Soft Computing – Probabilistic Reasoning

- Introduction to Bayesian Networks-

Prof. Matteo Matteucci – *matteo.matteucci@polimi.it*

Course Syllabus (Tentative)

Probability basics (fast and furious)

- Frequentists vs Bayesians
- Joint and Naive Distributions

Probabilistic graphical models

- Directed graphical models (Bayesian Networks)
- Conditional independence and d-separation
- Inference in directed graphical models

Dynamical graphical models

- Markov chains
- Hidden Markov models

Learning directed graphical models …

Beyond Independence …

We are thankful to the independence hypothesis because:

- It makes computation possible
- It yields optimal classifiers when satisfied
- It gives good enough generalization to our Bayes Classifiers

<u>Seldom satisfied in practice, as attributes are often correlated!</u>

To overcome this limitation we can describe a probability distribution via:

- *Conditional Independence Assumptions* that apply on subsets of them
- A set of conditional probabilities with their priors

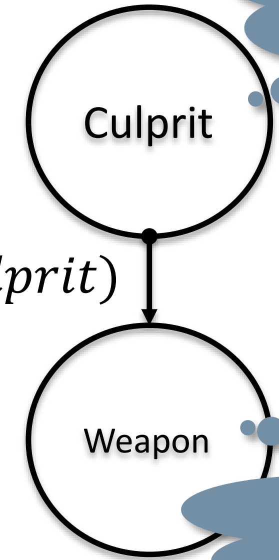
*<u>These models are often referred also as **Graphical Models**</u>*

Beyond Independence …

To overcome this limitation we can describe a probability distribution via:

- *Conditional Independence Assumptions* that apply on subsets of them
- A set of conditional probabilities with their priors

1 number to describe
$P(Culprit = Butler)$



$P(Culprit)$

$P(Weapon|Culprit)$

	Pistol	Knife	Poker
Cook	4%	52%	24%
Butler	16%	2%	2%

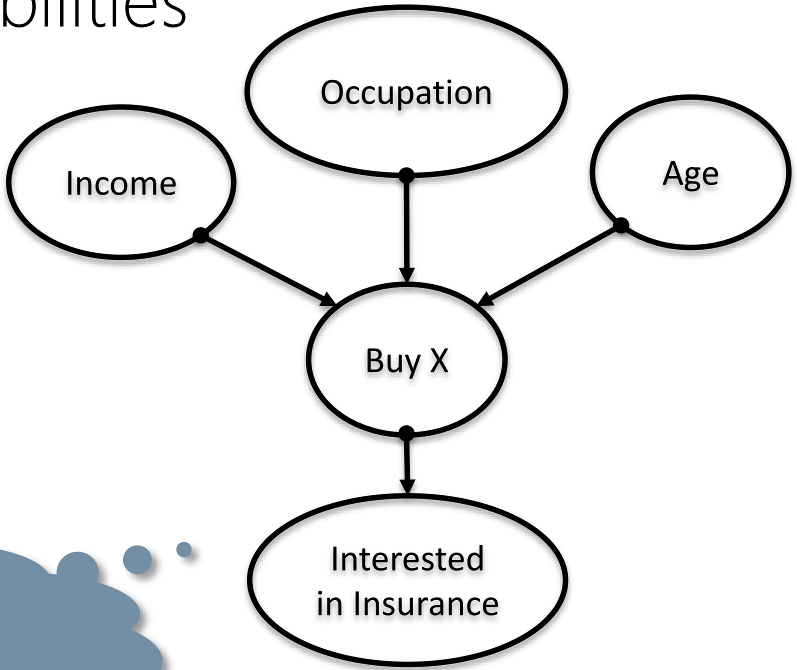
4 numbers to describe
$P(Weapon|Culprit = Butler)$
and $P(Weapon|Culprit = Cook)$

Bayesian Belief Networks

A *Bayesian Belief Networks*, or *Bayesian Network*, is a method to describe the joint probability distribution of a set of variables

Let $x_1, x_2, \dots, x_N$ be a set of variables a Bayesian Network can tell any combination of these probabilities

- Age, Occupation and Income determine if customer will buy this product.
- Given that customer buys product, whether there is interest in insurance is now independent of Age, Occupation, Income.



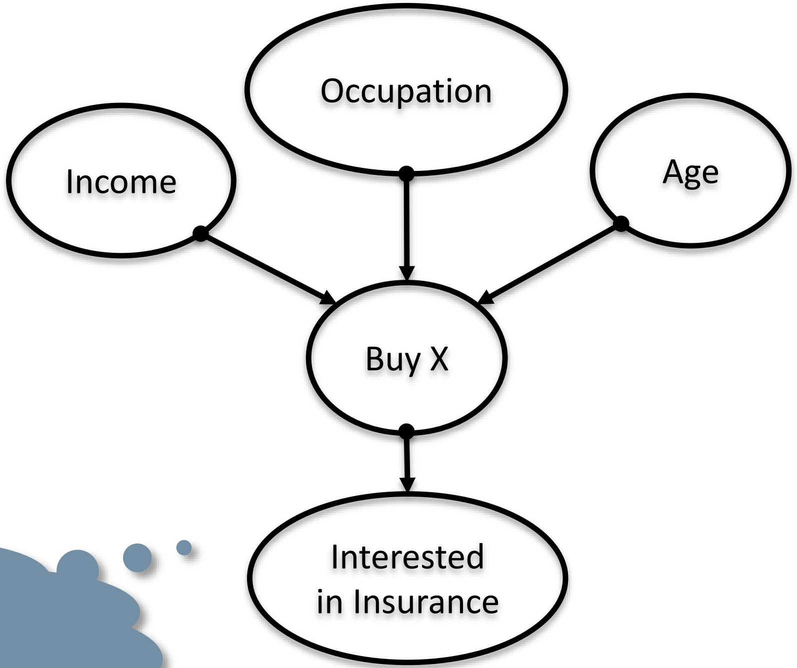
Some sort of independence should be there …

Bayesian Belief Networks

A *Bayesian Belief Networks*, or *Bayesian Network*, is a method to describe the joint probability distribution of a set of variables

Let $x_1, x_2, \dots, x_N$ be a set of variables a Bayesian Network can tells any combination probability.

- The full joint distribution would require $2^N - 1 = 2^5 - 1 = 31$ parameters
- To represent the probabilities in the network we need only the priors and conditionals $3 * 1 + 1 * 2^3 + 1 * 2^1 = 13$ parameters



Some sort of independence should be there …

Conditional Independence

We say X_1 is *Conditionally Independent* from X_2 given X_3 if the probability of X_1 is independent of X_2 given some knowledge about X_3:

$$P(X_1|X_2, X_3) = P(X_1|X_3)$$

The same can be said for a set of variables: X_1, X_2, X_3 is independent from Y_1, Y_2, Y_3 given Z_1, Z_2, Z_3:

$$P(X_1, X_2, X_3|Y_1, Y_2, Y_3, Z_1, Z_2, Z_3) = P(X_1, X_2, X_3|Z_1, Z_2, Z_3)$$

Note: there is a soubtle difference with respect to independence!

Conditional Independence Example (Part 1)

Martin and Norman toss the same coin. Let be A "Norman's outcome", and B "Martin's outcome". Assume the coin might be biased; in this case A and B are not independent: observing that B is Heads causes us to increase our belief in A being Heads.

$$P(A|B) \neq P(A)$$

Variables A and B are both dependent on C "The coin is biased towards Heads with probability θ". Once we know for C then any evidence about B cannot change our belief about A.

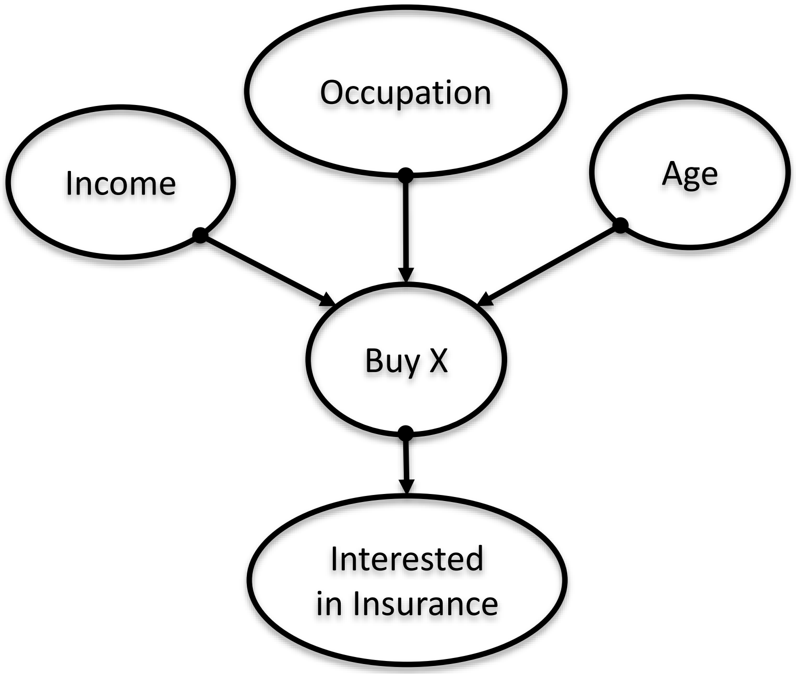
$$P(A|B, C) = P(A|C)$$

Bayesian Belief Network Ingredients

A *Bayesian Network* is a compact representation of the joint probability distribution via explicit indication of conditional independences:

- A Directed Acyclic Graph (DAG) where Nodes represent random variables and Edges represent direct influence
- Conditional Probability Distributions (CPD) for priors and "influenced" variables

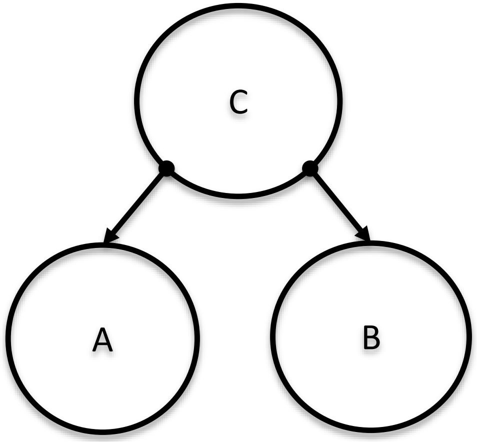
$$P(Income),\ P(Occupation),\ P(Age)$$
$$P(Buy\ X|Income, Occupation, Age)$$
$$P(Interested\ in\ Insurance|Buy\ X)$$



Conditional Independence Example (Part 2)

Martin and Norman toss the same coin. Let be A "Norman's outcome", and B "Martin's outcome". Assume they are both dependent on C "The coin is biased towards Heads with probability $\theta = 0.9$".

