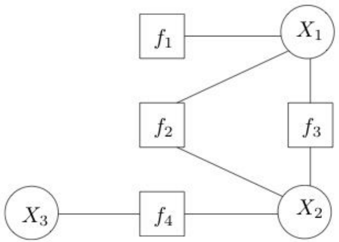
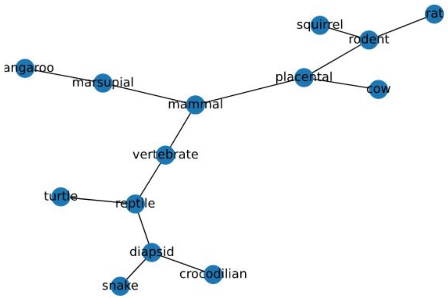
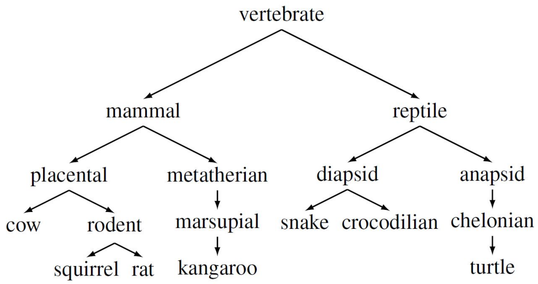




Graphical Models
Belief Propagation Algorithm

Clément Bonnet, Charbel-Raphaël Segerie



March 31st, 2021

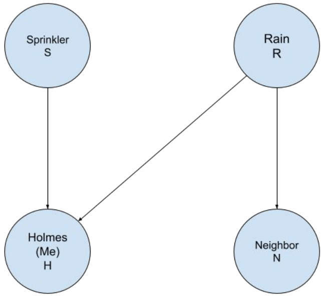
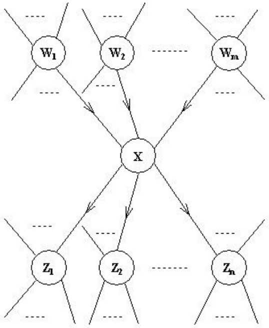
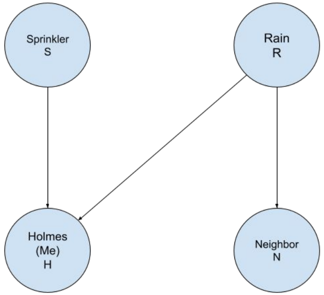


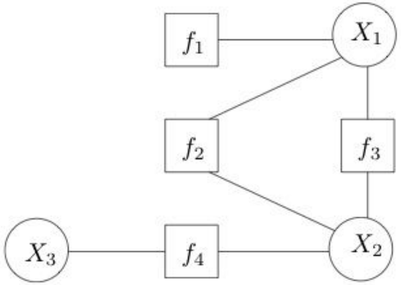
Table of Contents

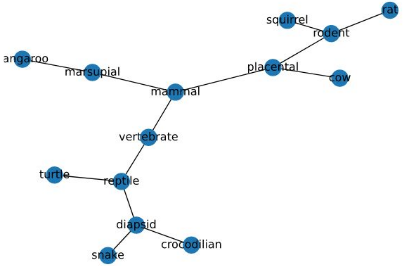
1. Belief Propagation on Bayesian Networks
    a. Theory
    b. Case Study



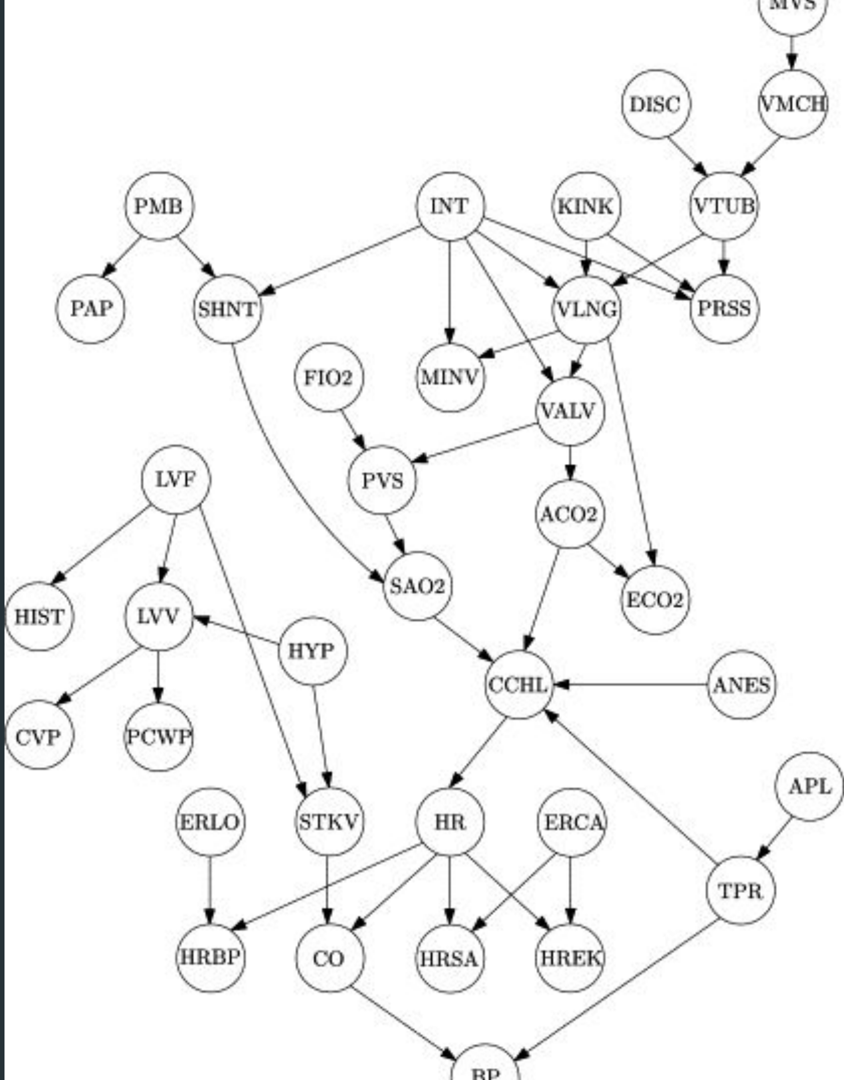


2. Belief Propagation on Factor graphs
    a. Theory
    b. Taxonomy Induction

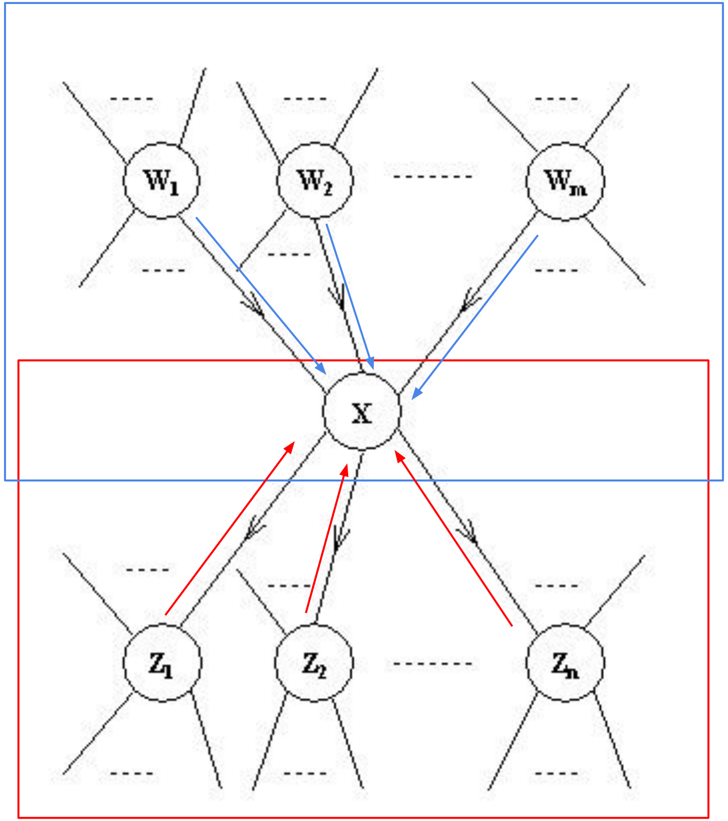




Bayesian Networks



Bayesian Networks: Propagation Rules



Acyclic graph, but messages flow in both direction.

