EECS 391
Intro to AI

Assignment Project Exam Help

Inference in Bayes Nets

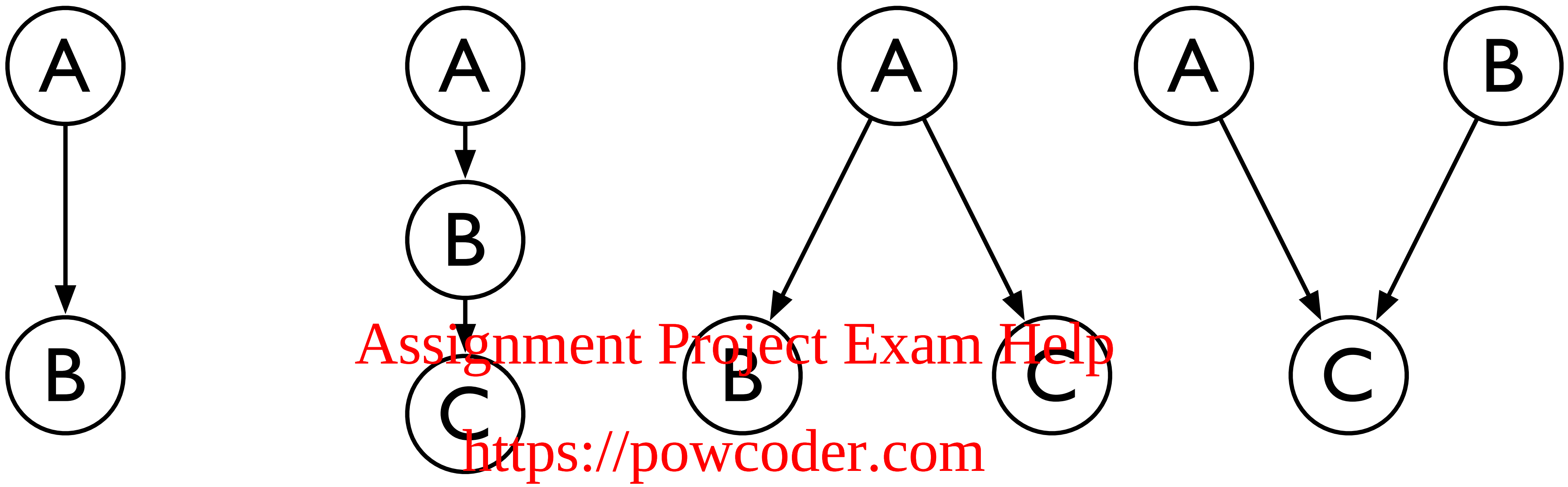
https://powcoder.com

Add WeChat powcoder

L14 Thu Oct 25

Recap: Modeling causal relationships with belief networks

Direct cause | Indirect cause | Common cause | Common effect



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Direct cause	Indirect cause	Common cause	Common effect
P(B|A)	P(B|A) P(C|B)	P(B|A) P(C|A)	P(C|A,B)

Defining the belief network

- Each link in the graph represents a conditional relationship between nodes.
- To compute the inference, we must specify the conditional probabilities.
- Let's start with the open door. What do we specify?

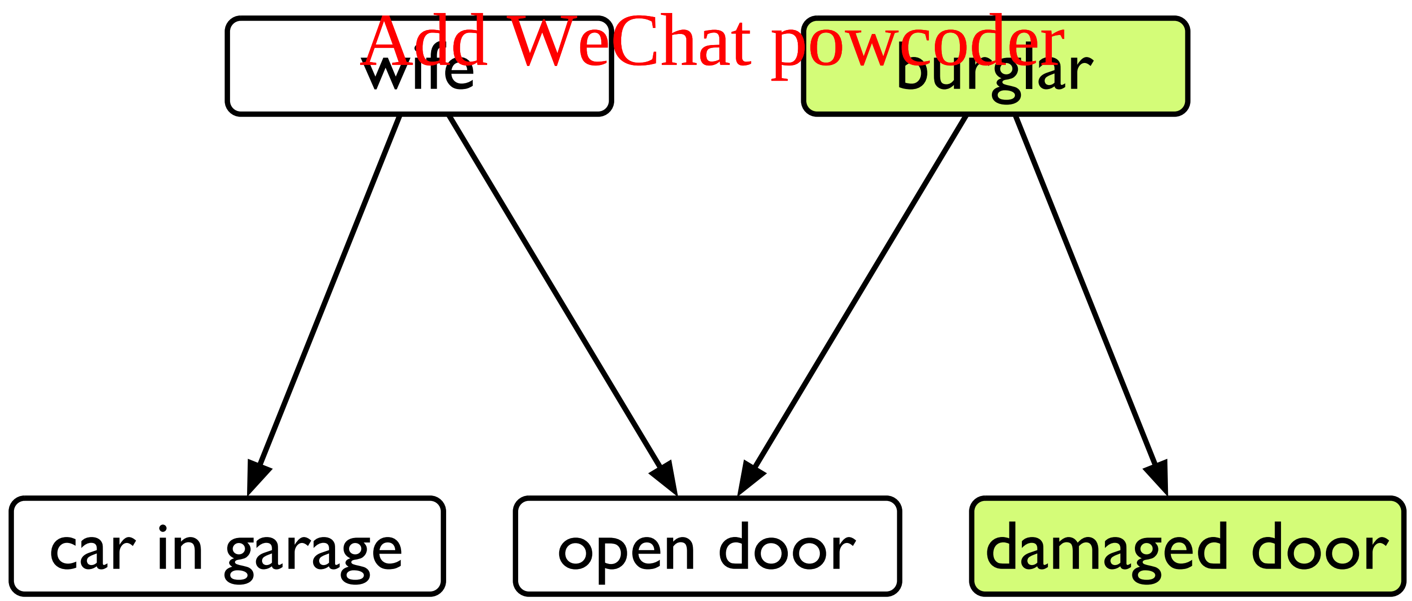
W	B	P(O|W,B)
F	F	0.01
F	T	0.25
T	F	0.05
T	T	0.75

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

P(W)
0.05

P(B)
0.001

Finally, we specify the remaining conditionals



W	P(C|W)
F	0.01
T	0.95

B	P(D|B)
F	0.001
T	0.5

Now what?

