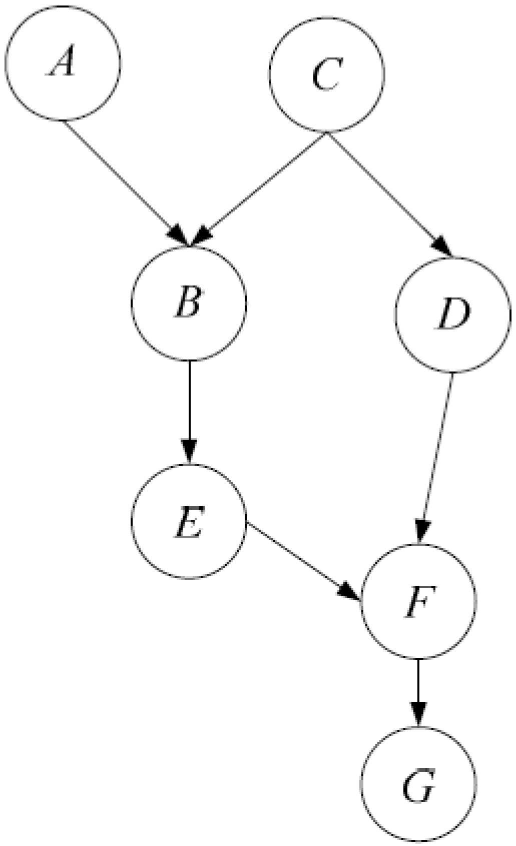
mini-exercise

- Write the joint distribution for this model (DAG)



mini-exercise

- Write the joint distribution for this model (DAG)



P(A)P(C)P(B|A,C)P(E|B)P(D|C)P(F|E,D)P(G|F)

mini-exercise

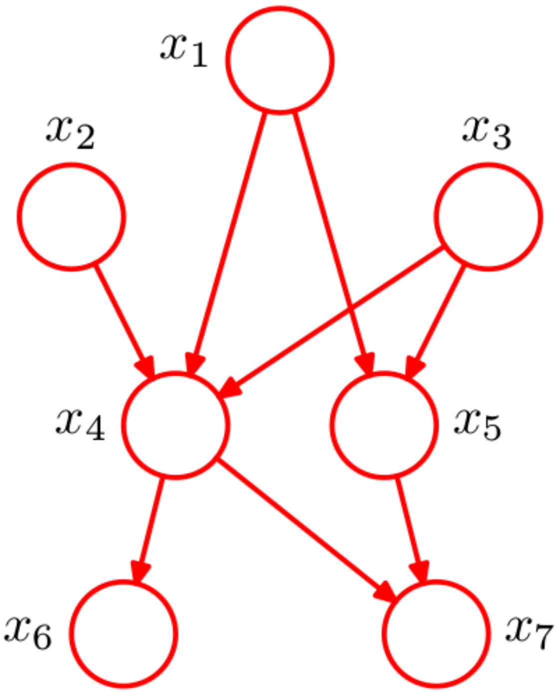
$$p(x_1)p(x_2)p(x_3)p(x_4|x_1,x_2,x_3)p(x_5|x_1,x_3)p(x_6|x_4)p(x_7|x_4,x_5)$$

- Draw the model (DAG) that describes this joint distribution

mini-exercise

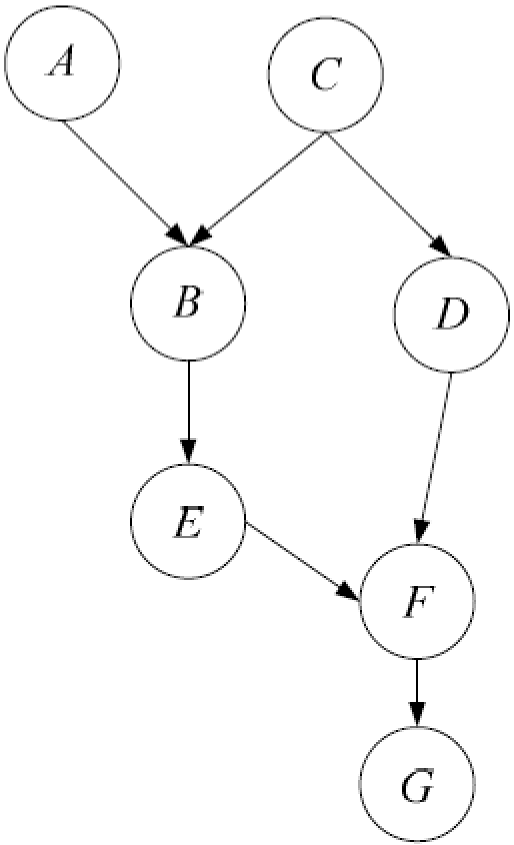
$$p(x_1)p(x_2)p(x_3)p(x_4|x_1,x_2,x_3)p(x_5|x_1,x_3)p(x_6|x_4)p(x_7|x_4,x_5)$$