message Child -> Parent

Likelihood $\lambda(x) = \mathbb{P}(Z_1, \ldots, Z_n | X = x) = \prod_{j=1}^{n} \lambda_{Z_j}(x)$

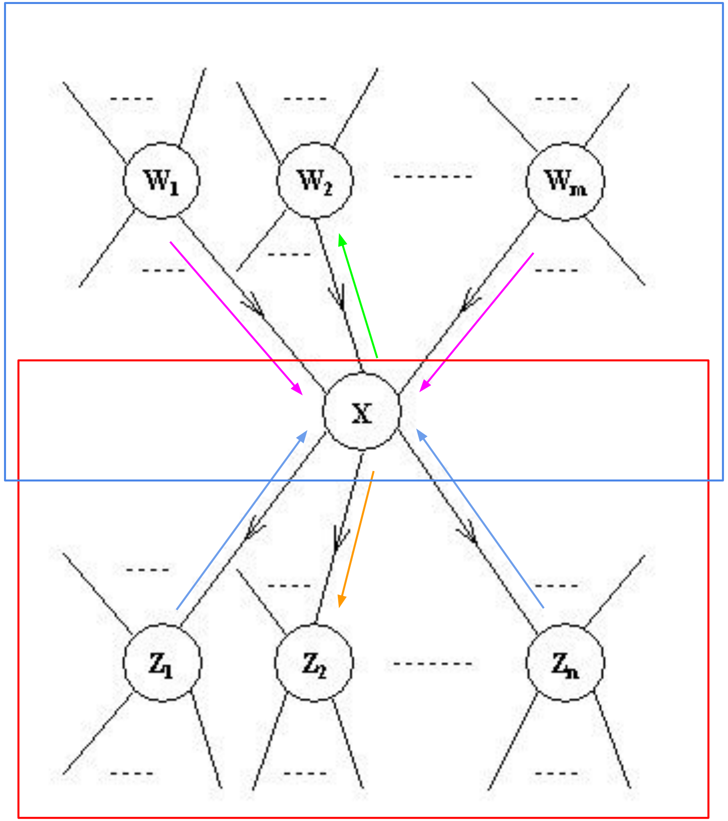
message Parent-> Child

Prior $\pi(x) = \sum_{\mathbf{w} \in \{0,1\}^m} \mathbb{P}(X = x | \mathbf{w}) \prod_{k=1}^{m} \pi_X(w_k)$

Belief

$$
\begin{aligned}
BEL(x) &\equiv \alpha\lambda(x)\pi(x) \\
&= \alpha \boxed{\left[\prod_{j=1}^{n} \lambda_{Z_j}(x)\right]} \boxed{\left[\sum_{\mathbf{w}} P(x \mid \mathbf{w}) \prod_{k=1}^{n} \pi_X(w_k)\right]}
\end{aligned}
$$

Bayesian Networks: Updates



Model of the world specified by the user

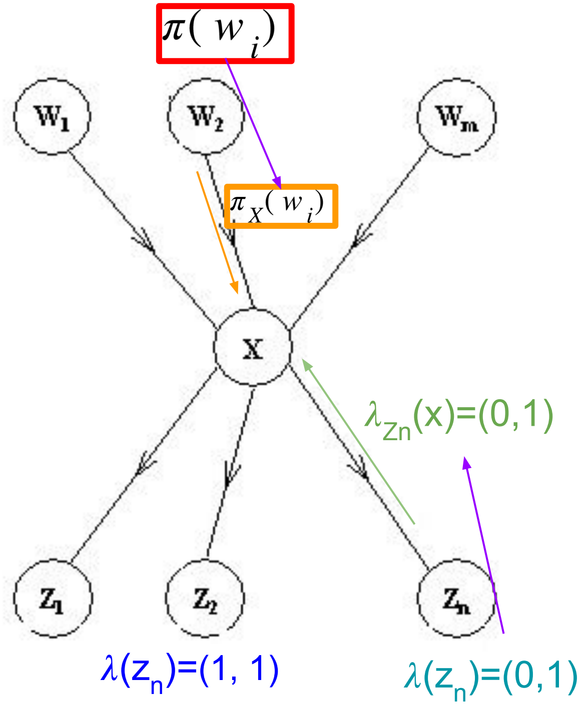
Likelihood:

$$\overrightarrow{\lambda_X(w_i)} \propto \sum_x \boxed{\lambda(x)} \sum_{\substack{\mathbf{w}' \in \{0,1\}^m \\ w'_i = w_i}} \mathbb{P}(X = x|\mathbf{w}') \prod_{k \neq i} \overrightarrow{\pi_X(w'_k)}$$

Prior:

$$\overrightarrow{\pi_{Z_j}(x)} \propto \boxed{\pi(x)} \prod_{k \neq j} \overrightarrow{\lambda_{Z_k}(x)}$$

Bayesian Networks: Boundary Conditions



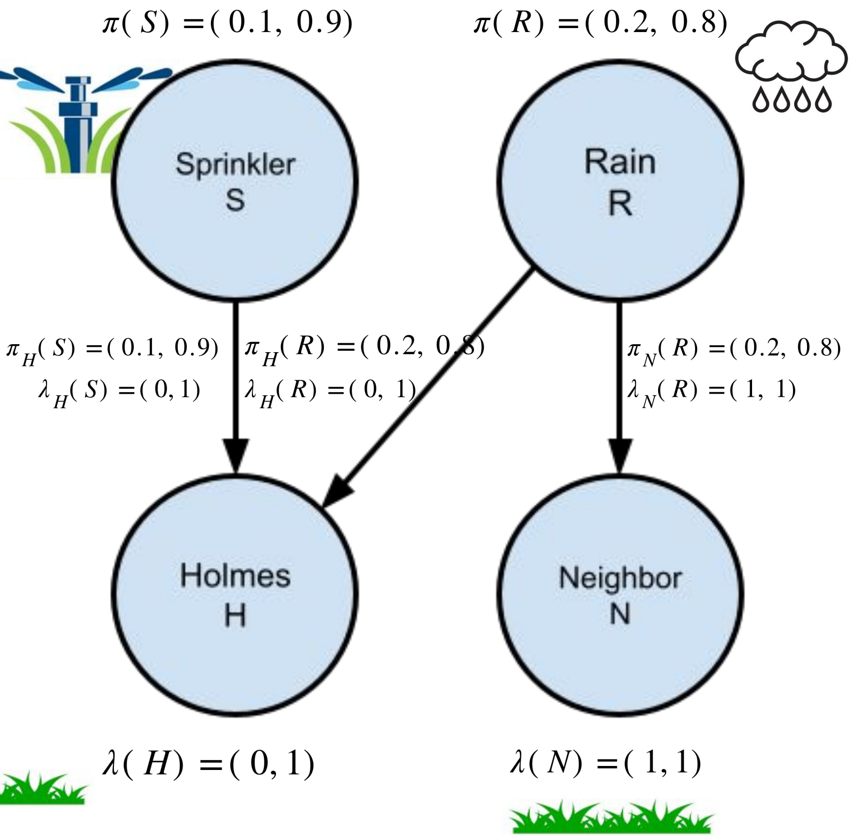
Example of binary states

Root nodes: If W is a node with no parents, we set the value of the prior message $\boldsymbol{\pi}_X$(w) equal to the prior probabilities.

Anticipatory nodes: If X is a childless node that has not been instantiated, we set the λ(z) value as a vector of all 1's

Evidence nodes: If evidence X=i is obtained, we set the λ(z) value to (0,...,0,1,0,...0) with 1 at the i_{th} position, and we set the likelihood message equal to the evidence node.