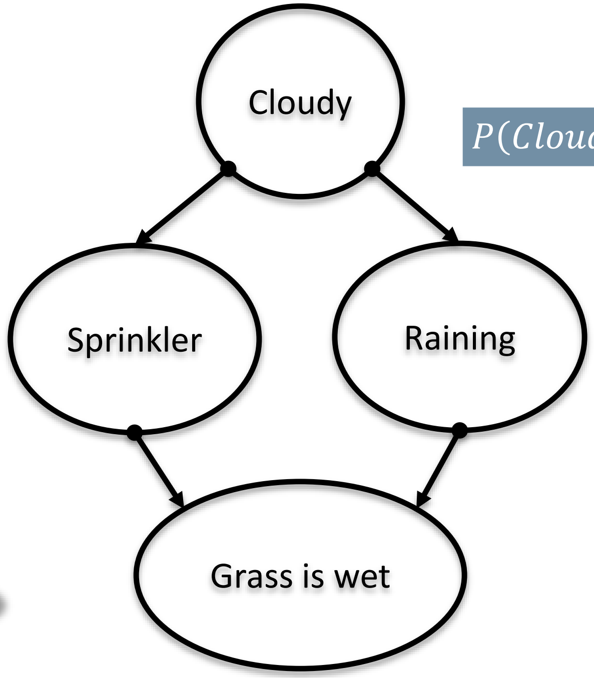
- You do not know if the coin is biased
- Martin tosses the coin and gets Tail
- What do you think about Norman toss?
- Now someone tells you the coin is biased
- What do you think about Norman toss?



The Sprinkler Example: Modeling

The event "Grass is wet" ($\boldsymbol{W = true}$) has two possible causes: either the water "*Sprinkler*" is on ($\boldsymbol{S = true}$) or it is "*Raining*" ($\boldsymbol{R = true}$).

Rows sum to 1



	True
$P(Cloudy)$	50%

Prior Probability

	$C = True$	$C = False$
$P(Sprinkler)$	50%	90%

	$C = True$	$C = False$
$P(Raining)$	50%	20%

Conditional Probability

	$S = True, R = True$	$S = True, R = False$	$S = False, R = True$	$S = False, R = False$
$P(Grass\ is\ wet)$	99%	90%	90%	0%

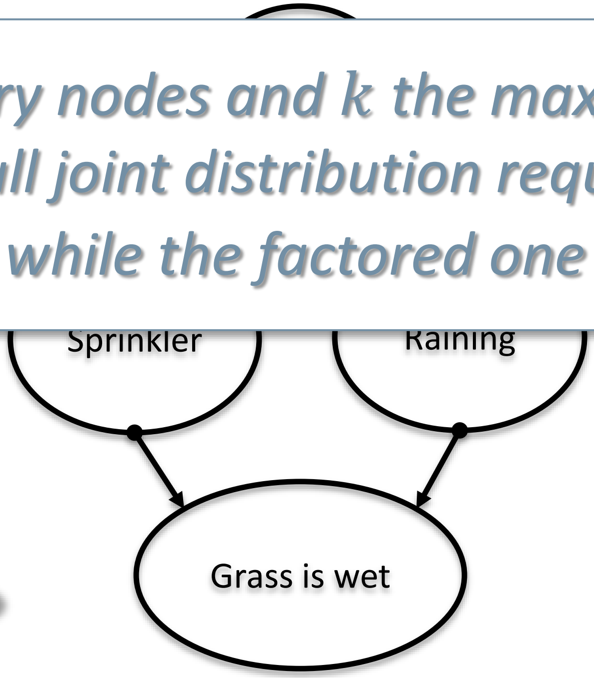
The Sprinkler Example: Modeling

The event "Grass is wet" ($\boldsymbol{W = true}$) has two possible causes: either the water "*Sprinkler*" is on ($\boldsymbol{S = true}$) or it is "*Raining*" ($\boldsymbol{R = true}$).

Rows sur

With N binary nodes and k the maximum node fan-in, the full joint distribution requires $O(2^N)$ parameters while the factored one $O(N * 2^k)$

Prior Probability



	$C = True$	$C = False$
$P(Sprinkler)$	50%	90%

	$C = True$	$C = False$
$P(Raining)$	50%	20%

Conditional Probability

	$S = True, R = True$	$S = True, R = False$	$S = False, R = True$	$S = False, R = False$
$P(Grassi\ is\ wet)$	99%	90%	90%	0%

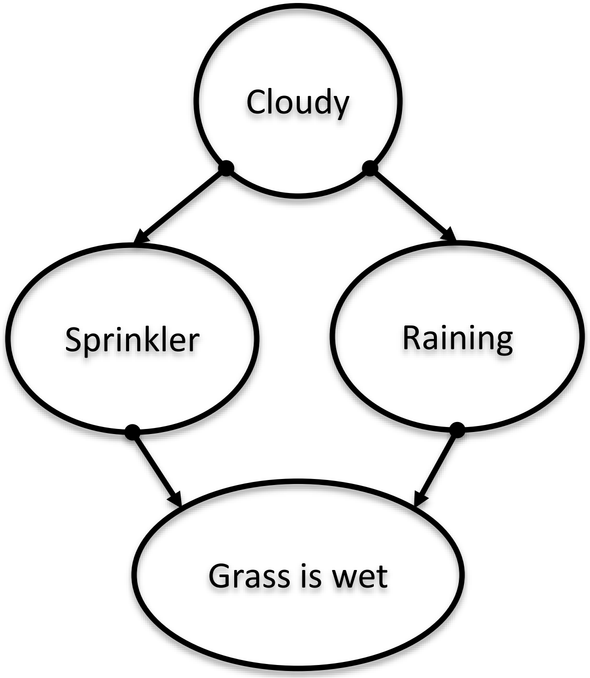
The Sprinkler Exampe: Joint Probability

The simplest conditional independence encoded in a Bayesian network states: "a node is independent of its ancestors given its parents."

Using the chain rule we get the joint probability:

$$
\begin{aligned}
P(C,S,R,W) &= P(W|S,R,C)P(S,R,C) \\
&= P(W|S,R)P(S,R,C) \\
&= P(W|S,R)P(S|R,C)P(R,C) \\
&= P(W|S,R)P(S|C)P(R,C) \\
&= P(W|S,R)P(S|C)P(R|C)P(C)
\end{aligned}
$$

This is like the biased coin ;-)



The Sprinkler Example: Making Inference

We observe the fact that the grass is wet. There are two possible causes for this: (a) the sprinkler is on or (b) it is raining. Which is more likely?

We know how to compute the joint distribution:

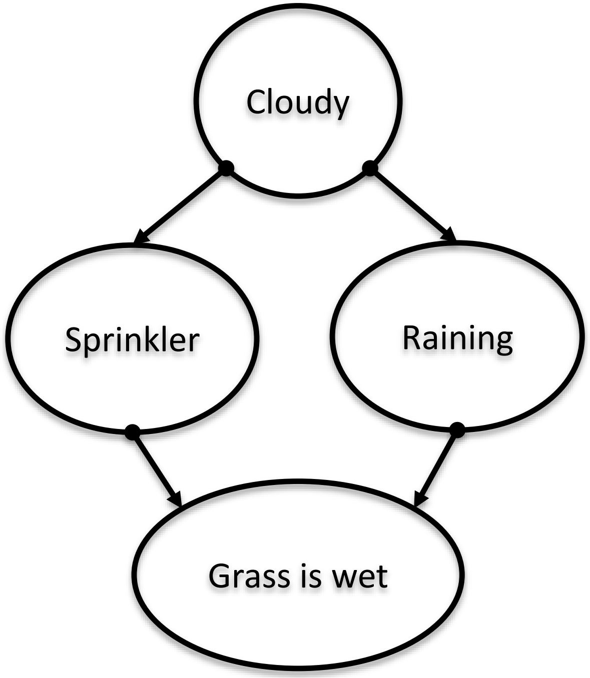
- The posterior probability can be computed as

$$P(E_1|E_2) = \frac{P(E_1 \wedge E_2)}{P(E_2)} = \frac{\sum_{row \sim E_1 \wedge E_2} P(row)}{\sum_{row \sim E_2} P(row)}$$