Calculating probabilities using the joint distribution

- What the probability that the door is open, it is my wife and not a burglar, we see the car in the garage, and the door is not damaged?
- Mathematically, we want to compute the expression: $P(o,w,\neg b,c,\neg d)$ = ?
- We can just repeatedly apply the rule relating joint and conditional probabilities.
  - $P(x,y) = P(x|y)\, P(y)$

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Summary of inference with the joint probability distribution

- The complete (probabilistic) relationship between variables is specified by the joint probability:

$$P(X_1, X_2, \ldots, X_n)$$
$$= P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

- All conditional and marginal distributions can be derived from this using the basic rules of probability, the sum rule and the product rule

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

$$P(X) = \sum_Y P(X, Y)$$ sum rule, "marginalization"

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

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

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

Calculating probabilities using the joint distribution

- The probability that the door is open, it is my wife and not a burglar, we see the car in the garage, and the door is not damaged.
- $P(o,w,\neg b,c,\neg d) = P(o|w,\neg b,c,\neg d)P(w,\neg b,c,\neg d)$

$$= P(o|w,\neg b)P(w,\neg b,c,\neg d)$$

$$= P(o|w,\neg b)P(c|w,\neg b,\neg d)P(w,\neg b,\neg d)$$

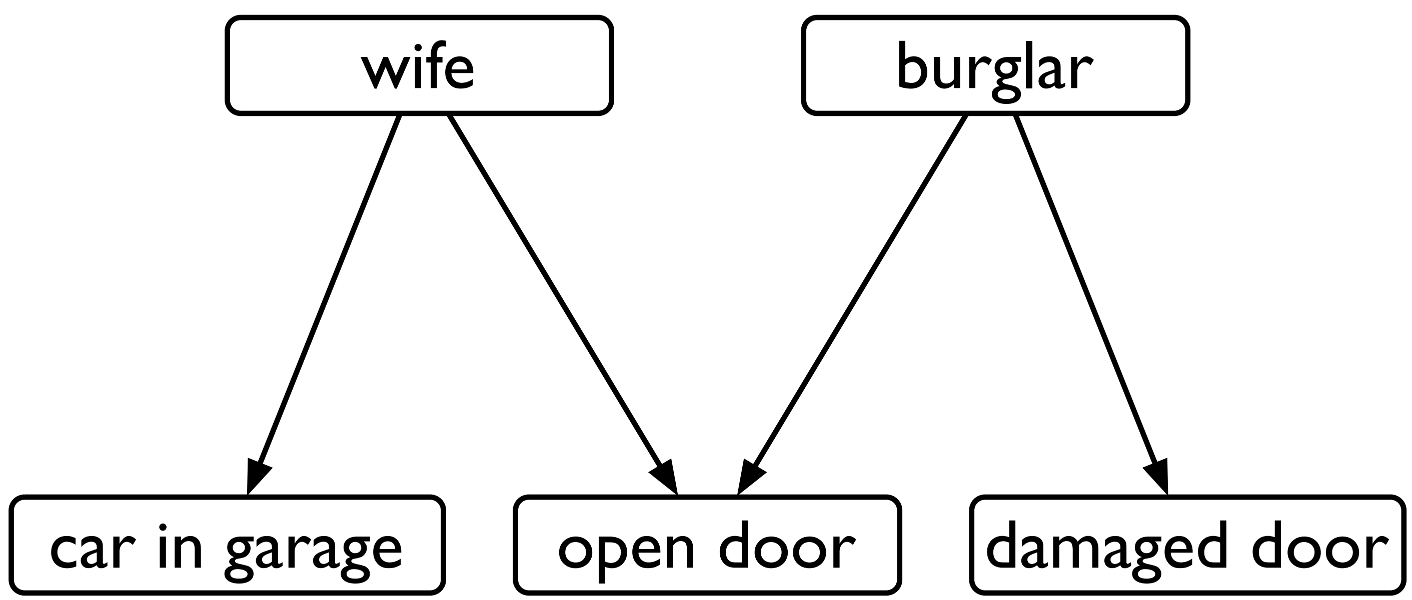
$$= P(o|w,\neg b)P(c|w)P(w,\neg b,\neg d)$$

$$= P(o|w,\neg b)P(c|w)P(\neg d|w,\neg b)P(w,\neg b)$$

$$= P(o|w,\neg b)P(c|w)P(\neg d|\neg b)P(w,\neg b)$$

$$= P(o|w,\neg b)P(c|w)P(\neg d|\neg b)P(w)P(\neg b)$$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Calculating probabilities using the joint distribution

- P(o,w,¬b,c,¬d) = P(o|w,¬b)P(c|w)P(¬d|¬b)P(w)P(¬b)

  = 0.05 × 0.95 × 0.999 × 0.05 × 0.999 = 0.0024

- This is essentially the probability that my wife is home and leaves the door open.

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

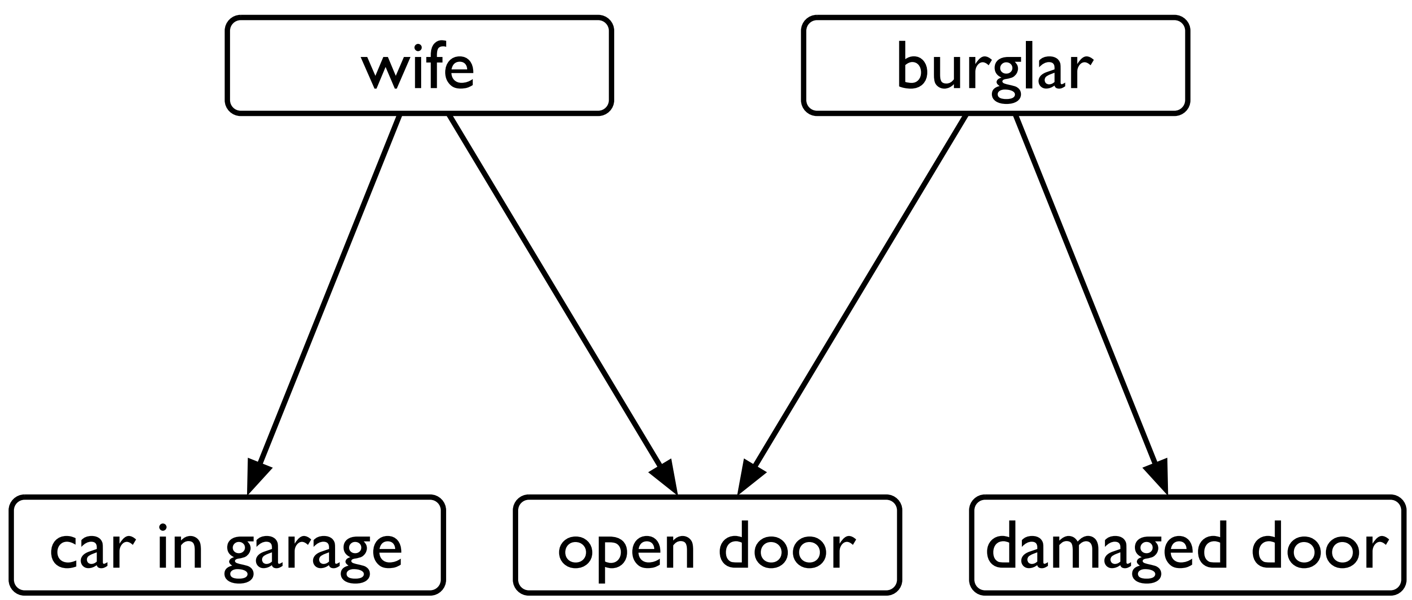
W	B	P(O|W,B)
F	F	0.01
F	T	0.25
T	F	0.05
T	T	0.75