Bayesian Networks: Case Study → The Wet Grass Example: Priors



$\mathbb{P}(N \mid R)$	$N = 0$	$N = 1$
$R = 0$	0.8	0.2
$R = 1$	0	1

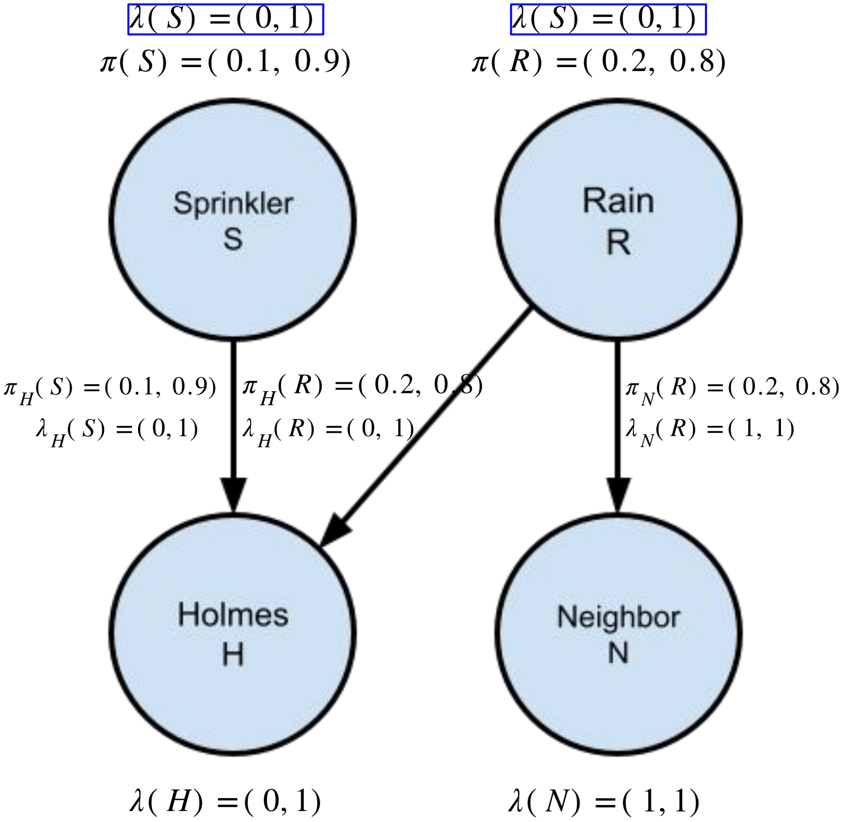
Model of the world

$\mathbb{P}(H \mid R, S)$	$H = 0$	$H = 1$
$R = 1, S = 1$	0	1
$R = 0, S = 1$	0.1	0.9
$R = 1, S = 0$	0	1
$R = 0, S = 0$	1	0

X	$\mathrm{BEL}(x)$	$\pi(x)$	$\lambda(x)$
R	-	(0.2, 0.8)	-
S	-	(0.1, 0.9)	-
N	-	-	(1, 1)
H	-	-	(0, 1)

Boundary Conditions

Bayesian Networks: Case Study → The Wet Grass Example: Likelihood



message Child -> Parent

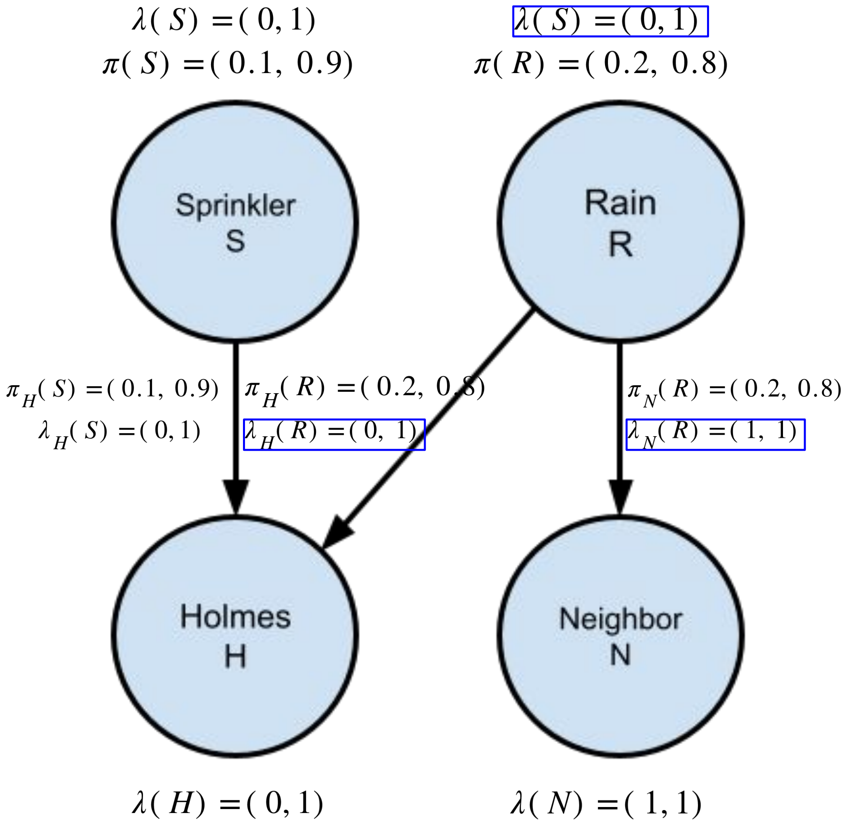
Likelihood $\boxed{\lambda(x) = \mathbb{P}(Z_1, \ldots, Z_n | X = x) = \prod_{j=1}^{n} \lambda_{Z_j}(x)}$

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	-	(0.2, 0.8)	-
S	-	(0.1, 0.9)	-
N	-	-	(1, 1)
H	-	-	(0, 1)

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	(0, 0.8)	(0.2, 0.8)	(0, 1)
S	(0, 0.9)	(0.1, 0.9)	(0, 1)
N	(0.16, 0.84)	(0.16, 0.84)	(1, 1)
H	(0, 0.962)	(0.038, 0.962)	(0, 1)

Propagation of belief

Bayesian Networks: Case Study → The Wet Grass Example: Likelihood



message Child -> Parent

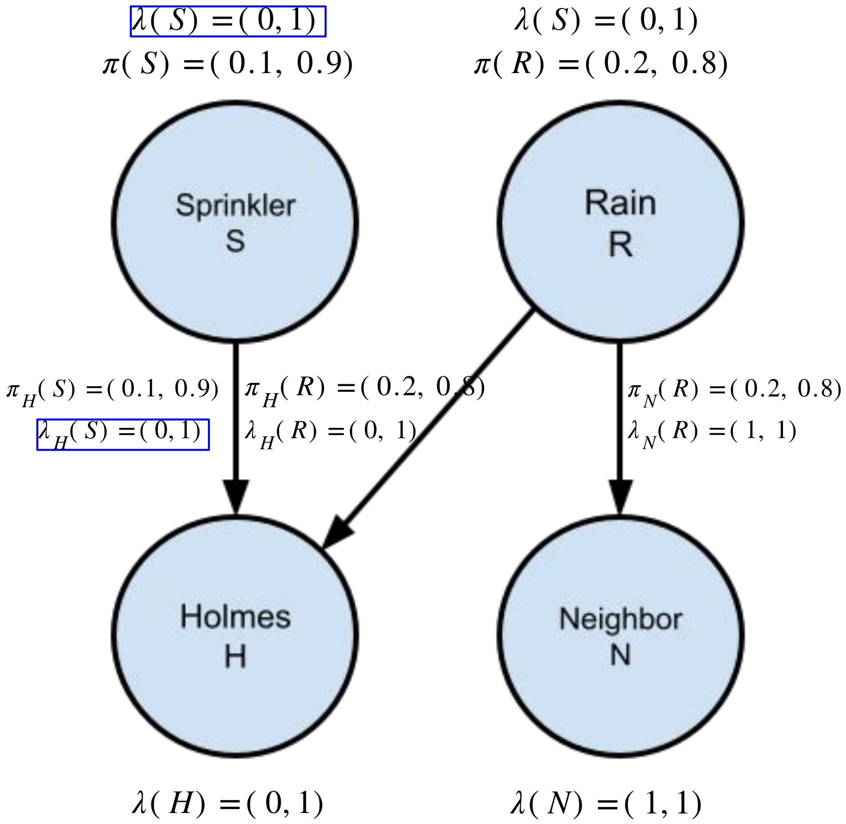
Likelihood $$\boxed{\lambda(x) = \mathbb{P}(Z_1, \ldots, Z_n | X = x) = \prod_{j=1}^{n} \lambda_{Z_j}(x)}$$