- Draw the model (DAG) that describes this joint distribution



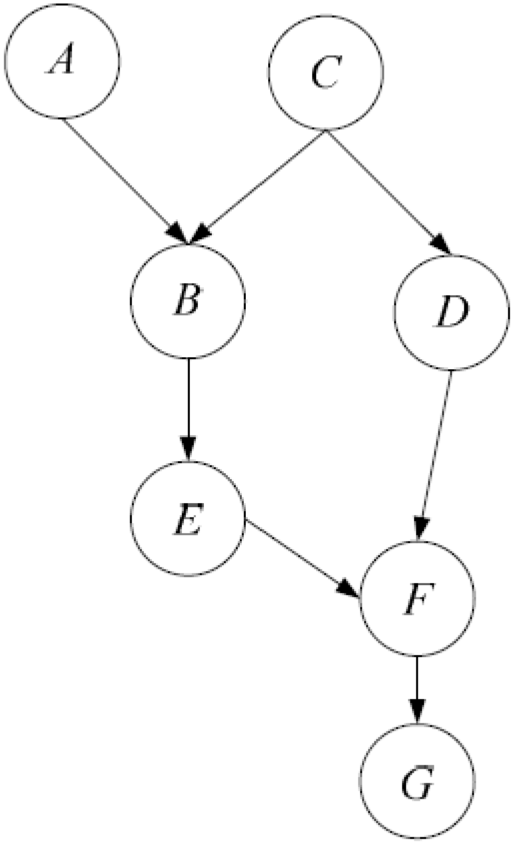
d-Separation

- A general way to describe if two (sets of) nodes are independent
- Two (sets of) nodes are d-separated if all paths between them are blocked



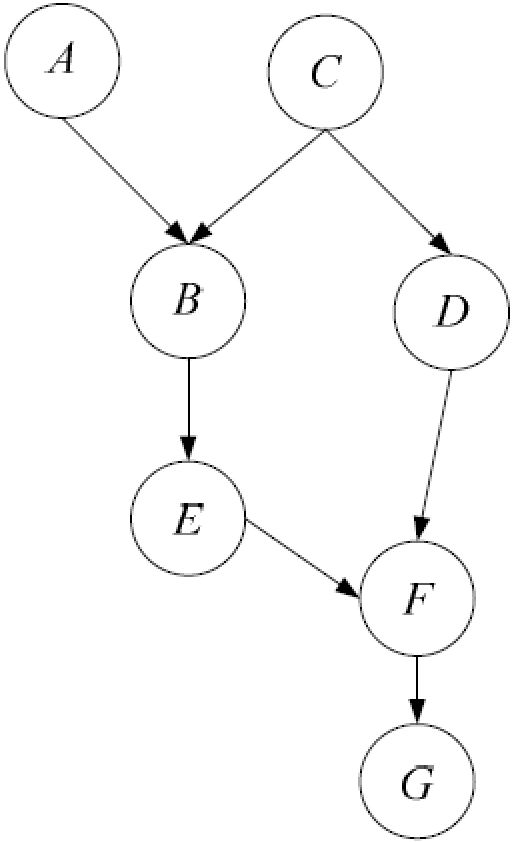
d-Separation

- A path from node *X* to node *Y* is blocked if
  a) Edges on the path meet head-to-tail (case 1) or tail-to-tail (case 2) and the node between is observed
  b) Edges on the path meet head-to-head (case 3) and neither that node nor any of its descendants are observed
- If all paths are blocked, *X* and *Y* are d-separated (conditionally independent) given the observations.



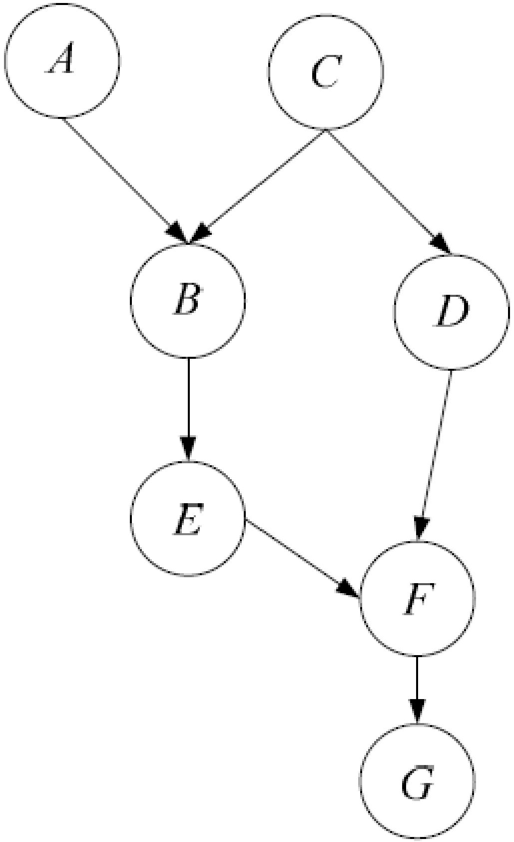
d-Separation: Are these paths blocked?