P(W)
0.05

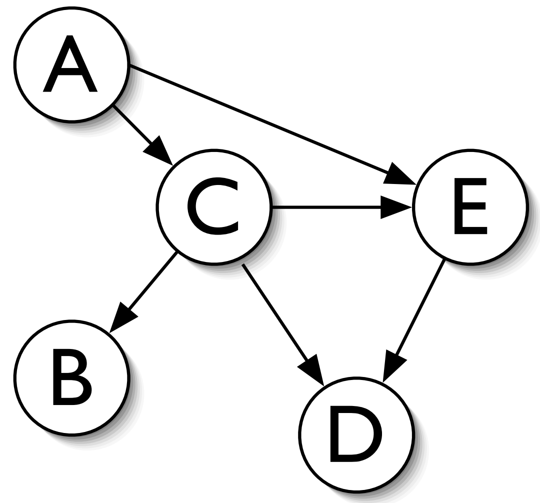
P(B)
0.001

W	P(C|W)
F	0.01
T	0.95

B	P(D|B)
F	0.001
T	0.5



Calculating probabilities in a general Bayesian belief network



- Note that by specifying all the conditional probabilities, we have also specified the joint probability. For the directed graph above:

Assignment Project Exam Help

P(A,B,C,D,E) = P(A) P(B|C) P(C|A) P(D|C,E) P(E|A,C)

https://powcoder.com

- The general expression is:

Add WeChat powcoder

$$
\begin{aligned}
P(x_1, \ldots, x_n) &\equiv P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n) \\
&= \prod_{i=1}^{n} P(x_i | \mathrm{parents}(X_i))
\end{aligned}
$$

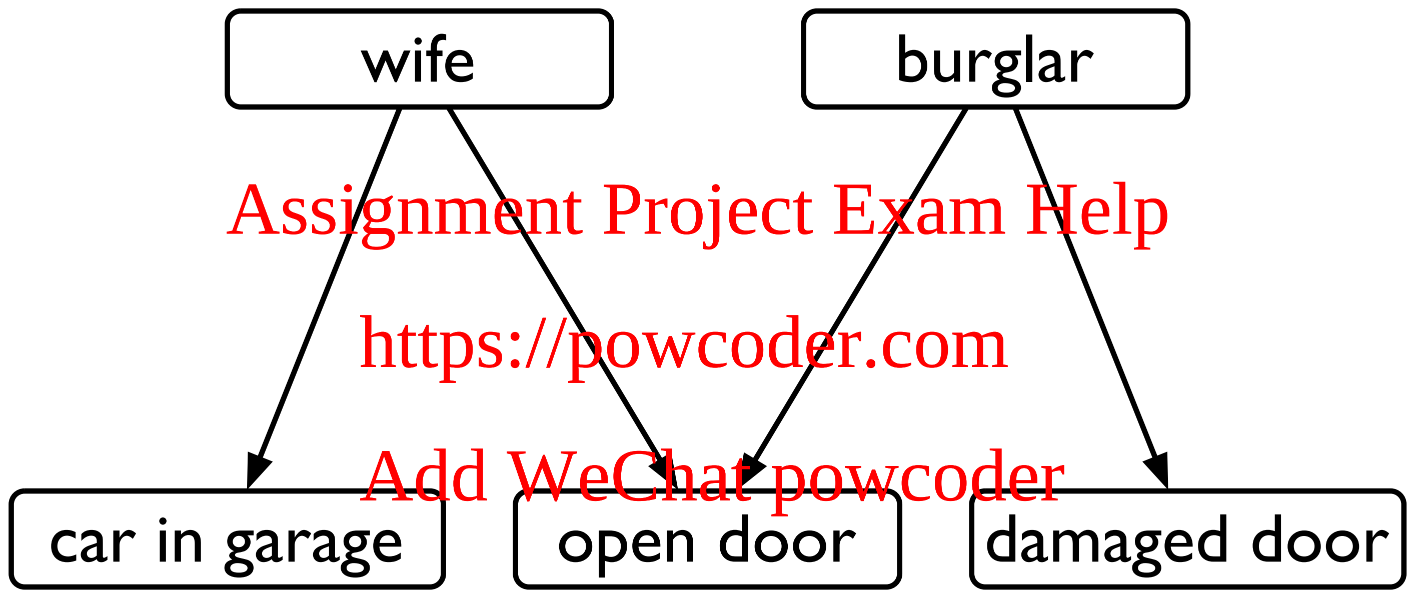
- With this we can calculate (in principle) the probability of any joint probability.
- This implies that we can also calculate any conditional probability.

For the burglar model

W	B	P(O|W,B)
F	F	0.01
F	T	0.25
T	F	0.05
T	T	0.75

P(W)
0.05

P(B)
0.001



W	P(C|W)
F	0.01
T	0.95

B	P(D|B)
F	0.001
T	0.5

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- The structure of this model allows a simple expression for the joint probability

$$P(x_1, \ldots, x_n) \equiv P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$$
$$= \prod_{i=1}^{n} P(x_i|\text{parents}(X_i))$$
$$\Rightarrow P(o, c, d, w, b) = P(c|w)P(o|w,b)P(d|b)P(w)P(b)$$

What if we want a simpler probabilistic question?

- How do we calculate P(b|o), i.e. the probability of a burglar given we see the open door?
- This is not an entry in the joint distribution. We had:

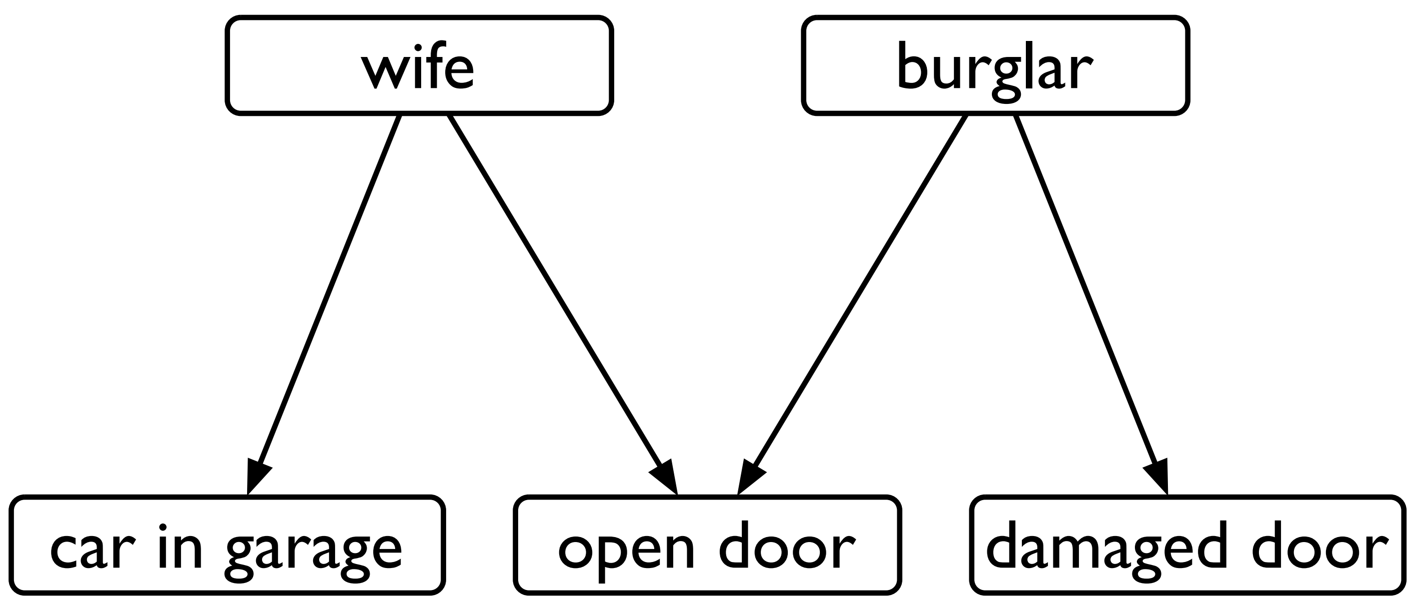
$$P(o,w,\neg b,c,\neg d) = P(o|w,\neg b)P(c|w)P(\neg d|\neg b)P(w)P(\neg b)$$
$$= \quad 0.05 \times 0.95 \times 0.999 \times 0.05 \times 0.999 = 0.0024$$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