- What is the value of

$$P(S|W) = P(S,W)/P(W) = \sum_{C,R} P(C,S,R,W)/P(W)$$

$$= \frac{\sum_{C,R} P(C,S,R,W)}{\sum_{S,R,C} P(C,S,R,W)} = \cdots$$



The Sprinkler Example: Making Inference

C	S	R	W	P(C,S,R,W)	P(C,S,R,W)
0	0	0	0	0.5*0.1*0.8*1	0.04
0	0	0	1	0.5*0.1*0.8*0	0
0	0	1	0	0.5*0.1*0.2*0.1	0.001
0	0	1	1	0.5*0.1*0.2*0.9	0.009
0	1	0	0	0.5*0.9*0.8*0.1	0.036
0	1	0	1	0.5*0.9*0.8*0.9	0.324
0	1	1	0	0.5*0.9*0.2*0.01	0.0009
0	1	1	1	0.5*0.9*0.2*0.99	0.0891
1	0	0	0	0.5*0.5*0.5*1	0.125
1	0	0	1	0.5*0.5*0.5*0	0
1	0	1	0	0.5*0.5*0.5*0.1	0.0125
1	0	1	1	0.5*0.5*0.5*0.9	0.1125
1	1	0	0	0.5*0.5*0.5*0.1	0.0125
1	1	0	1	0.5*0.5*0.5*0.9	0.1125
1	1	1	0	0.5*0.5*0.5*0.01	0.00125
1	1	1	1	0.5*0.5*0.5*0.99	0.12375

Should sum up to 1!

$$P(S|W) = \frac{\sum_{C,R} P(C,S,R,W)}{\sum_{S,R,C} P(C,S,R,W)}$$

The Sprinkler Example: Making Inference

C	S	R	W	P(C,S,R,W)	P(C,S,R,W)
0	0	0	0	0.5*0.1*0.8*1	0.04
0	0	0	1	0.5*0.1*0.8*0	0
0	0	1	0	0.5*0.1*0.2*0.1	0.001
0	0	1	1	0.5*0.1*0.2*0.9	0.009
0	1	0	0	0.5*0.9*0.8*0.1	0.036
0	1	0	1	0.5*0.9*0.8*0.9	0.324
0	1	1	0	0.5*0.9*0.2*0.01	0.0009
0	1	1	1	0.5*0.9*0.2*0.99	0.0891
1	0	0	0	0.5*0.5*0.5*1	0.125
1	0	0	1	0.5*0.5*0.5*0	0
1	0	1	0	0.5*0.5*0.5*0.1	0.0125
1	0	1	1	0.5*0.5*0.5*0.9	0.1125
1	1	0	0	0.5*0.5*0.5*0.1	0.0125
1	1	0	1	0.5*0.5*0.5*0.9	0.1125
1	1	1	0	0.5*0.5*0.5*0.01	0.00125
1	1	1	1	0.5*0.5*0.5*0.99	0.12375

$$P(S|W) = \frac{\sum_{C,R} P(C,S,R,W)}{\sum_{S,R,C} P(C,S,R,W)}$$

The Sprinkler Example: Making Inference

C	S	R	W	P(C,S,R,W)	P(C,S,R,W)
0	0	0	0	0.5*0.1*0.8*1	0.04
0	0	0	1	0.5*0.1*0.8*0	0
0	0	1	0	0.5*0.1*0.2*0.1	0.001
0	0	1	1	0.5*0.1*0.2*0.9	0.009
0	1	0	0	0.5*0.9*0.8*0.1	0.036
0	1	0	1	0.5*0.9*0.8*0.9	0.324
0	1	1	0	0.5*0.9*0.2*0.01	0.0009
0	1	1	1	0.5*0.9*0.2*0.99	0.0891
1	0	0	0	0.5*0.5*0.5*1	0.125
1	0	0	1	0.5*0.5*0.5*0	0
1	0	1	0	0.5*0.5*0.5*0.1	0.0125
1	0	1	1	0.5*0.5*0.5*0.9	0.1125
1	1	0	0	0.5*0.5*0.5*0.1	0.0125
1	1	0	1	0.5*0.5*0.5*0.9	0.1125
1	1	1	0	0.5*0.5*0.5*0.01	0.00125
1	1	1	1	0.5*0.5*0.5*0.99	0.12375

$$P(S|W) = \frac{0.64935}{\sum_{S,R,C} P(C,S,R,W)}$$

The Sprinkler Example: Making Inference

C	S	R	W	P(C,S,R,W)	P(C,S,R,W)
0	0	0	0	0.5*0.1*0.8*1	0.04
0	0	0	1	0.5*0.1*0.8*0	0
0	0	1	0	0.5*0.1*0.2*0.1	0.001
0	0	1	1	0.5*0.1*0.2*0.9	0.009
0	1	0	0	0.5*0.9*0.8*0.1	0.036
0	1	0	1	0.5*0.9*0.8*0.9	0.324
0	1	1	0	0.5*0.9*0.2*0.01	0.0009
0	1	1	1	0.5*0.9*0.2*0.99	0.0891
1	0	0	0	0.5*0.5*0.5*1	0.125
1	0	0	1	0.5*0.5*0.5*0	0
1	0	1	0	0.5*0.5*0.5*0.1	0.0125
1	0	1	1	0.5*0.5*0.5*0.9	0.1125
1	1	0	0	0.5*0.5*0.5*0.1	0.0125
1	1	0	1	0.5*0.5*0.5*0.9	0.1125
1	1	1	0	0.5*0.5*0.5*0.01	0.00125
1	1	1	1	0.5*0.5*0.5*0.99	0.12375

$$P(S|W) = \frac{0.64935}{0.77085} = 0.8424$$

The Sprinkler Example: Making Inference

We observe the fact that the grass is wet. There are two possible causes for this: (a) the sprinkler is on or (b) it is raining. Which is more likely?

We know how to compute the joint distribution:

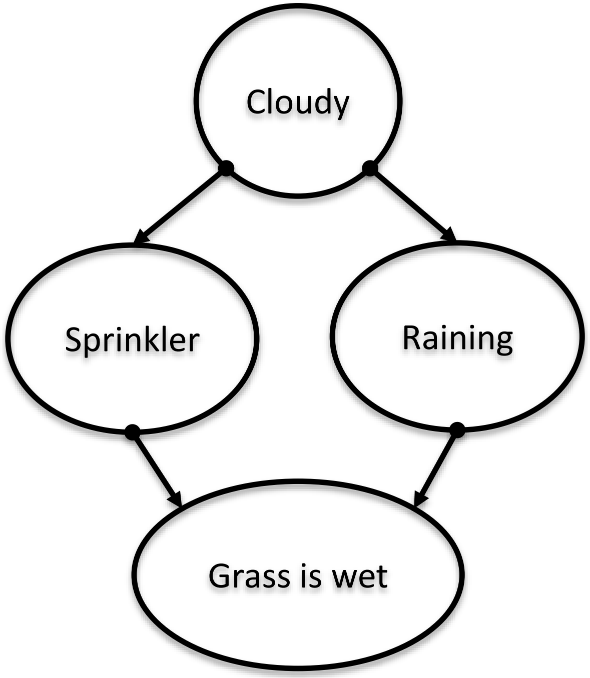
- The posterior probability can be computed as

$$P(E_1|E_2) = \frac{P(E_1 \wedge E_2)}{P(E_2)} = \frac{\sum_{row\sim E_1 \wedge E_2} P(row)}{\sum_{row\sim E_2} P(row)}$$

- What is the value of

$$P(R|W) = P(R,W)/P(W) = \sum_{S,C} P(C,S,R,W)/P(W)$$

$$= \frac{\sum_{S,C} P(C,S,R,W)}{\sum_{S,R,C} P(C,S,R,W)} = \cdots$$



The Sprinkler Example: Making Inference

C	S	R	W	P(C,S,R,W)	P(C,S,R,W)
0	0	0	0	0.5*0.1*0.8*1	0.04
0	0	0	1	0.5*0.1*0.8*0	0
0	0	1	0	0.5*0.1*0.2*0.1	0.001
0	0	1	1	0.5*0.1*0.2*0.9	0.009
0	1	0	0	0.5*0.9*0.8*0.1	0.036
0	1	0	1	0.5*0.9*0.8*0.9	0.324
0	1	1	0	0.5*0.9*0.2*0.01	0.0009
0	1	1	1	0.5*0.9*0.2*0.99	0.0891
1	0	0	0	0.5*0.5*0.5*1	0.125
1	0	0	1	0.5*0.5*0.5*0	0
1	0	1	0	0.5*0.5*0.5*0.1	0.0125
1	0	1	1	0.5*0.5*0.5*0.9	0.1125
1	1	0	0	0.5*0.5*0.5*0.1	0.0125
1	1	0	1	0.5*0.5*0.5*0.9	0.1125
1	1	1	0	0.5*0.5*0.5*0.01	0.00125
1	1	1	1	0.5*0.5*0.5*0.99	0.12375