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	-	(0.2, 0.8)	-
S	-	(0.1, 0.9)	-
N	-	-	(1, 1)
H	-	-	(0, 1)

Propagation of belief

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	(0, 0.8)	(0.2, 0.8)	(0, 1)
S	(0, 0.9)	(0.1, 0.9)	(0, 1)
N	(0.16, 0.84)	(0.16, 0.84)	(1, 1)
H	(0, 0.962)	(0.038, 0.962)	(0, 1)

Bayesian Networks: Case Study → The Wet Grass Example: Likelihood



message Child -> Parent

Likelihood $$\boxed{\lambda(x) = \mathbb{P}(Z_1, \ldots, Z_n | X = x) = \prod_{j=1}^{n} \lambda_{Z_j}(x)}$$

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	-	(0.2, 0.8)	-
S	-	(0.1, 0.9)	-
N	-	-	(1, 1)
H	-	-	(0, 1)

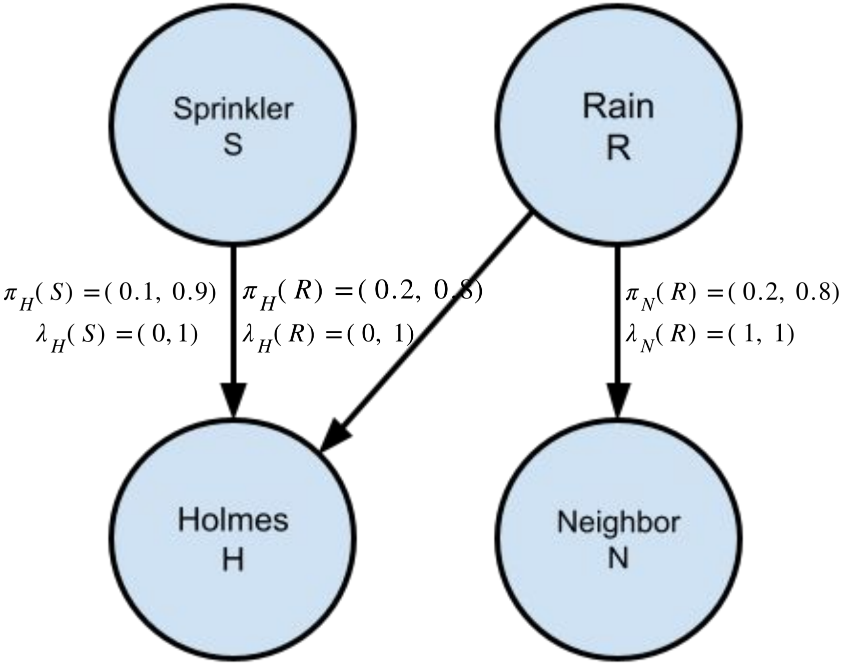
Propagation of belief

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	(0, 0.8)	(0.2, 0.8)	(0, 1)
S	(0, 0.9)	(0.1, 0.9)	(0, 1)
N	(0.16, 0.84)	(0.16, 0.84)	(1, 1)
H	(0, 0.962)	(0.038, 0.962)	(0, 1)

Bayesian Networks: Case Study: The Wet Grass Example: Priors messages

$\lambda(S) = (0, 1)$
$\pi(S) = (0.1,\ 0.9)$

$\lambda(S) = (0, 1)$
$\pi(R) = (0.2,\ 0.8)$



$\lambda(H) = (0, 1)$
$\boxed{\pi(H) = (0.038,\ 0.962)}$

$\lambda(N) = (1, 1)$
$\boxed{\pi(N) = (0.16,\ 0.84)}$

$\mathbb{P}(N \mid R)$	$N=0$	$N=1$
$R=0$	0.8	0.2
$R=1$	0	1

Model of the world

$\mathbb{P}(H \mid R, S)$	$H=0$	$H=1$
$R=1, S=1$	0	1
$R=0, S=1$	0.1	0.9
$R=1, S=0$	0	1
$R=0, S=0$	1	0

$$\text{Prior}\quad \pi(x) = \sum_{\mathbf{w} \in \{0,1\}^m} \mathbb{P}(X = x \mid \mathbf{w}) \prod_{k=1}^{m} \pi_X(w_k)$$

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	-	(0.2, 0.8)	-
S	-	(0.1, 0.9)	-
N	-	-	(1, 1)
H	-	-	(0, 1)

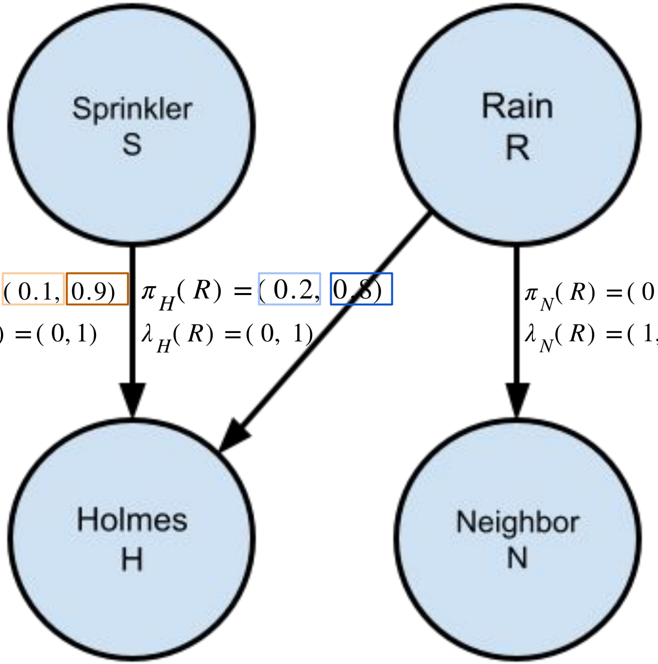
X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	(0, 0.8)	(0.2, 0.8)	(0, 1)
S	(0, 0.9)	(0.1, 0.9)	(0, 1)
N	(0.16, 0.84)	(0.16, 0.84)	(1, 1)
H	(0, 0.962)	(0.038, 0.962)	(0, 1)