W	B	P(O|W,B)
F	F	0.01
F	T	0.25
T	F	0.05
T	T	0.75

P(W)
0.05

P(B)
0.001



W	P(C|W)
F	0.01
T	0.95

B	P(D|B)
F	0.001
T	0.5

Calculating conditional probabilities

- So, how *do* we compute P(b|o)?
- Repeatedly apply laws of probability (factorization, marginalization, etc).
- Using the joint we can compute any conditional probability too
- The conditional probability of any one subset of variables given another disjoint subset is

$$P(S_1|S_2) = \frac{P(S_1 \wedge S_2)}{P(S_2)} = \frac{\sum p \in S_1 \wedge S_2}{\sum p \in S_2}$$

Assignment Project Exam Help

https://powcoder.com

where p ∈ S is shorthand for all the entries of the joint matching subset S.

Add WeChat powcoder

- How many terms are in this sum? 2^N

The number of terms in the sums is *exponential* in the number of variables.

In fact, general querying Bayes nets is NP complete.

Variable elimination on the burglary network

- We could do straight summation:

$$p(b|o) = \alpha\, p(o, w, b, c, d)$$
$$= \alpha \sum_{w,c,d} p(o|w, b)p(c|w)p(d|b)p(w)p(b)$$

- But: the number of terms in the sum is *exponential* in the non-evidence variables.
- This is bad, and we can do much better.
- We start by observing that we can pull out many terms from the summation.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Variable elimination

- When we've pulled out all the redundant terms we get:

$$p(b|o) = \alpha\, p(b) \sum_d p(d|b) \sum_w p(w) p(o|w,b) \sum_c p(c|w)$$

- We can also note the last term sums to one. In fact, every variable that is not an ancestor of a query variable or evidence variable is *irrelevant* to the query, so we get

$$p(b|o) = \alpha\, p(b) \sum_d p(d|b) \sum_w p(w) p(o|w,b)$$

Assignment Project Exam Help

which contains far fewer terms. In general, *complexity is **linear** in the # of CPT entries.*

https://powcoder.com

Add WeChat powcoder

Variable elimination

- We can go even further.
- If we exchange the sums we get (with all the steps):

$$
\begin{aligned}
p(b|o) &= \alpha \sum_d p(d|b) \sum_w p(w)p(o|w,b) \\
&= \alpha \sum_d \sum_w p(d|b)p(w)p(o|w,b) \\
&= \alpha \sum_w \sum_d p(d|b)p(w)p(o|w,b) \\
&= \alpha \sum_w p(w)p(o|w,b) \sum_d p(d|b) \\
&= \alpha \sum_w p(w)p(o|w,b) \cdot 1
\end{aligned}
$$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- We could have also achieved this by a more direct path.

Variable elimination

- When we've pulled out all the redundant terms we get:

$$p(b|o) = \alpha \sum_{w} p(w)p(o|w, b)$$

- which contains far fewer terms than the original expression.
- In general, *complexity is **linear** in the # of CPT entries.*
- This method is called *variable elimination*.

Assignment Project Exam Help

  - if # of parents is bounded, also linear in the number of nodes.

https://powcoder.com

  - the expressions are evaluated in right-to-left order (bottom-up in the network)
  - intermediate results are stored

Add WeChat powcoder

  - sums over each are done only for those expressions that depend on the variable
- Note: for multiply connected networks, variable elimination can have exponential complexity in the worst case.

General inference questions in Bayesian networks

- For queries in Bayesian networks, we divide variables into three classes:
  - evidence variables: $e = \{e_1, \ldots, e_m\}$ what you know
  - query variables: $x = \{x_1, \ldots, x_n\}$ what you want to know
  - non-evidence variables: $y = \{y_1, \ldots, y_l\}$ what you don't care about
- The complete set of variables in the network is $\{e \cup x \cup y\}$.
- Inferences in Bayesian networks consist of computing p(x|e), the posterior probability of the query given the evidence:

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

$$p(x|e) = \frac{p(x,e)}{p(e)} = \alpha p(x,e) = \alpha \sum_y p(x,e,y)$$

- This computes the marginal distribution p(x,e) by summing the joint over all values of y.
- Recall that the joint distribution is defined by the product of the conditional pdfs:

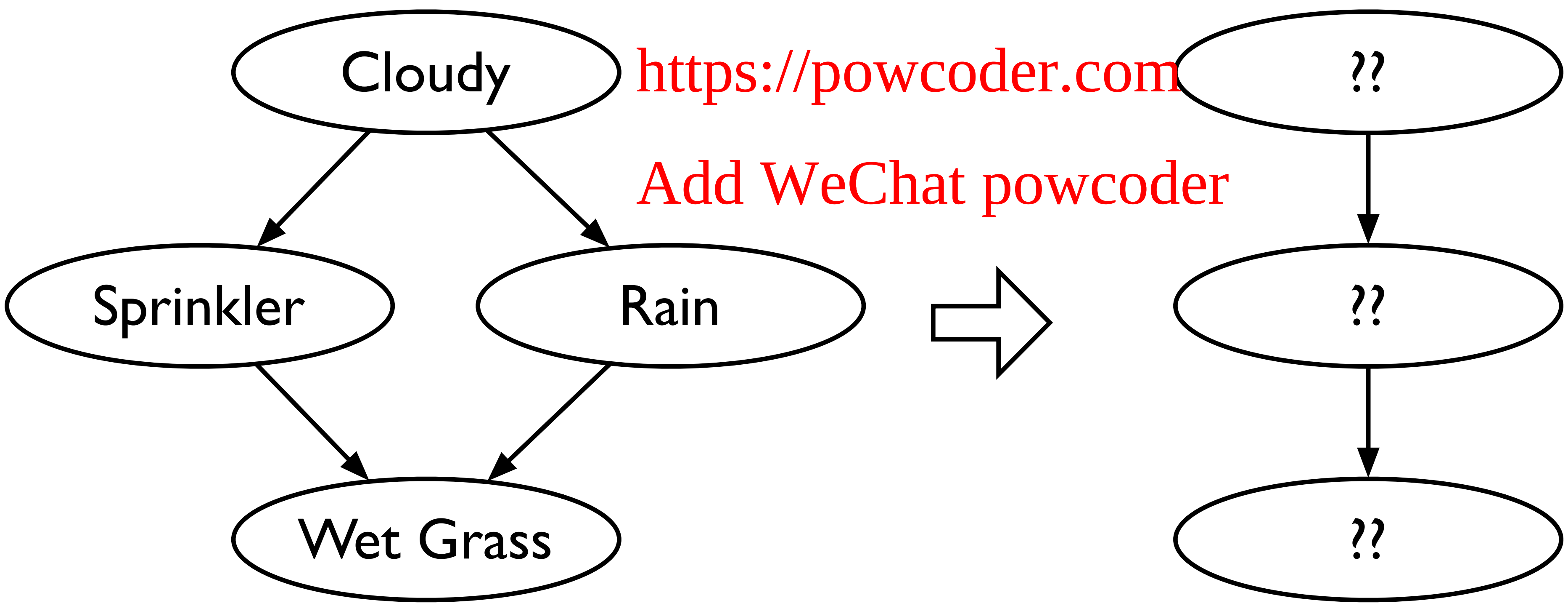
$$p(z) = \prod_{i=1} P(z_i|\text{parents}(z_i))$$