$$P(R|W) = \frac{\sum_{C,S} P(C,S,R,W)}{\sum_{S,R,C} P(C,S,R,W)}$$

The Sprinkler Example: Making Inference

C	S	R	W	P(C,S,R,W)	P(C,S,R,W)
0	0	0	0	0.5*0.1*0.8*1	0.04
0	0	0	1	0.5*0.1*0.8*0	0
0	0	1	0	0.5*0.1*0.2*0.1	0.001
0	0	1	1	0.5*0.1*0.2*0.9	0.009
0	1	0	0	0.5*0.9*0.8*0.1	0.036
0	1	0	1	0.5*0.9*0.8*0.9	0.324
0	1	1	0	0.5*0.9*0.2*0.01	0.0009
0	1	1	1	0.5*0.9*0.2*0.99	0.0891
1	0	0	0	0.5*0.5*0.5*1	0.125
1	0	0	1	0.5*0.5*0.5*0	0
1	0	1	0	0.5*0.5*0.5*0.1	0.0125
1	0	1	1	0.5*0.5*0.5*0.9	0.1125
1	1	0	0	0.5*0.5*0.5*0.1	0.0125
1	1	0	1	0.5*0.5*0.5*0.9	0.1125
1	1	1	0	0.5*0.5*0.5*0.01	0.00125
1	1	1	1	0.5*0.5*0.5*0.99	0.12375

$$P(R|W) = \frac{0.33435}{\sum_{S,R,C} P(C,S,R,W)}$$

The Sprinkler Example: Making Inference

C	S	R	W	P(C,S,R,W)	P(C,S,R,W)
0	0	0	0	0.5*0.1*0.8*1	0.04
0	0	0	1	0.5*0.1*0.8*0	0
0	0	1	0	0.5*0.1*0.2*0.1	0.001
0	0	1	1	0.5*0.1*0.2*0.9	0.009
0	1	0	0	0.5*0.9*0.8*0.1	0.036
0	1	0	1	0.5*0.9*0.8*0.9	0.324
0	1	1	0	0.5*0.9*0.2*0.01	0.0009
0	1	1	1	0.5*0.9*0.2*0.99	0.0891
1	0	0	0	0.5*0.5*0.5*1	0.125
1	0	0	1	0.5*0.5*0.5*0	0
1	0	1	0	0.5*0.5*0.5*0.1	0.0125
1	0	1	1	0.5*0.5*0.5*0.9	0.1125
1	1	0	0	0.5*0.5*0.5*0.1	0.0125
1	1	0	1	0.5*0.5*0.5*0.9	0.1125
1	1	1	0	0.5*0.5*0.5*0.01	0.00125
1	1	1	1	0.5*0.5*0.5*0.99	0.12375

$$P(R|W) = \frac{0.33435}{0.77085} = 0.4337$$

The Sprinkler Example: Making Inference

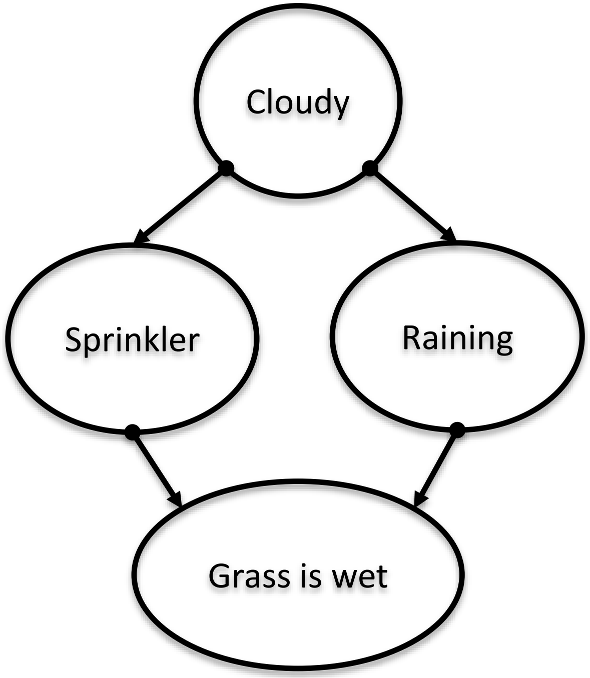
We observe the fact that the grass is wet. There are two possible causes for this: (a) the sprinkler is on or (b) it is raining. Which is more likely?

We know how to compute the joint distribution:

- The posterior probability can be computed as

$$P(E_1|E_2) = \frac{P(E_1 \wedge E_2)}{P(E_2)} = \frac{\sum_{row \sim E_1 \wedge E_2} P(row)}{\sum_{row \sim E_2} P(row)}$$

- $P(S|W) = 0.8424$
- $P(R|W) = 0.4337$



The most likely reason for the grass do be wet is …

The Sprinkler Example: Explaining Away

In Sprinkler Example the two causes *"compete"* to *"explain"* the observed data. Hence S and R become conditionally dependent given that their common child, W, is observed.

Example: Suppose the grass is wet, but we know that it is raining. Then the posterior probability of sprinkler being on becomes:

$$P(S|W,R) = \frac{P(S,W|R)}{P(W|R)} = \frac{\sum_C P(C,S,W|R)}{\sum_{S,C} P(C,S,W|R)}$$

The Sprinkler Example: Making Inference

C	S	R	W	P(C,S,R,W)	P(C,S,R,W)
0	0	0	0	0.5*0.1*0.8*1	0.04
0	0	0	1	0.5*0.1*0.8*0	0
0	0	1	0	0.5*0.1*0.2*0.1	0.001
0	0	1	1	0.5*0.1*0.2*0.9	0.009
0	1	0	0	0.5*0.9*0.8*0.1	0.036
0	1	0	1	0.5*0.9*0.8*0.9	0.324
0	1	1	0	0.5*0.9*0.2*0.01	0.0009
0	1	1	1	0.5*0.9*0.2*0.99	0.0891
1	0	0	0	0.5*0.5*0.5*1	0.125
1	0	0	1	0.5*0.5*0.5*0	0
1	0	1	0	0.5*0.5*0.5*0.1	0.0125
1	0	1	1	0.5*0.5*0.5*0.9	0.1125
1	1	0	0	0.5*0.5*0.5*0.1	0.0125
1	1	0	1	0.5*0.5*0.5*0.9	0.1125
1	1	1	0	0.5*0.5*0.5*0.01	0.00125
1	1	1	1	0.5*0.5*0.5*0.99	0.12375

$$P(S|W,R) = \frac{\sum_C P(C,S,W|R)}{\sum_{S,C} P(C,S,W|R)}$$

$$= \frac{0.201475}{\sum_{S,C} P(C,S,W|R)}$$

The Sprinkler Example: Making Inference

C	S	R	W	P(C,S,R,W)	P(C,S,R,W)
0	0	0	0	0.5*0.1*0.8*1	0.04
0	0	0	1	0.5*0.1*0.8*0	0
0	0	1	0	0.5*0.1*0.2*0.1	0.001
0	0	1	1	0.5*0.1*0.2*0.9	0.009
0	1	0	0	0.5*0.9*0.8*0.1	0.036
0	1	0	1	0.5*0.9*0.8*0.9	0.324
0	1	1	0	0.5*0.9*0.2*0.01	0.0009
0	1	1	1	0.5*0.9*0.2*0.99	0.0891
1	0	0	0	0.5*0.5*0.5*1	0.125
1	0	0	1	0.5*0.5*0.5*0	0
1	0	1	0	0.5*0.5*0.5*0.1	0.0125
1	0	1	1	0.5*0.5*0.5*0.9	0.1125
1	1	0	0	0.5*0.5*0.5*0.1	0.0125
1	1	0	1	0.5*0.5*0.5*0.9	0.1125
1	1	1	0	0.5*0.5*0.5*0.01	0.00125
1	1	1	1	0.5*0.5*0.5*0.99	0.12375

$$P(S|W,R) = \frac{\sum_C P(C,S,W|R)}{\sum_{S,C} P(C,S,W|R)}$$

$$= \frac{0.201475}{\sum_{S,C} P(C,S,W|R)} = \frac{0.201475}{0.33435}$$

$$= 0.6026$$

The Sprinkler Example: Explaining Away

In Sprinkler Example the two causes *"compete"* to *"explain"* the observed data. Hence S and R become conditionally dependent given that their common child, W, is observed.

<u>*Example*</u>: Suppose the grass is wet, but we know that it is raining. Then the posterior probability of sprinkler being on becomes:

$$P(S|W,R) = \frac{P(S,W|R)}{P(W|R)} = \frac{\sum_C P(C,S,W|R)}{\sum_{S,C} P(C,S,W|R)}$$

This means that the probability of Sprinkler goes down

$$P(S|W,R) \leq P(S|W)$$

This is called Explaining Away

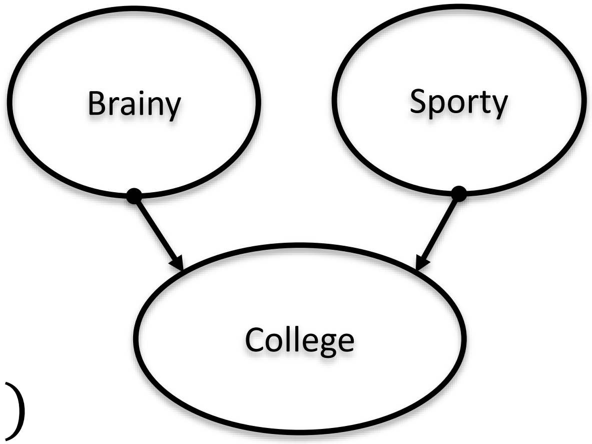
Explaining Away Phenomenon

Explaining Away is known in Statstics as *Berkson's Paradox*, or *Selection Bias*, and it describes two variable which become dependent because you observe a third one.

<u>*Example*</u>: Consider a college which admits students who are either *Brainy* or *Sporty* (or both!). Let $\boldsymbol{C}$ denote the event "admitted to College", which is $\mathbf{True}$ if a student is either Brainy ($\boldsymbol{B}$) or Sporty ($\boldsymbol{S}$).