Propagation of belief

Bayesian Networks: Case Study: The Wet Grass Example: Priors messages

$\lambda(S) = (0, 1)$
$\pi(S) = (0.1,\ 0.9)$

$\lambda(S) = (0, 1)$
$\pi(R) = (0.2,\ 0.8)$



$\pi_H(S) = (0.1, \boxed{0.9})$
$\lambda_H(S) = (0, 1)$

$\pi_H(R) = (0.2, \boxed{0.8})$
$\lambda_H(R) = (0,\ 1)$

$\pi_N(R) = (0.2,\ 0.8)$
$\lambda_N(R) = (1,\ 1)$

$\lambda(H) = (0, 1)$
$\pi(H) = (\boxed{0.038},\ 0.962)$

$\lambda(N) = (1, 1)$
$\pi(N) = (0.16,\ 0.84)$

$\mathbb{P}(N\|R)$	$N=0$	$N=1$
$R=0$	0.8	0.2
$R=1$	0	1

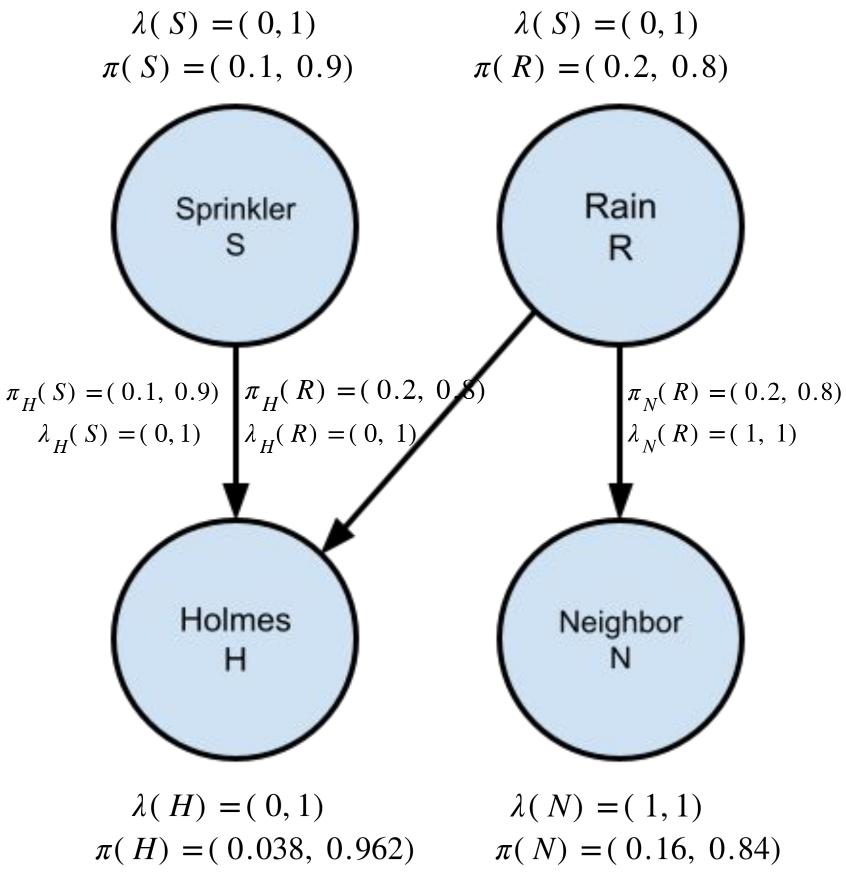
Model of the world

$\mathbb{P}(H\|R,S)$	$H=0$	$H=1$
$R=1, S=1$	0	1
$R=0, S=1$	0.1	0.9
$R=1, S=0$	0	1
$R=0, S=0$	1	0

Prior $\pi(x) = \sum_{\mathbf{w} \in \{0,1\}^m} \mathbb{P}(X=x|\mathbf{w}) \prod_{k=1}^{m} \pi_X(w_k)$

$$
\begin{aligned}
\boxed{\pi(H=0)} &= \mathbb{P}(H=0|R=0,S=0)\pi_H(R=0)\pi_H(S=0) \\
&+ \mathbb{P}(H=0|R=0,S=1)\pi_H(R=0)\pi_H(S=1) \\
&+ \mathbb{P}(H=0|R=1,S=0)\pi_H(R=1)\pi_H(S=0) \\
&+ \mathbb{P}(H=0|R=1,S=1)\pi_H(R=1)\pi_H(S=1) \\
&= 1 \times (0.2 \times 0.1) + 0.1 \times (0.2 \times 0.9) \\
&+ 0 \times (0.8 \times 0.1) + 0 \times (0.8 \times 0.9) \\
&= 0.02 + 0.018 = 0.038
\end{aligned}
$$

Bayesian Networks: Case Study: The Wet Grass Example: Beliefs



$\mathbb{P}(N\|R)$	$N=0$	$N=1$
$R=0$	0.8	0.2
$R=1$	0	1

Model of the world

$\mathbb{P}(H\|R,S)$	$H=0$	$H=1$
$R=1, S=1$	0	1
$R=0, S=1$	0.1	0.9
$R=1, S=0$	0	1
$R=0, S=0$	1	0

Belief $\boxed{BEL(x) \equiv \alpha\lambda(x)\pi(x)}$

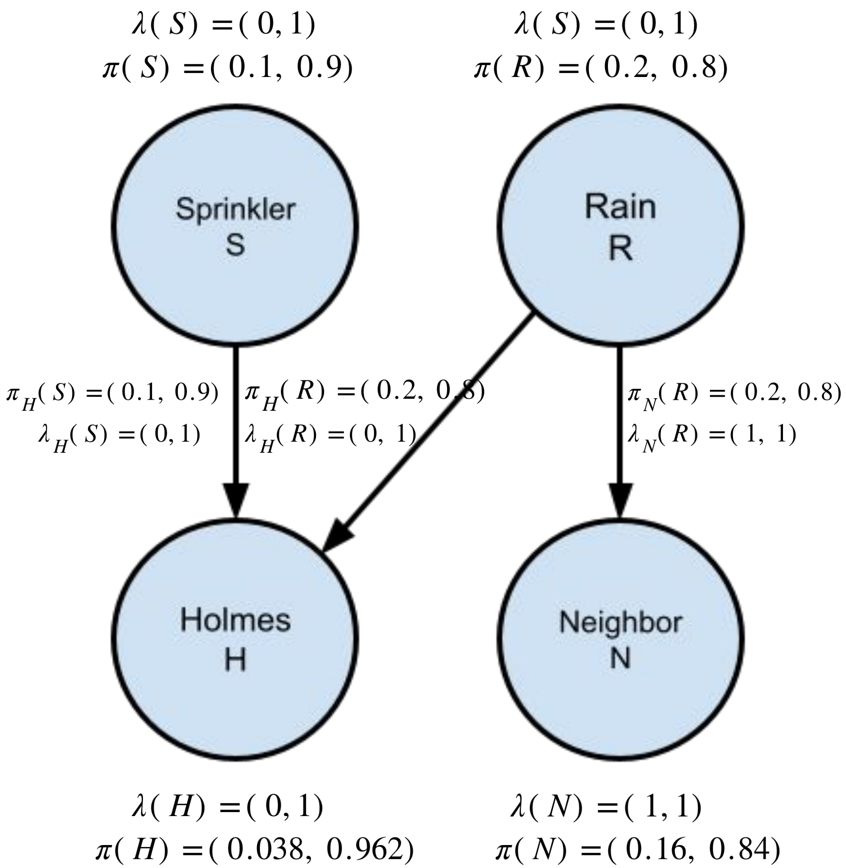
X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	-	(0.2, 0.8)	-
S	-	(0.1, 0.9)	-
N	-	-	(1, 1)
H	-	-	(0, 1)

Product coordinate by coordinate

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	(0, 0.8)	(0.2, 0.8)	(0, 1)
S	(0, 0.9)	(0.1, 0.9)	(0, 1)
N	(0.16, 0.84)	(0.16, 0.84)	(1, 1)
H	(0, 0.962)	(0.038, 0.962)	(0, 1)

Propagation of belief

Bayesian Networks: Case Study: The Wet Grass Example: Beliefs



Model of the world

$\mathbb{P}(N\|R)$	$N = 0$	$N = 1$
$R = 0$	0.8	0.2
$R = 1$	0	1

$\mathbb{P}(H\|R, S)$	$H = 0$	$H = 1$
$R = 1, S = 1$	0	1
$R = 0, S = 1$	0.1	0.9
$R = 1, S = 0$	0	1
$R = 0, S = 0$	1	0

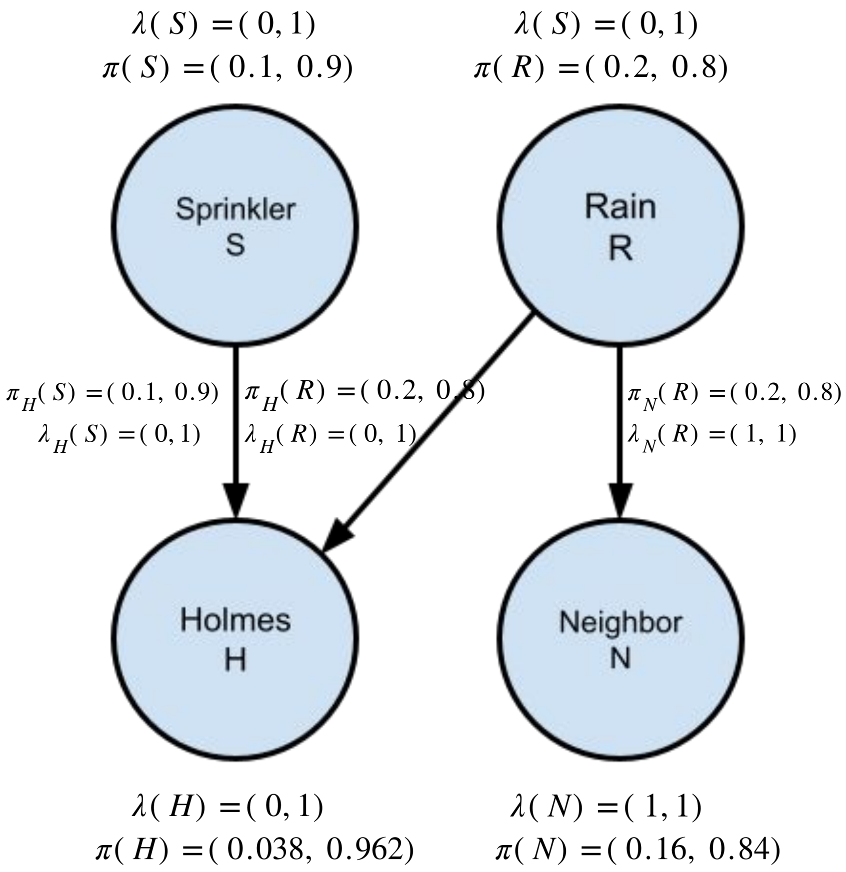
Belief $\boxed{BEL(x) \equiv \alpha\lambda(x)\pi(x)}$