where the product is taken over all variables in the network.

Another approach: Simplify model using clustering algorithms

- Inference is efficient if you have a *polytree*, ie a singly connected network.
- But what if you don't?
- Idea: Convert a non-singly connected network to an equivalent singly connected network.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

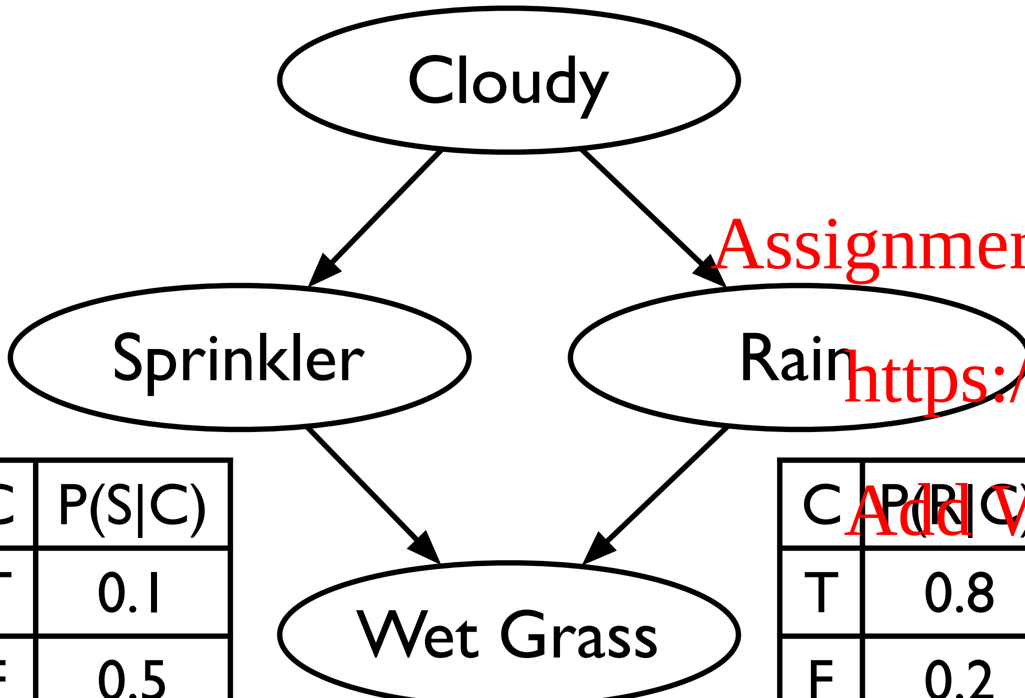


What should go into the nodes?

Clustering or join tree algorithms

Idea: merge multiply connected nodes into a single, higher-dimensional case.

P(C)
0.5



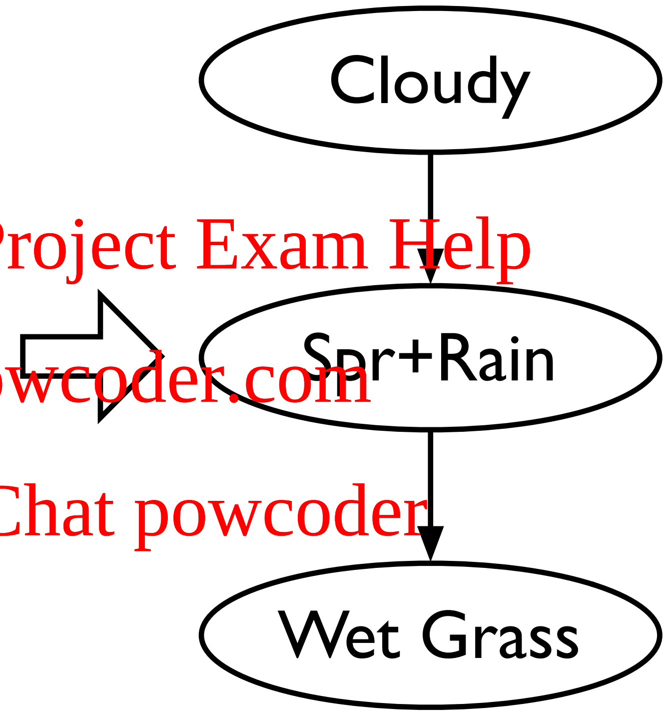
C	P(S|C)
T	0.1
F	0.5

C	P(R|C)
T	0.8
F	0.2

S	R	P(W|S,R)
T	T	0.99
T	F	0.9
F	T	0.9
F	F	0.01

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

P(C)
0.5



	P(S+R=x|C)			
C	TT	TF	FT	FF
T	0.08	0.02	0.72	0.18
F	0.1	0.4	0.1	0.4

S+R	P(W|S+R)
T+T	0.99
T+F	0.9
F+T	0.9
F+F	0.01

Can take exponential time to construct CPTs
But approximate algorithms usu. give reasonable solutions.

So what do we do?