Suppose in population, $\boldsymbol{B}$ and $\boldsymbol{S}$ are independent.

In *College*, being *Brainy* makes you less likely to be *Sporty*, because either are sufficient to explain $\boldsymbol{C}$



$$P(S = True|C = True, B = True) \leq P(S = True|B = True)$$

Bottom-Up vs Top-Down Reasoning

Looking at the simple Sprinkler Example we can already see the two kind of reasoning we can make with Bayesian Networks:

- Bottom-Up: we had evidence of an effect ($\boldsymbol{Grass\ is\ Wet}$), and inferred the most likely cause; it goes from effects to causes as in diagnostic systems;
- Top-Down: we can compute the probability grass will be wet given that it is cloudy; this is predictive use of Bayesian Networks as "generative" models.

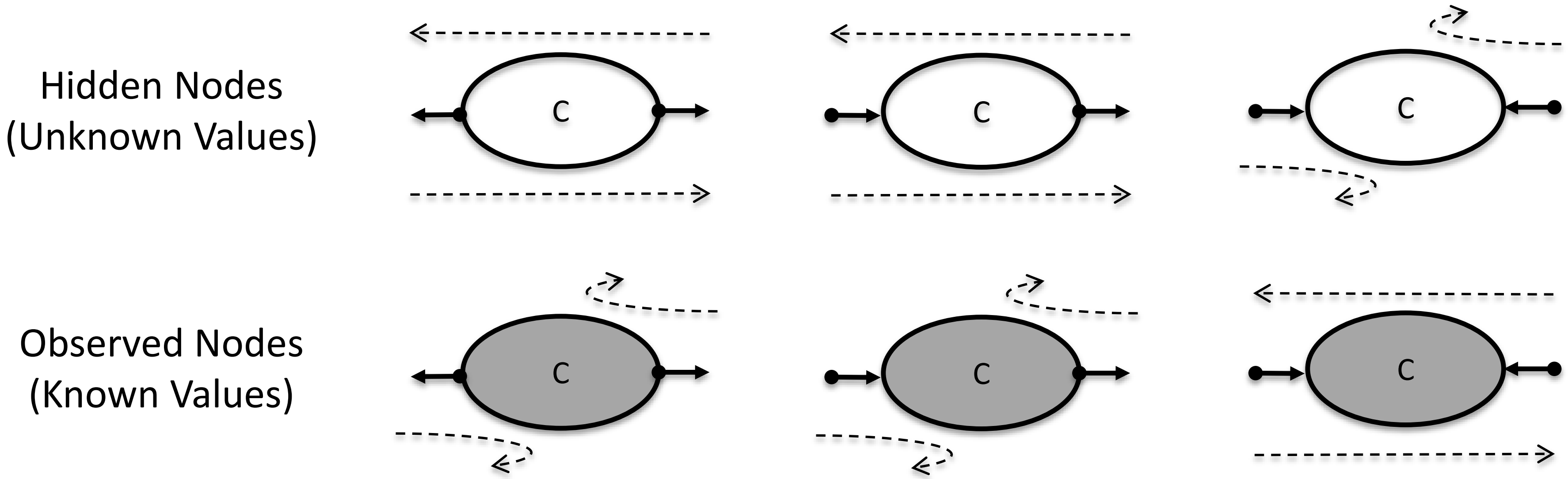
The most interesting property of Bayesian Networks is that they can be used to reason about causality on a solid mathematical basis:

- Question: Can we distinguish causation from mere correlation? So we don't need to make experiments to infer causality.
- Answer: Yes, sometimes, but we need to measure the relationships between at least three variables.

Causality: Models, Reasoning and Inference, Judea Pearl, 2000.

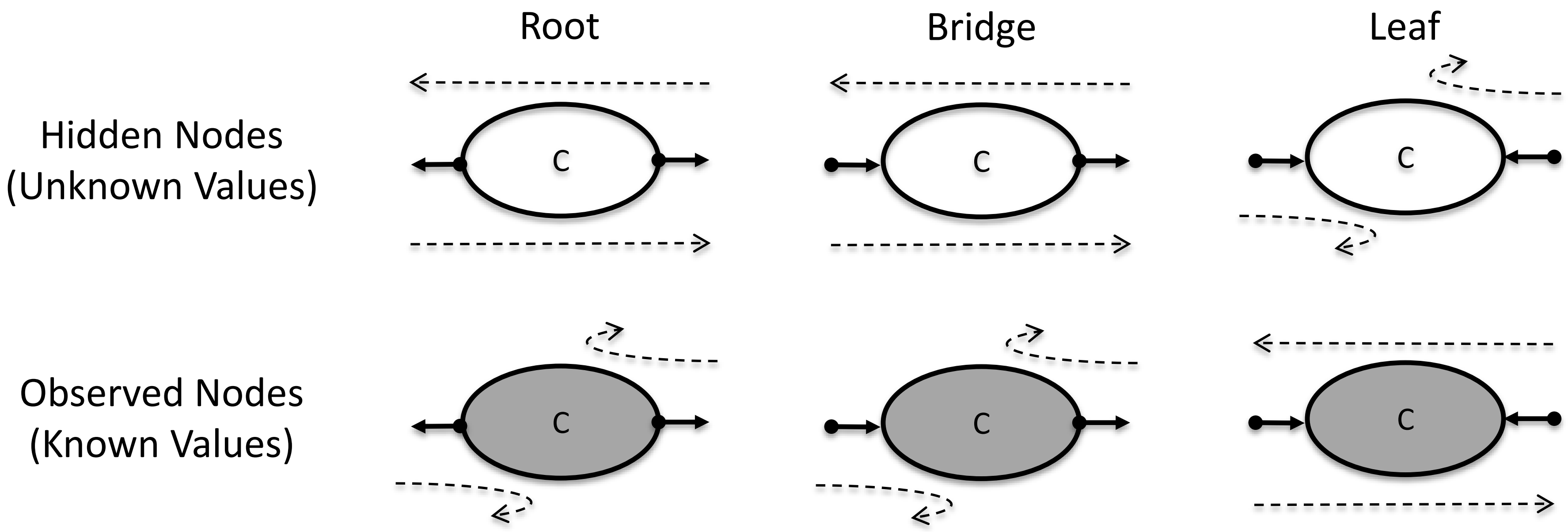
Conditional Independence in Bayesian Networks

Two (sets of) nodes $\boldsymbol{A}$ and $\boldsymbol{B}$ are conditionally independent (d-separated) given $\boldsymbol{C}$ <u>*if and only if*</u> all the path from $\boldsymbol{A}$ to $\boldsymbol{B}$ are shielded by $\boldsymbol{C}$.



The dotted arcs indicate direction of flow in the path, if they do not traverse the node then the path is shielded.

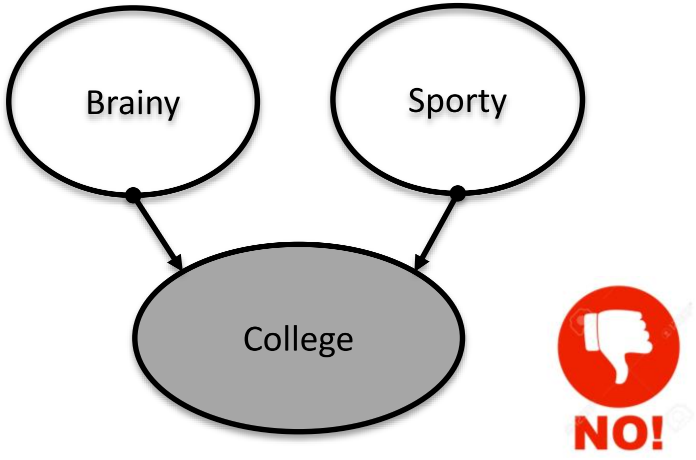
Conditional Independence in Bayesian Networks

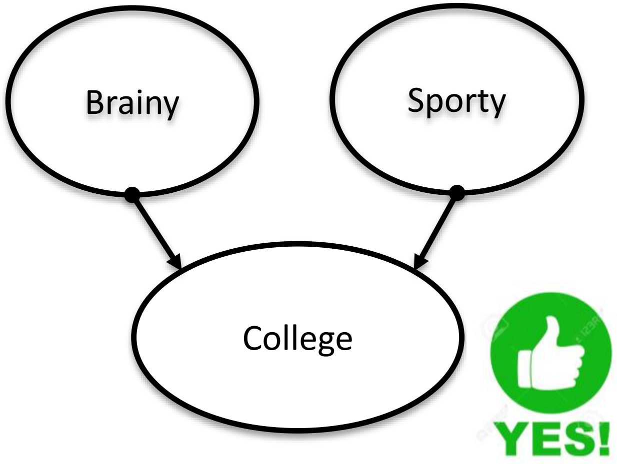


- *C is a "root":* if C is hidden, children are dependent due to a hidden common cause. If C is observed, they are conditionally independent;
- *C is a "leaf":* if C is hidden, its parents are marginally independent, but if C is observed, the parents become dependent (Explaining Away);
- *C is a "bridge":* nodes upstream and downstream of C are dependent if and only if C is hidden, because conditioning breaks the graph at that point.