- BCDF given C?
- BEFG given F?
- BEFD given G?



d-Separation: Are these paths blocked?

- BCDF given C?
  -blocked
- BEFG given F?
  -blocked
- BEFD given G?
  -not blocked



Generative models

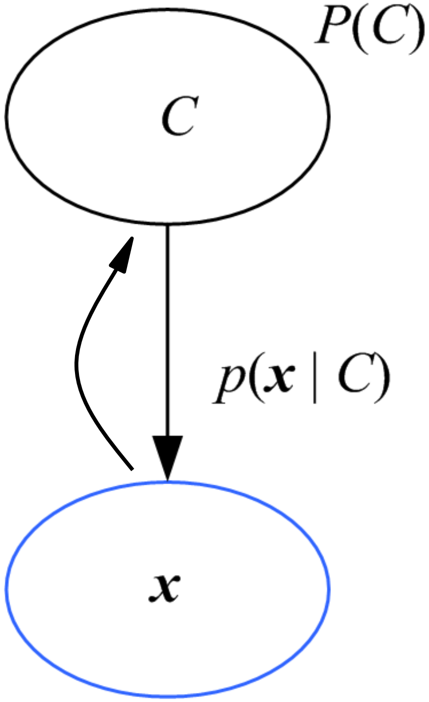
- Graphical models can represent the processes that generate data

- Another way to look at
  - Classification
  - Regression
  - Etc

Classification

Can be visualized as a graphical model:

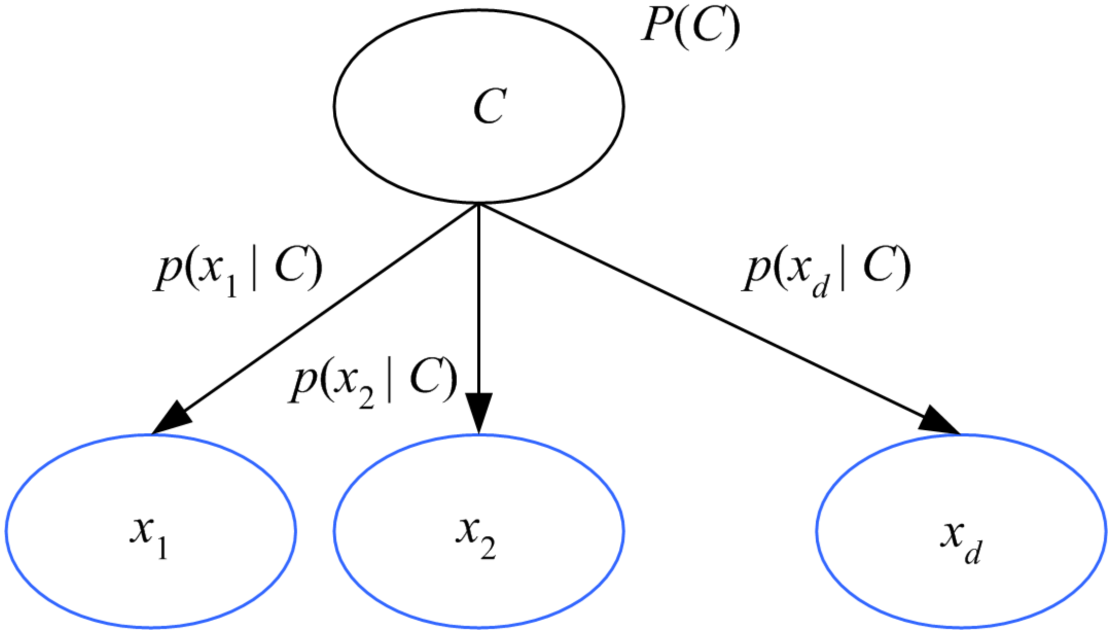
diagnostic
$P(C|x)$



Bayes' rule inverts the arc:

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

Naive Bayes' Classifier



Given C, x_j are independent:

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

Linear Regression

Linear regression as a graphical (Bayesian) model