- The are special cases of Bayes nets for which there are fast, exact algorithms:
  - variable elimination
  - belief propagation
- There are also many approximations:
  - stochastic (MCMC) approximations
  - approximations

Assignment Project Exam Help

https://powcoder.com

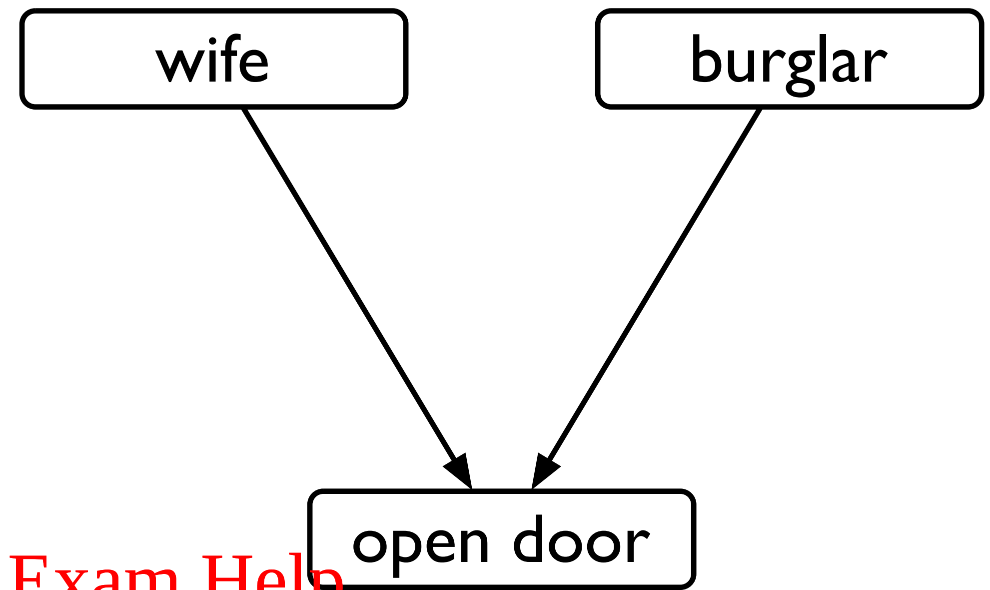
Add WeChat powcoder

The complexity of multi-cause models

- In the models above, we specified the joint conditional density by hand.
- This specified the probability of a variable given each possible value of the causal nodes.
- Can this be specified in a more generic way?
- Can we avoid having to specify every entry in the joint conditional pdf?
- For this we need to specify:

  P(X | parents(X))

- The number of parameters (table entries) scales exponentially with the number of causes



W	B	P(O|W,B)
F	F	0.01
F	T	0.25
T	F	0.05
T	T	0.75

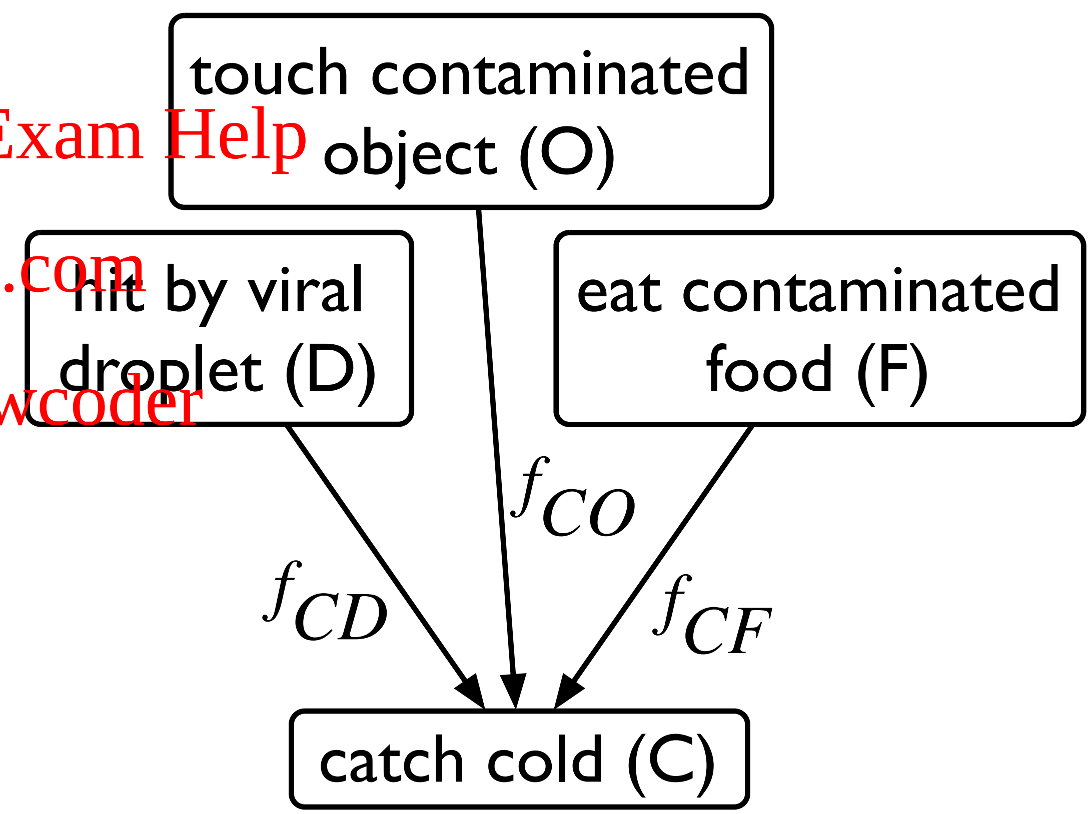
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Beyond tables: modeling causal relationships using Noisy-OR

- We assume each cause C_j can produce effect E_i with probability f_{ij}.
- The noisy-OR model assumes the parent causes of effect E_i contribute independently.
- The probability that none of them caused effect E_i is simply the product of the probabilities that each one *did not* cause E_i.
- The probability that any of them caused E_i is just one minus the above, i.e.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

$$
\begin{aligned}
P(E_i|\mathrm{par}(E_i)) &= P(E_i|C_1,\ldots,C_n) \\
&= 1 - \prod_i (1 - P(E_i|C_j)) \\
&= 1 - \prod_i (1 - f_{ij})
\end{aligned}
$$