Examples on d-separtion

Simply using d-separation:





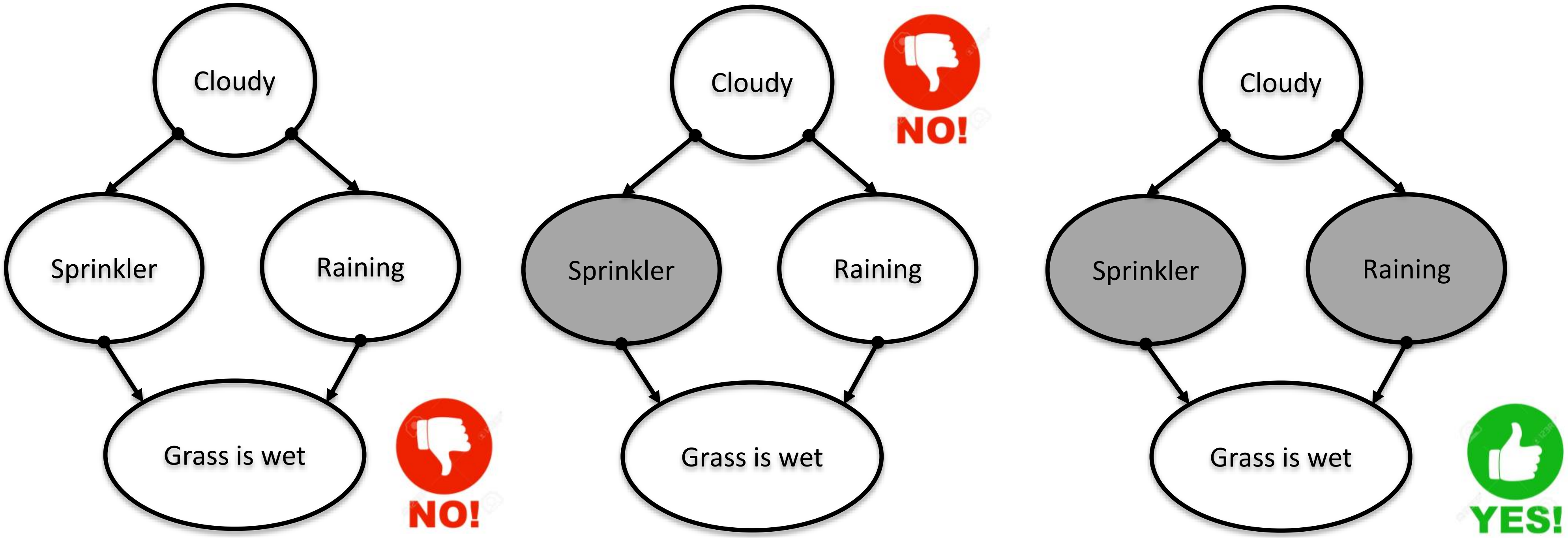
$$P(B, S|C) = P(B|C) * P(S|C)$$

$$P(B, S) = P(B) * P(S)$$

Examples on d-separtion

Simply using d-separation:

$$P(W,C|S) = P(W|S) * P(C|S)$$



$$P(W,C) = P(W) * P(C)$$

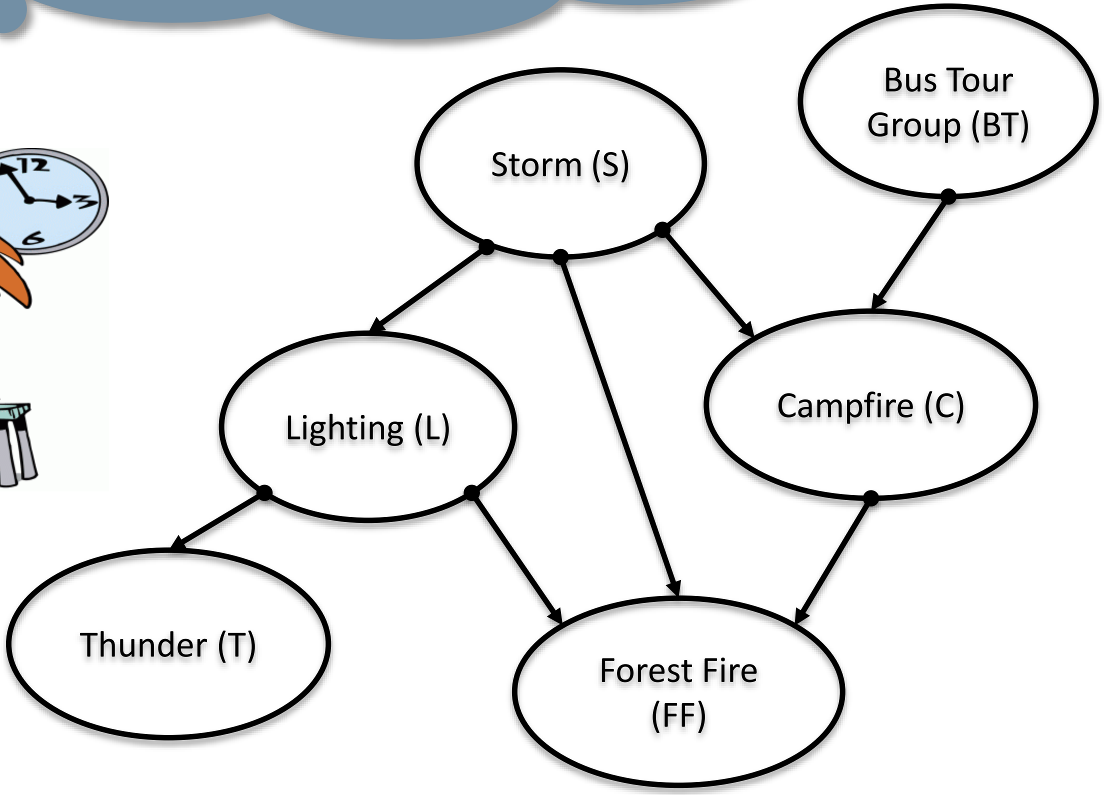
$$P(W,C|S,R) = P(W|S,R) * P(C|S,R)$$

Examples on d-separtion

What can you say about …

You will see how this come handy at inference time!!!

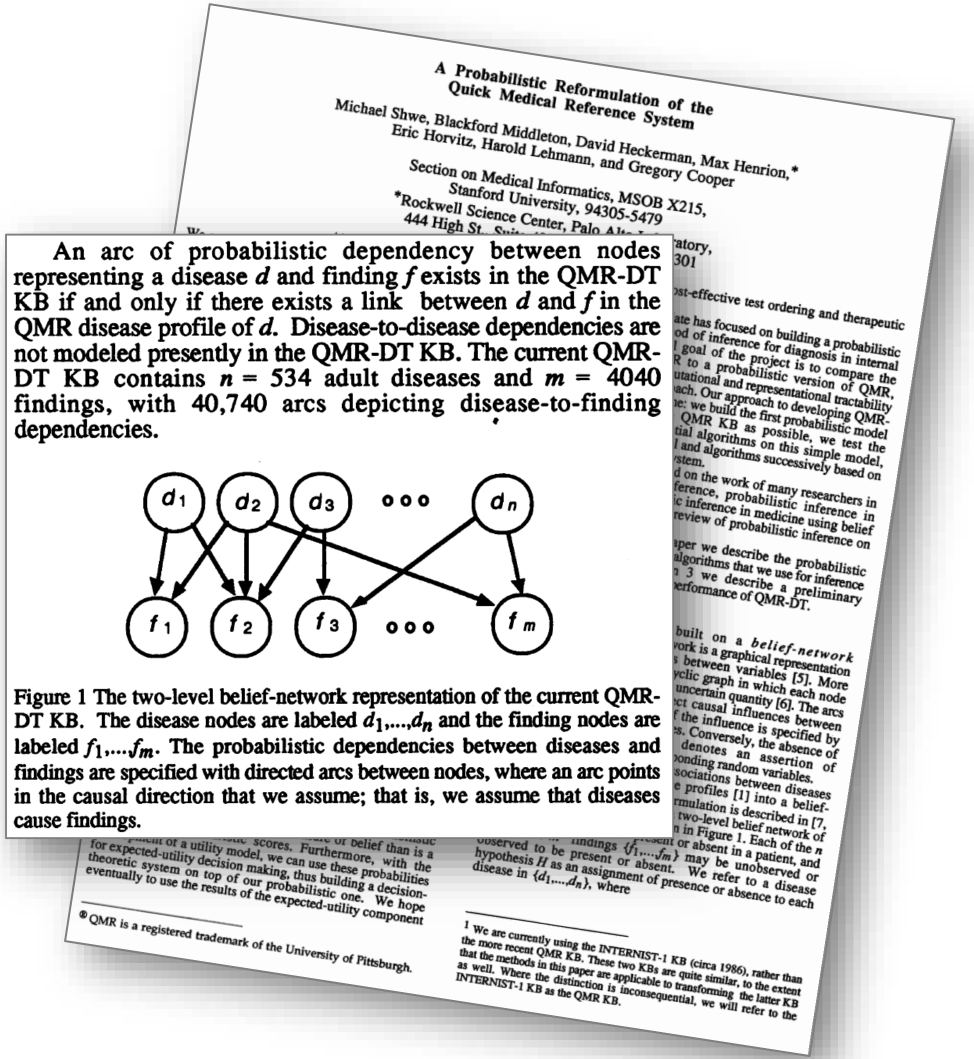
- $P(FF|L,C)$
- $P(L,C|FF,BT)$
- $P(T|BT)$
- $P(S|T,L,FF)$
- $P(C|FF,L)$
- $P(FF|T,C,S)$
- $P(FF,T|L)$
- $P(FF,T,L,C,S,BT)$
- …



Bayesian Network Applications (1)

A famous example is the reformulation of the Quick Medical Reference model (1990)

- Top layer represents 534 hidden disease;
- Bottom layer represents 4040 possibly observed symptoms;
- The goal is to infer the posterior probability of each disease given all the symptoms (which can be present, absent or unknown).