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	-	(0.2, 0.8)	-
S	-	(0.1, 0.9)	-
N	-	-	(1, 1)
H	-	-	(0, 1)

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	(0, 0.8)	(0.2, 0.8)	(0, 1)
S	(0, 0.9)	(0.1, 0.9)	(0, 1)
N	(0.16, 0.84)	(0.16, 0.84)	(1, 1)
H	(0, 0.962)	(0.038, 0.962)	(0, 1)

Propagation of belief

Bayesian Networks: Case Study: The Wet Grass Example: Beliefs



$\mathbb{P}(N\|R)$	$N=0$	$N=1$
$R=0$	0.8	0.2
$R=1$	0	1

Model of the world

$\mathbb{P}(H\|R,S)$	$H=0$	$H=1$
$R=1, S=1$	0	1
$R=0, S=1$	0.1	0.9
$R=1, S=0$	0	1
$R=0, S=0$	1	0

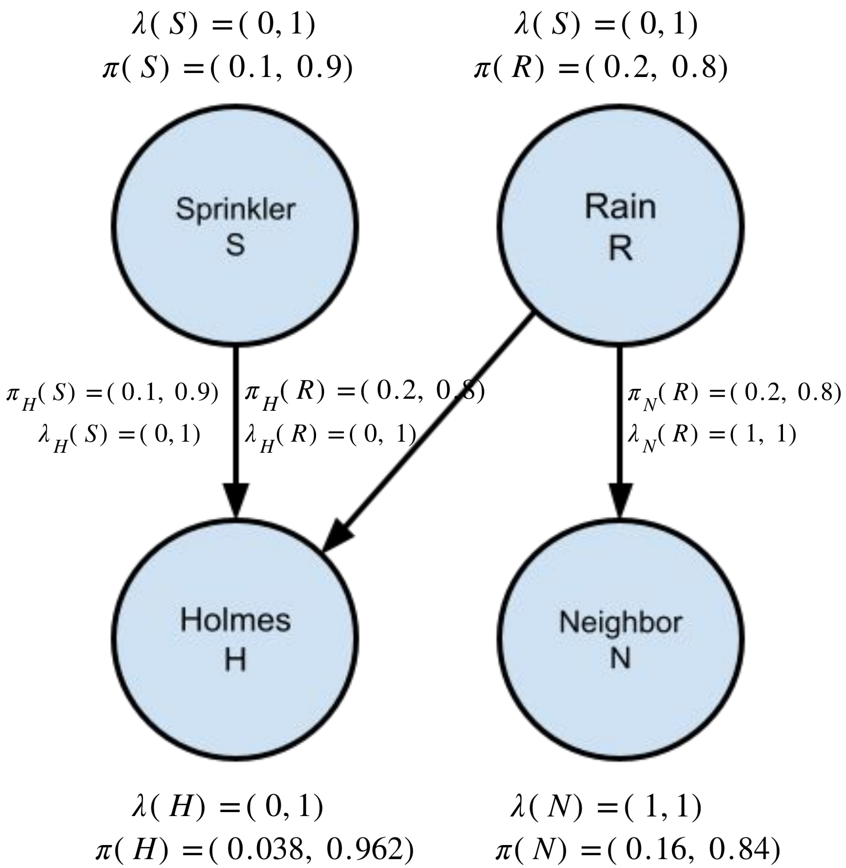
Belief $\boxed{BEL(x) \equiv \alpha\lambda(x)\pi(x)}$

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	-	(0.2, 0.8)	-
S	-	(0.1, 0.9)	-
N	-	-	(1, 1)
H	-	-	(0, 1)

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	(0, 0.8)	(0.2, 0.8)	(0, 1)
S	($\boxed{0,}$ 0.9)	($\boxed{0.1}$, 0.9)	($\boxed{0,}$ 1)
N	(0.16, 0.84)	(0.16, 0.84)	(1, 1)
H	(0, 0.962)	(0.038, 0.962)	(0, 1)

Propagation of belief

Bayesian Networks: Case Study: The Wet Grass Example: Beliefs



$\mathbb{P}(N\|R)$	$N=0$	$N=1$
$R=0$	0.8	0.2
$R=1$	0	1

Model of the world

$\mathbb{P}(H\|R,S)$	$H=0$	$H=1$
$R=1, S=1$	0	1
$R=0, S=1$	0.1	0.9
$R=1, S=0$	0	1
$R=0, S=0$	1	0

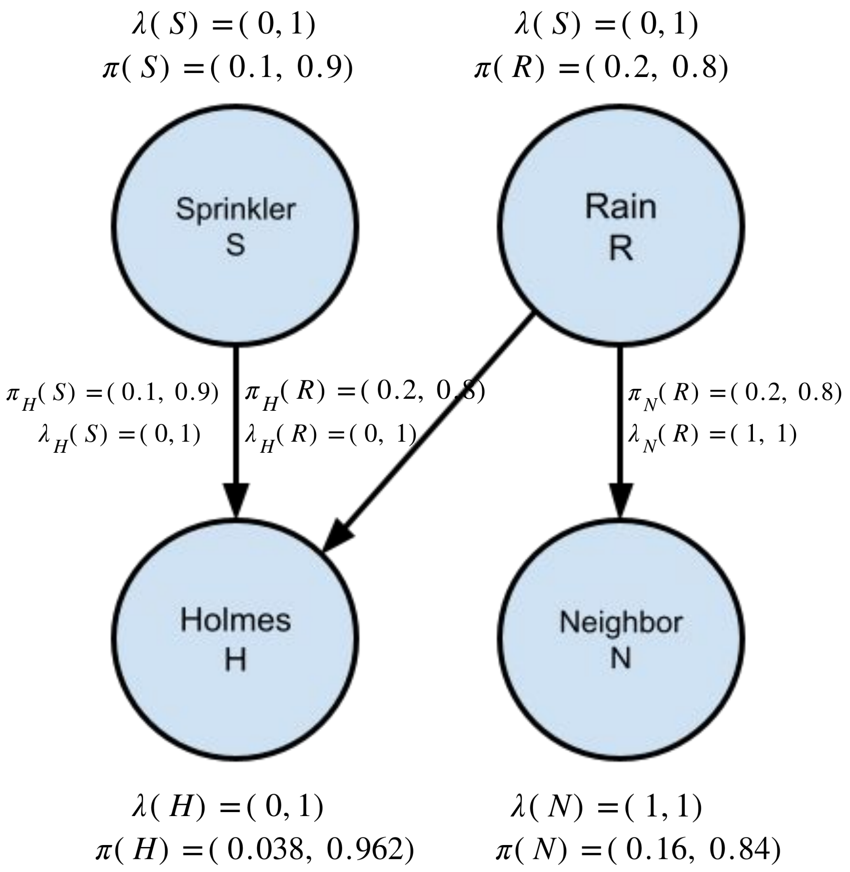
Belief $BEL(x) \equiv \alpha\lambda(x)\pi(x)$

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	-	(0.2, 0.8)	-
S	-	(0.1, 0.9)	-
N	-	-	(1, 1)
H	-	-	(0, 1)

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	(0, 0.8)	(0.2, 0.8)	(0, 1)
S	(0, 0.9)	(0.1, 0.9)	(0, 1)
N	(0.16, 0.84)	(0.16, 0.84)	(1, 1)
H	(0, 0.962)	(0.038, 0.962)	(0, 1)