$$
P(C|D,O,F) = 1 - (1-f_{CD})(1-f_{CO})(1-f_{CF})
$$

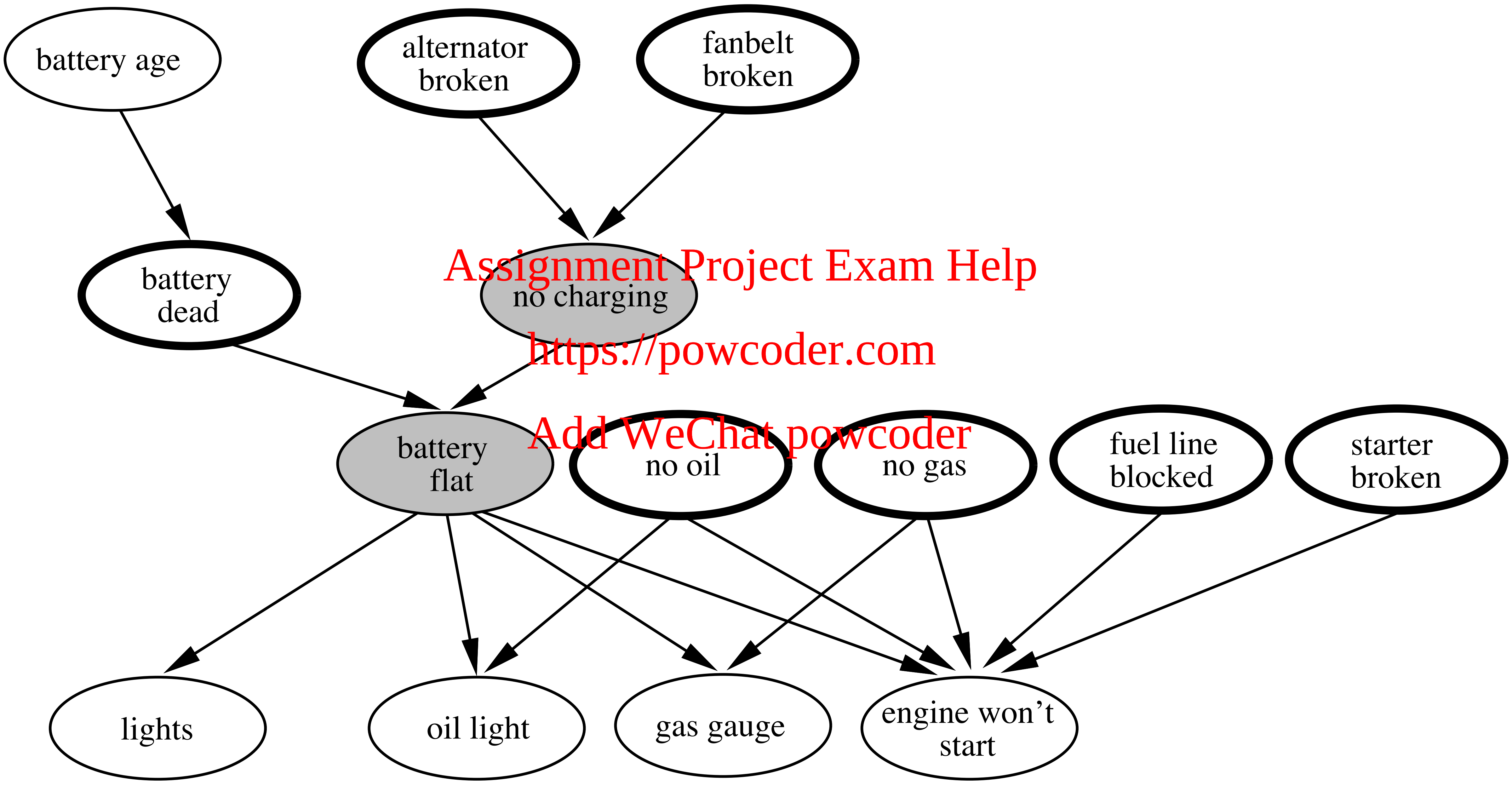
Another noisy-OR example

Table 2. Conditional probability table for $P(Fever \mid Cold, Flu, Malaria)$, as calculated from the noisy-OR model.

Cold	*Flu*	*Malaria*	$P(Fever)$	$P(\neg Fever)$
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	0.6
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

A more complex model with noisy-OR nodes

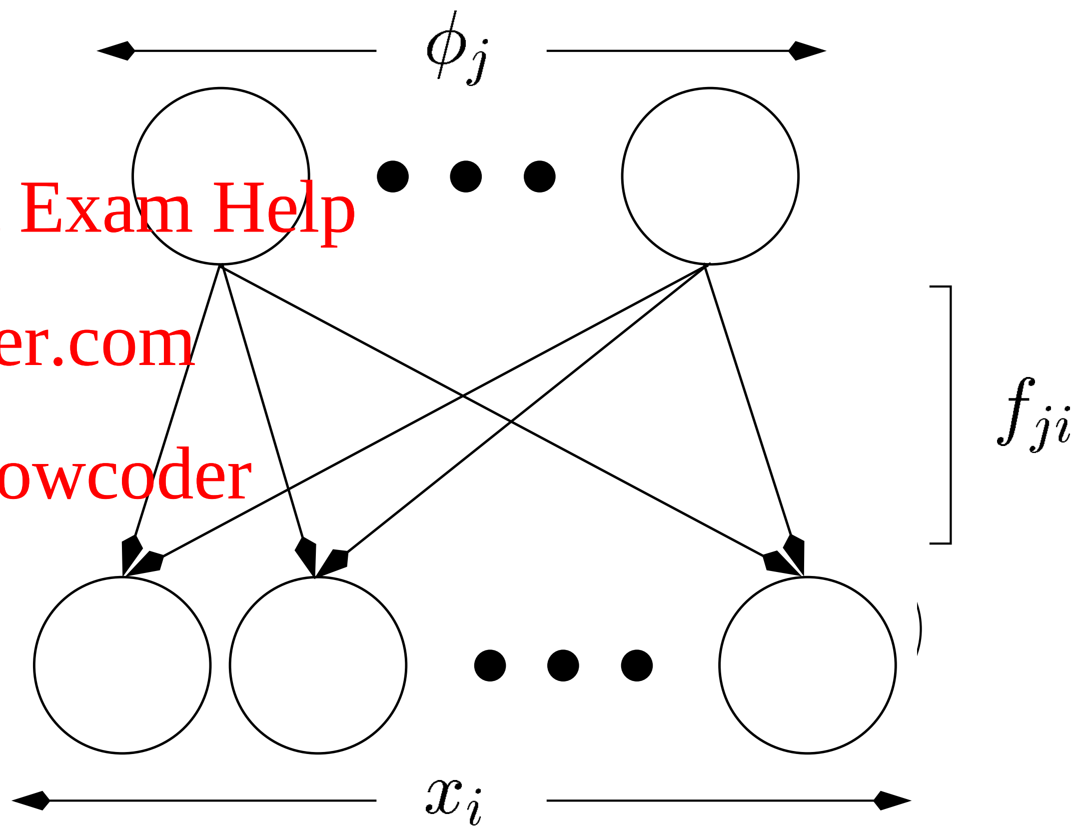


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

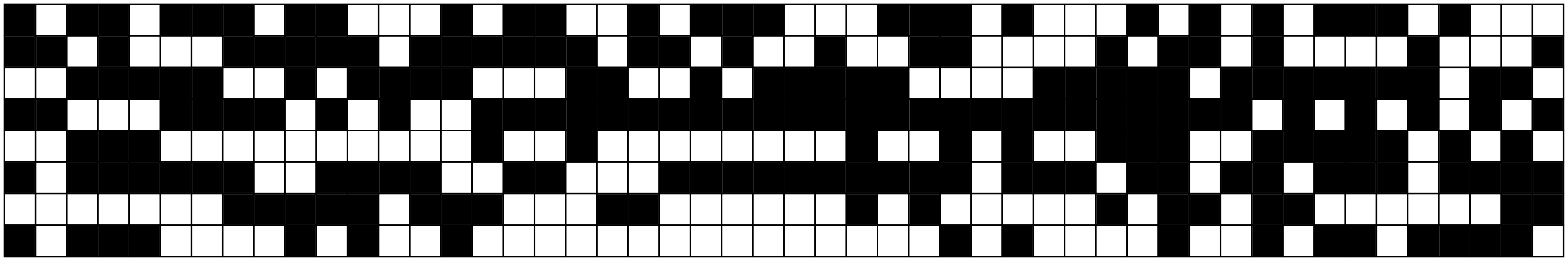
A general one-layer causal network

- Could either model causes and effects
- Or equivalently stochastic binary features.
- Each input x_i encodes the probability that the ith binary input feature is present.
- The set of features represented by ϕj is defined by weights f_{ij} which encode the probability that feature i is an instance of ϕ_j.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



The data: a set of stochastic binary patterns



Each column is a distinct eight-dimensional binary feature.