A Probabilistic Reformulation of the Quick Medical Reference System

Michael Shwe, Blackford Middleton, David Heckerman, Max Henrion,* Eric Horvitz, Harold Lehmann, and Gregory Cooper

Section on Medical Informatics, MSOB X215, Stanford University, 94305-5479

*Rockwell Science Center, Palo Alto ...

An arc of probabilistic dependency between nodes representing a disease d and finding f exists in the QMR-DT KB if and only if there exists a link between d and f in the QMR disease profile of d. Disease-to-disease dependencies are not modeled presently in the QMR-DT KB. The current QMR-DT KB contains n = 534 adult diseases and m = 4040 findings, with 40,740 arcs depicting disease-to-finding dependencies.

Figure 1 The two-level belief-network representation of the current QMR-DT KB. The disease nodes are labeled $d_1,\ldots,d_n$ and the finding nodes are labeled $f_1,\ldots,f_m$. The probabilistic dependencies between diseases and findings are specified with directed arcs between nodes, where an arc points in the causal direction that we assume; that is, we assume that diseases cause findings.

®QMR is a registered trademark of the University of Pittsburgh.

QMR-DT was so densely connected that exact inference was impossible.

Bayesian Network Applications (2)

The most widely used Bayesian Networks were embedded in Microsoft's products (2000):

- Answer Wizard in Office 95;
- Office Assistant in Office 97;
- Over 30 Technical Support Troubleshooters.

Check more on the Economist article (22/3/01) about Microsoft's Bayesian Networks (https://www.cs.ubc.ca/~murphyk/Bayes/econ.22mar01.html).

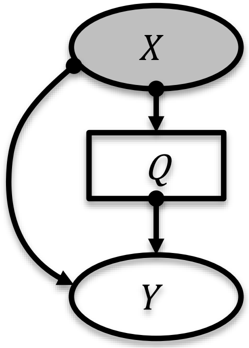
More modern applications exists being Bayesian Networks a general framework

Bayesian Networks as Unifying Framework

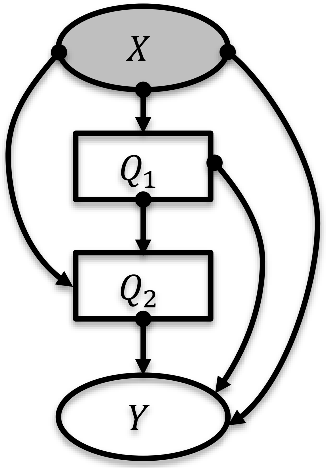
We can have Bayesian Networks with both real and discrete nodes:

- Discrete nodes with continuous parents, we use logistic/softmax distribution;
- Most common distribution for real nodes is Gaussian.

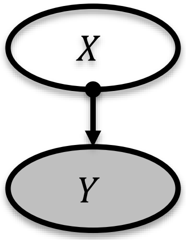
Using these we can obtain a rich toolbox for probabilistic modeling



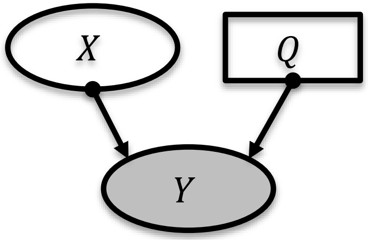
Mixture of Experts



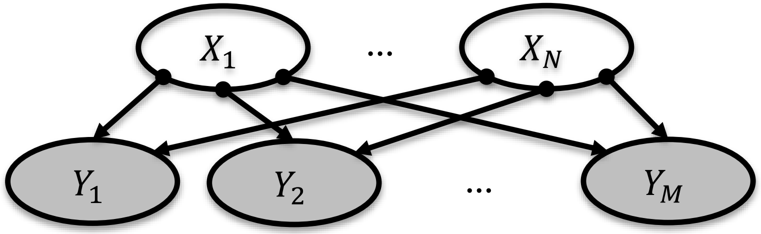
Hierarchical Mixture of Experts



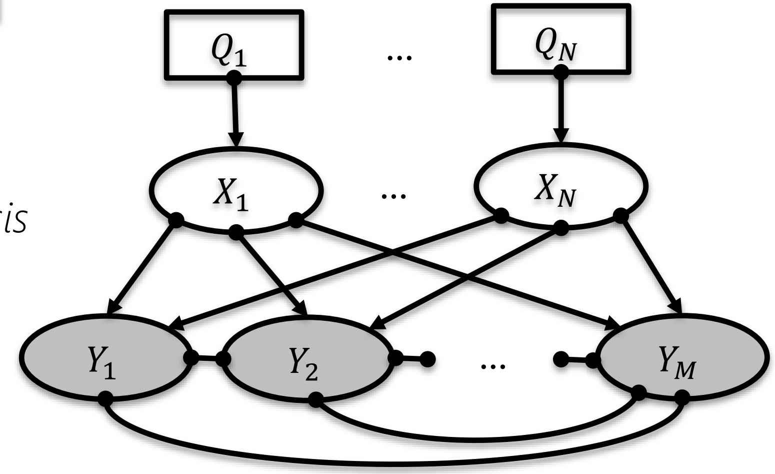
Factor Analysis / PCA



Mixture factor Analysis



Factor Analysis



Independent Factor Analysis

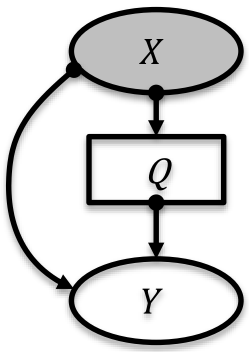
Bayesian Networks as Unifying Framework

We can have Bayesian Networks with both real and discrete nodes:

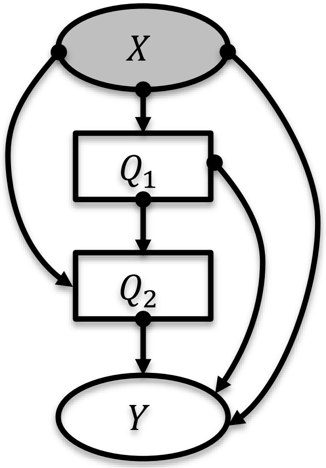
- Discrete nodes with continuous par
- Most common distribution f

S. Roweis, Z. Ghahramani. "A Unifying Review of Linear Gaussian Models", Neural Computation 11(2):305-345. 1999

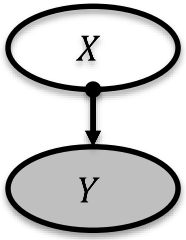
Using these we can obtain a rich tool



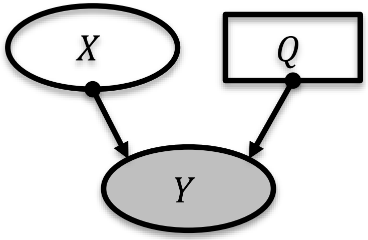
Mixture of Experts



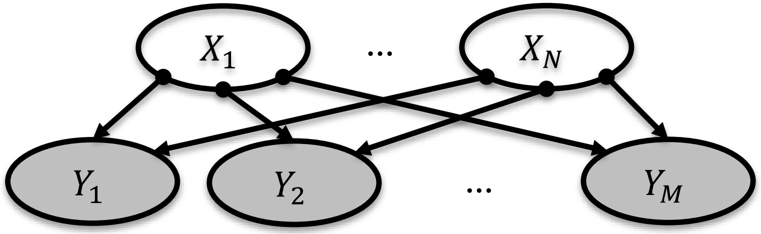
Hierarchical Mixture of Experts



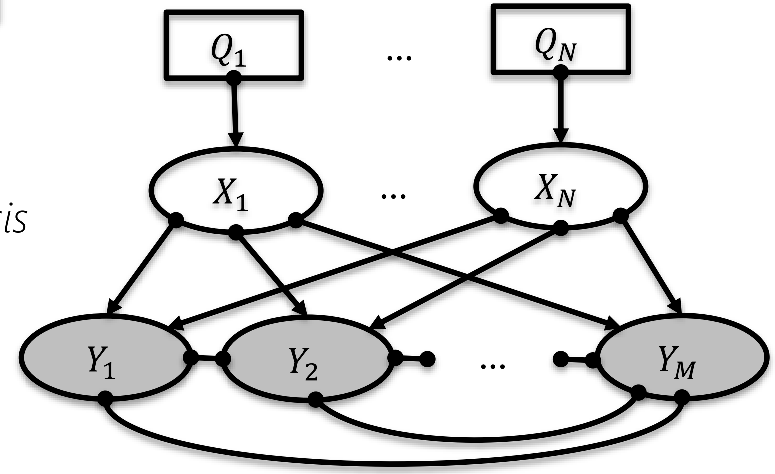
Factor Analysis / PCA



Mixture factor Analysis



Factor Analysis

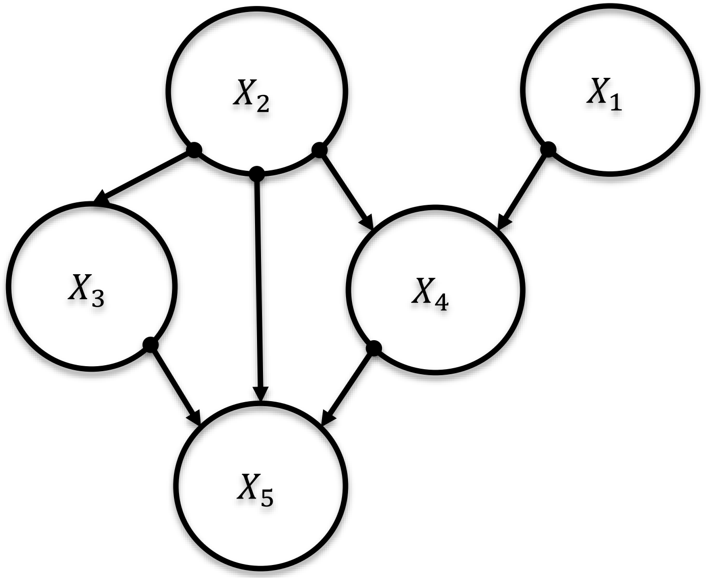


Independent Factor Analysis

Bayesian Networks Wrap-up!