Propagation of belief

Bayesian Networks: Case Study: The Wet Grass Example: Beliefs



$\mathbb{P}(N\mid R)$	$N=0$	$N=1$
$R=0$	0.8	0.2
$R=1$	0	1

Model of the world

$\mathbb{P}(H\mid R,S)$	$H=0$	$H=1$
$R=1, S=1$	0	1
$R=0, S=1$	0.1	0.9
$R=1, S=0$	0	1
$R=0, S=0$	1	0

Belief $\boxed{BEL(x) \equiv \alpha\lambda(x)\pi(x)}$

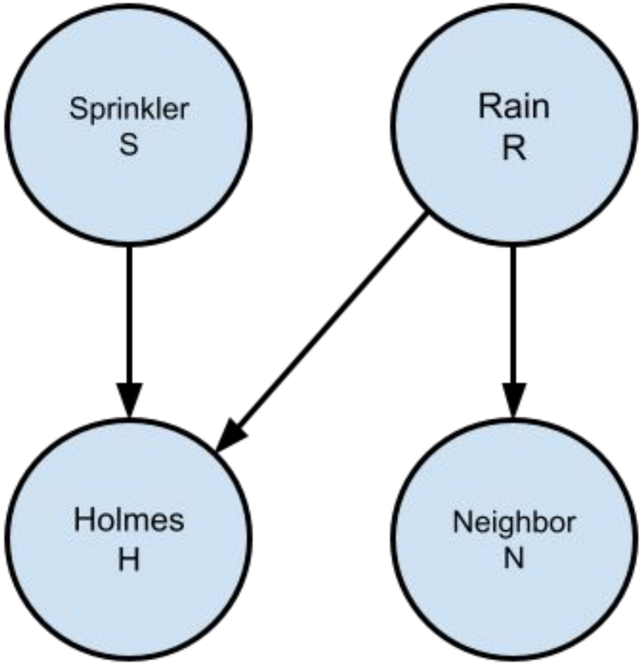
etc...

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	-	(0.2, 0.8)	-
S	-	(0.1, 0.9)	-
N	-	-	(1, 1)
H	-	-	(0, 1)

X	BEL(x)	$\pi(x)$	$\lambda(x)$
R	(0, 0.8)	(0.2, 0.8)	(0, 1)
S	(0, 0.9)	(0.1, 0.9)	(0, 1)
N	(0.16, 0.84)	(0.16, 0.84)	(1, 1)
H	(0, 0.962)	(0.038, 0.962)	(0, 1)

Propagation of belief

Bayesian Networks: Case Study: The Wet Grass Example: Results



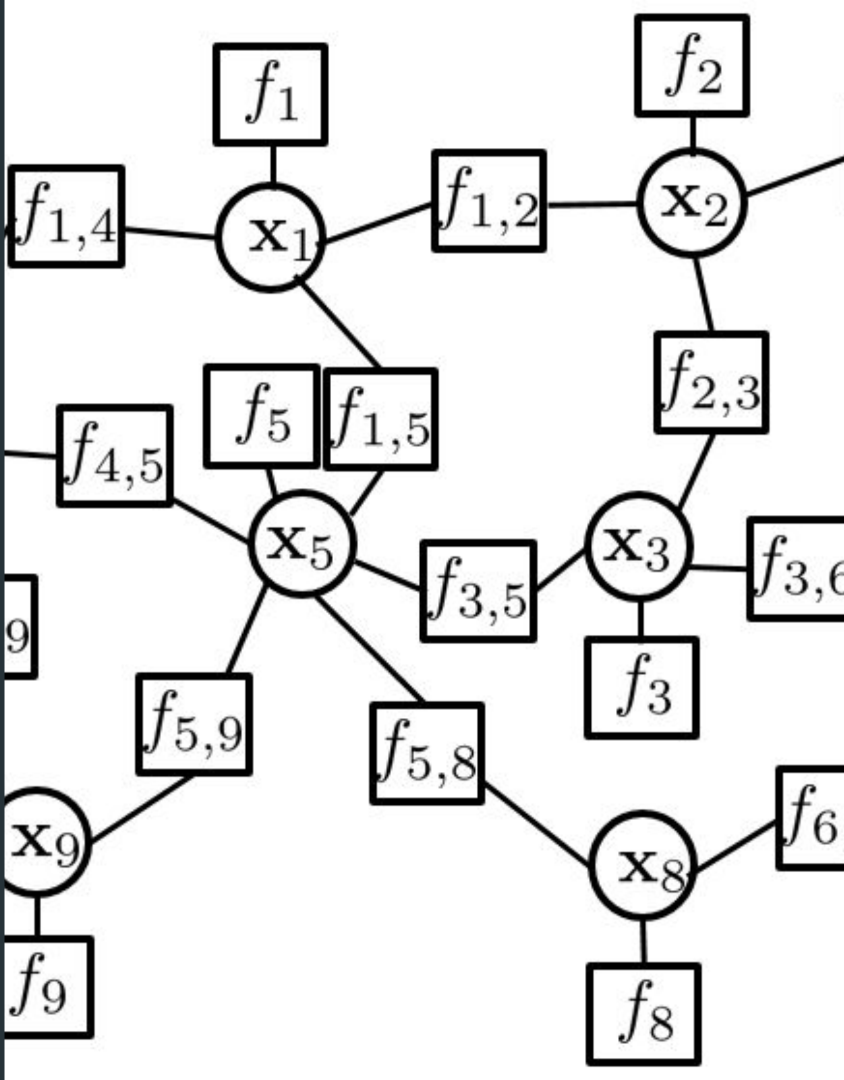
Holmes sees only his grass

Watson's grass is wet: 0.865
Holmes' grass is wet: 1
It has rained: 0.832
The sprinkler was left on: 0.917

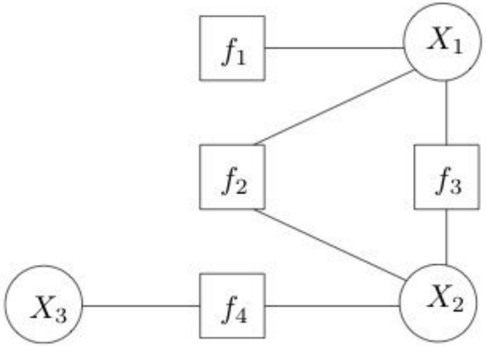
If Holmes also sees his neighbor's wet lawn

Watson's grass is wet: 1
Holmes' grass is wet: 1
It has rained: 0.961
The sprinkler was left on: 0.904

Factor Graphs



Factor Graphs: Theory



Factor Graph = Bipartite Graph

Factor node

Variable node

Total likelihood:

$$p(\mathbf{x}) = \prod_{f \in F} \mathbf{f}_f(\mathbf{x}_f) = f_1(X_1) f_2(X_1, X_2) f_3(X_1, X_2) f_4(X_2, X_3)$$

Propagation:

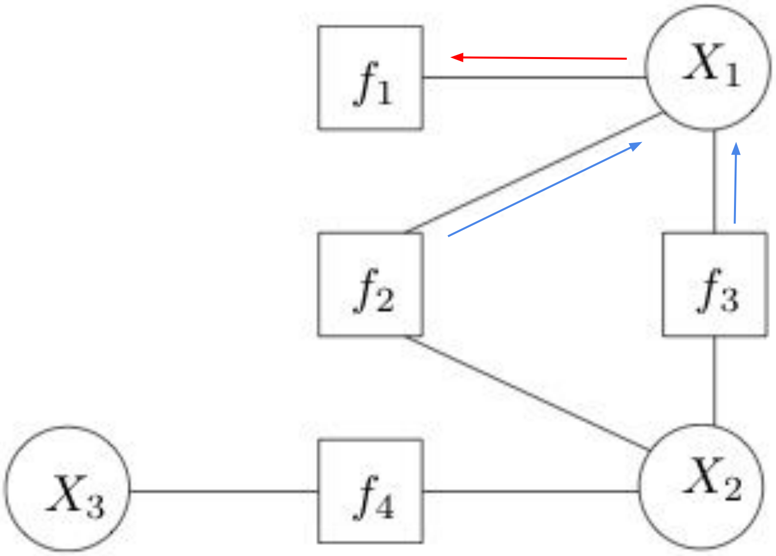
$$\mu_{v \to f}(x_v) = \prod_{f^\star \in N(v) \setminus \{f\}} \mu_{f^\star \to v}(x_v)$$

$$\mu_{f \to v}(x_v) = \sum_{\mathbf{x}'_f : x'_v = x_v} \mathbf{f}_f(\mathbf{x}'_f) \prod_{v^* \in N(f) \setminus \{v\}} \mu_{v^* \to f}(x'_{v^*})$$