There are five underlying causal feature patterns.

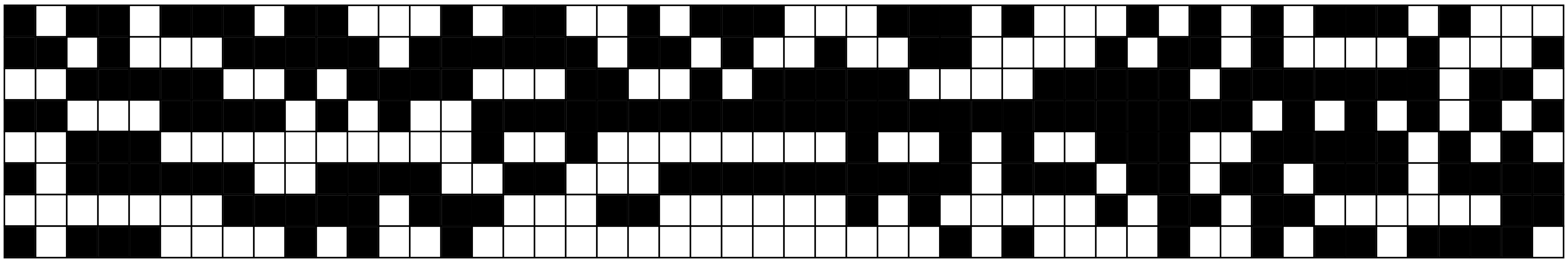
What are they?

Assignment Project Exam Help

https://powcoder.com

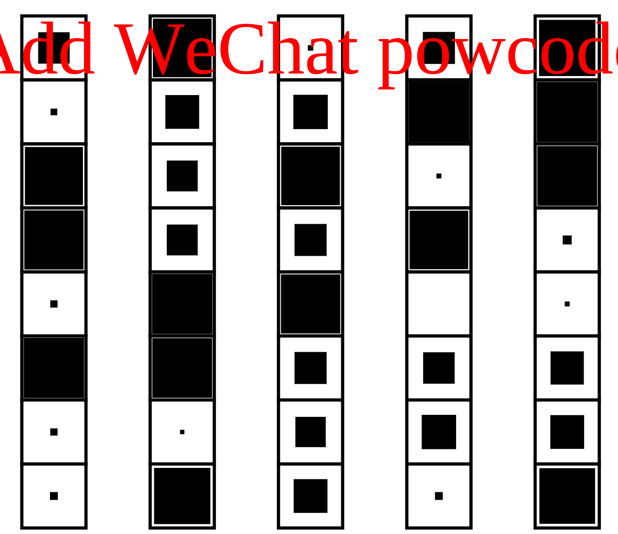
Add WeChat powcoder

The data: a set of stochastic binary patterns



Each column is a distinct eight-dimensional binary feature.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

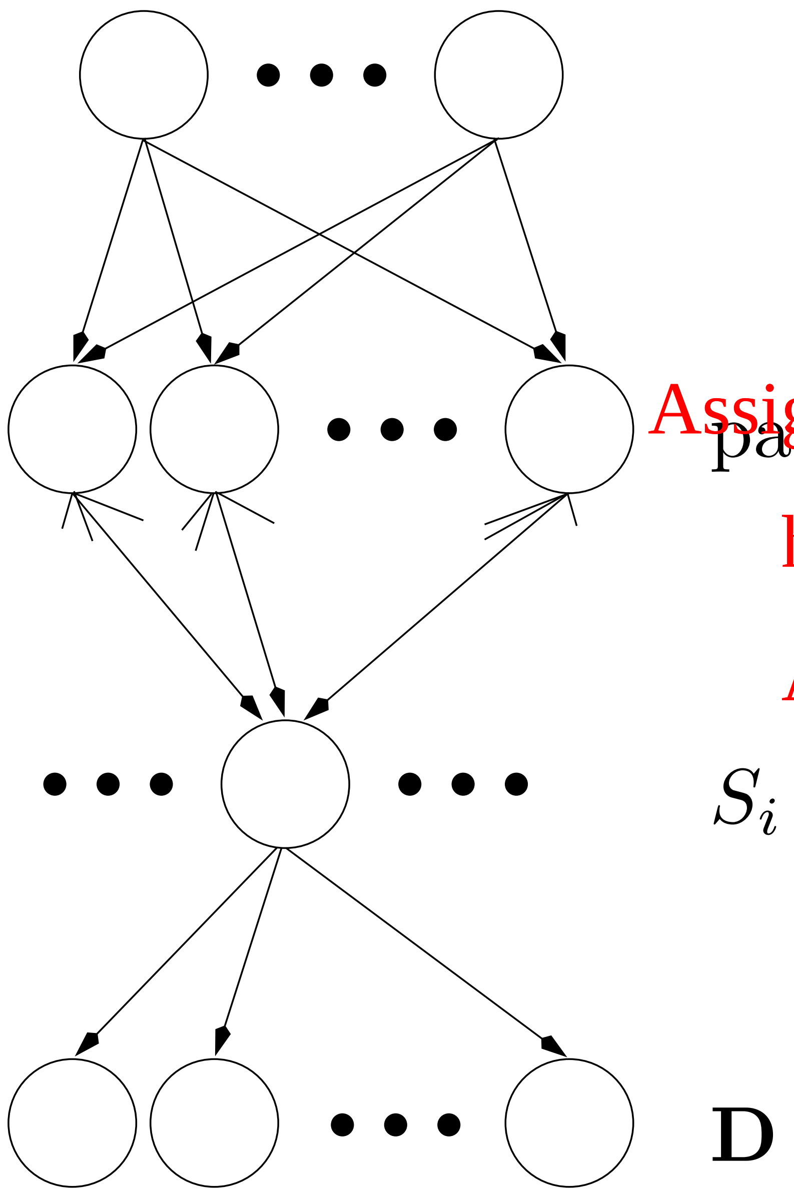


true hidden causes of the data

This is a *learning* problem, which we'll cover in later lecture.

Hierarchical Statistical Models

A Bayesian belief network:



The joint probability of binary states is

$$P(\mathbf{S}|\mathbf{W}) = \prod_i P(S_i|\mathrm{pa}(S_i), \mathbf{W})$$

The probability S_i depends only on its parents:

$$P(S_i|\mathrm{pa}(S_i), \mathbf{W}) = \begin{cases} h(\sum_j S_j w_{ji}) & \text{if } S_i = 1 \\ 1 - h(\sum_j S_j w_{ji}) & \text{if } S_i = 0 \end{cases}$$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

The function h specifies how causes are combined, $h(u) = 1 - \exp(-u)$, $u > 0$.

Main points:

- hierarchical structure allows model to form high order representations
- upper states are priors for lower states
- weights encode higher order features