Bayesian Network

- A Directed Acyclic Graph (DAG) where Nodes represent random variables and Edges represent direct influence
- Conditional Probability Distributions (CPD) for priors and "influenced" variables



Joint Distribution Factorization

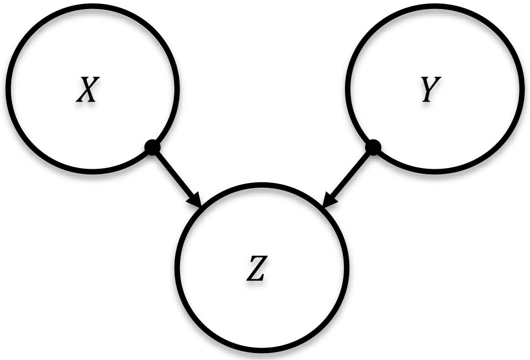
$$P(X) = P(X_1, X_2, \dots, X_N) = \prod_{k=1}^{N} P\big(X_k | parents(X_k)\big) = \prod_{k=1}^{N} P(X_k | pa_k)$$

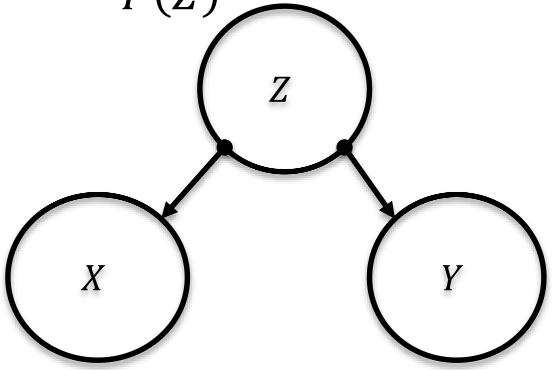
Bayesian Networks Wrap-up!

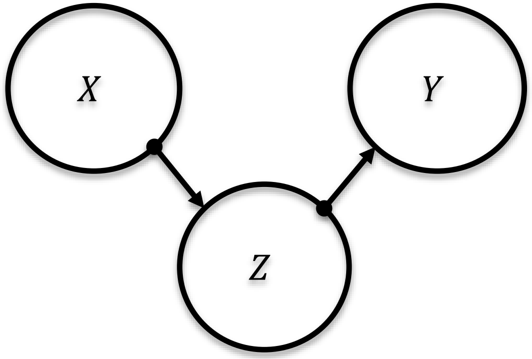
Independencies

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

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







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

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

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

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

The Sprinkler Example: Making Inference (Return)

We observe the fact that the grass is wet. There are two possible causes for this: (a) the sprinkler is on or (b) it is raining. Which is more likely?

We know how to compute the joint distribution:

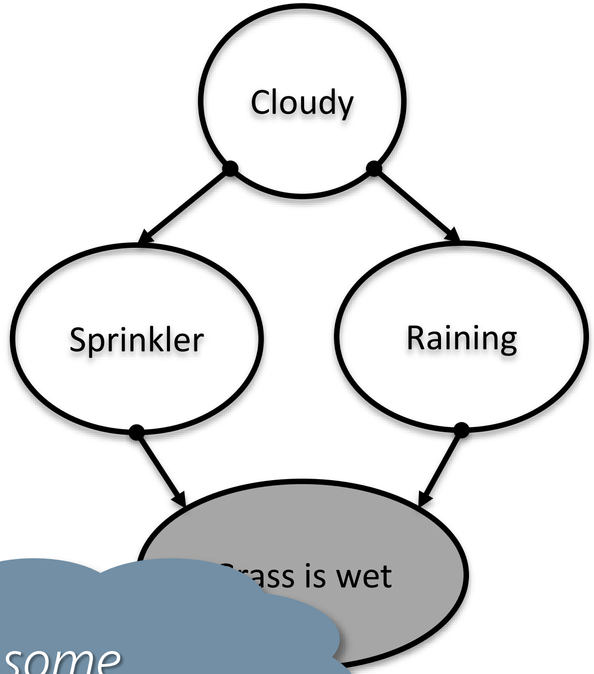
- The posterior probability can be computed as

$$P(E_1|E_2) = \frac{P(E_1 \wedge E_2)}{P(E_2)} = \frac{\sum_{row\sim E_1 \wedge E_2} P(row)}{\sum_{row\sim E_2} P(row)}$$

- What is the value of

$$P(S|W) = P(S,W)/P(W) = \sum_{C,R} P(C,S,R,W)/P(W)$$

$$= \frac{\sum_{C,R} P(C,S,R,W)}{\sum_{S,R,C} P(C,S,R,W)} = \cdots$$



Can we save some computation?

The Sprinkler Example: Making Inference (Return)

We observe the fact that the grass is wet. There are two possible causes for this: (a) the sprinkler is on or (b) it is raining. Which is more likely give that yesterday was cloudy?

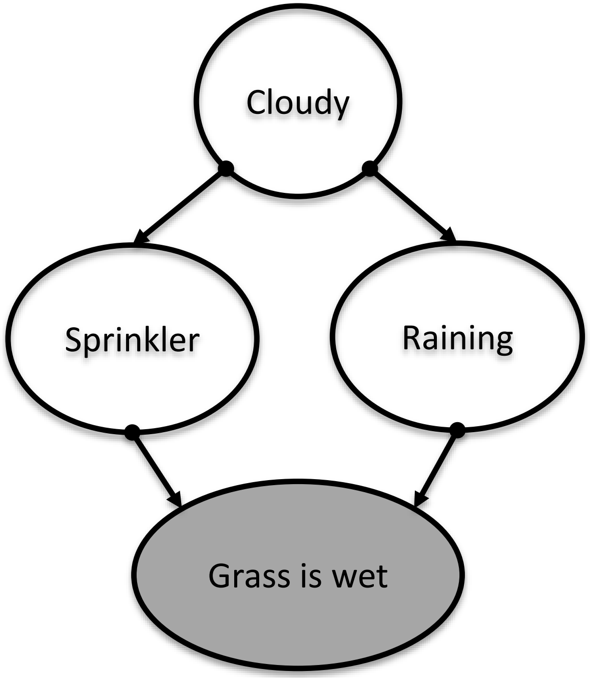
Let' workout some math!

$$P(S=T|W=T,C=T)=\frac{P(W=T|S=T,C=T)P(S|C=T)}{P(W=T|C=T)}=\frac{?*?}{?}$$

$$P(W=T|C=T)=\sum_{R,S}P(W=T,R,S|C=T)$$

$$=\sum_{R,S}P(W=T|R,S,C=T)P(R,S|C=T)$$

$$=\sum_{R,S}P(W=T|R,S)P(R|C=T)P(S|C=T)=\cdots$$

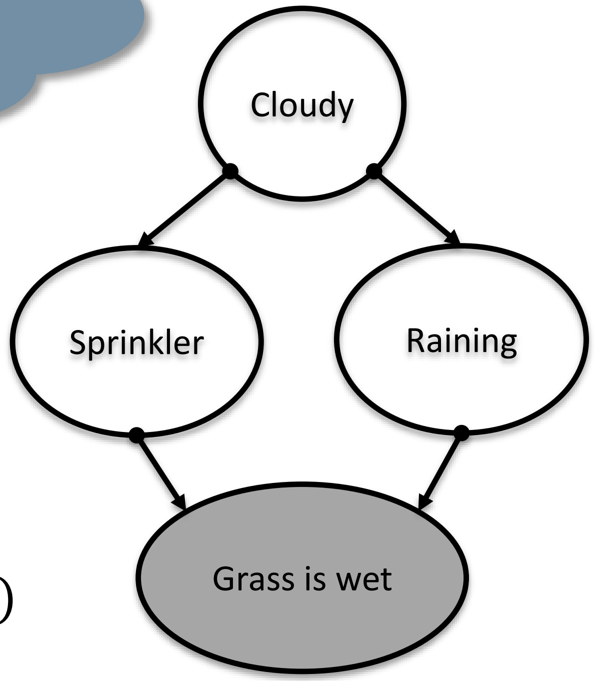


The Sprinkler Example: Making Inference (Return)

We observe the fact that the grass is wet. There are two possible causes for this: (a) the sprinkler is on or (b) it is raining. Which is more likely give that yesterday was cloudy?

Can we do it automatically / efficiently ?

Let’ workout some math!



$$P(S=T|W=T,C=T)=\frac{P(W=T|S=T,C=T)P(S|C=T)}{P(W=T|C=T)}=\frac{?*?}{?}$$

$$P(W=T|S=T,C=T)=\sum_R P(W=T,R|S=T,C=T)$$

$$=\sum_R P(W=T|R,S=T,C=T)P(R|S=T,C=T)$$

$$=\sum_R P(W=T|R,S=T)P(R|C=T)=\cdots$$