Marginals:

$$p_{X_v}(x_v) \propto \prod_{f \in N(v)} \mu_{f \to v}(x_v)$$

Factor Graphs: Example of Propagation: Variable-to-Factor Node



$$\mu_{v \to f}(x_v) = \prod_{f^\star \in N(v) \setminus \{f\}} \mu_{f^\star \to v}(x_v)$$

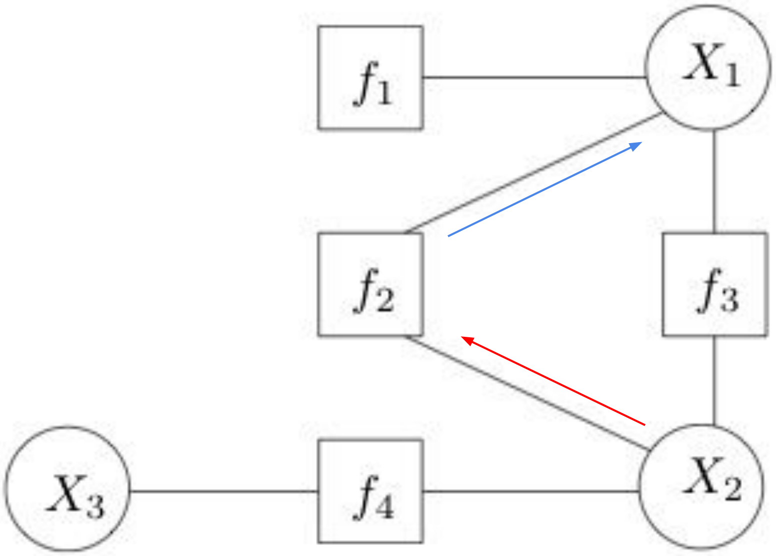
Messages from variable X to factor nodes f are vectors of size NbStates(X)

Messages from factor f to variable X are also vectors of size NbStates(X)

We just multiply coordinate by coordinate the two incoming messages

Factor Graphs: Example of Propagation: Factor-to-Variable Node

$$\overrightarrow{\mu_{f\to v}(x_v)} = \sum_{\mathbf{x}'_f : x'_v = x_v} \mathrm{f}_f(\mathbf{x}'_f) \prod_{v^* \in N(f)\setminus\{v\}} \overrightarrow{\mu_{v^*\to f}(x'_{v^*})}$$



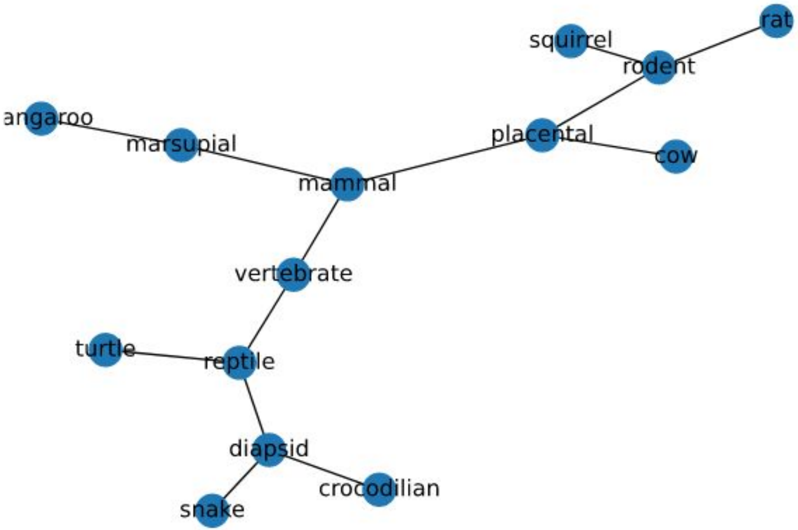
Messages from variable X to factor nodes f are vectors of size NbStates(X)

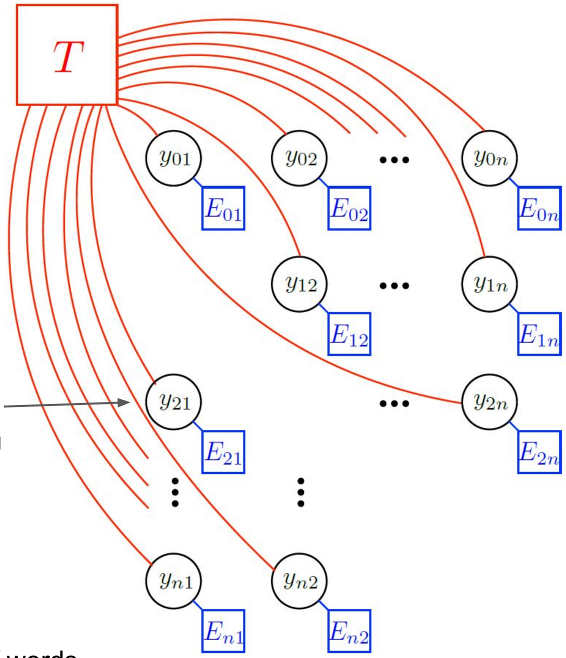
Messages from factor f to variable X are also vectors of size NbStates(X)

We just multiply coordinate by coordinate the incoming messages

Factor Graphs: Taxonomy Induction

From a corpus of abstracts on Wikipedia, and a selection of words (x_1, …, x_n), the objective is to discover the taxonomic structure.





1 if word x_2 is the parent of x_1

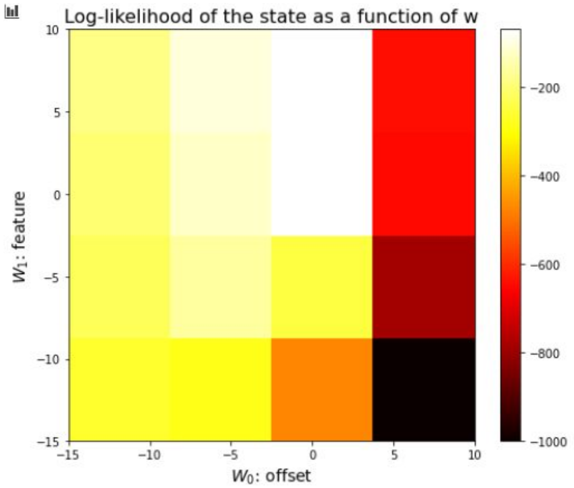
$$P(y|x) \propto \prod_F \phi_F(y)$$

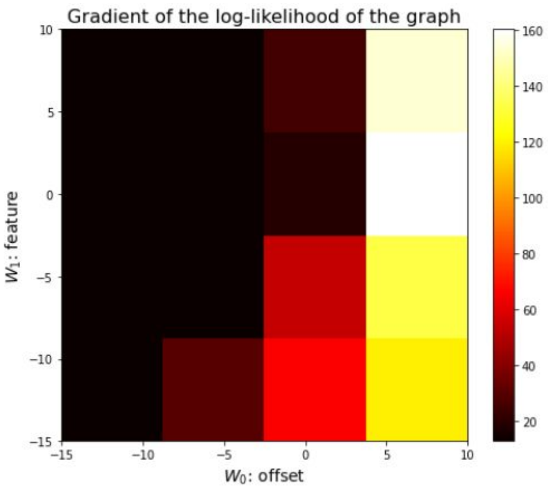
Embedding of the pair of words

$$\boxed{\phi_{E_{i,j}}(y_{i,j})} = \begin{cases} \exp(\mho(x_i, x_j) \cdot w) & \text{if } y_{i,j} = 1 \\ \exp(0) = 1 & \text{if } y_{i,j} = 0 \end{cases}$$

$$\boxed{\phi_T(y_{i,j})} = \begin{cases} 1 & \text{if } y \text{ forms a legal taxonomy tree} \\ 0 & \text{otherwise} \end{cases}$$

Factor Graphs: Taxonomy Induction → Training





- Trained on a taxonomic structure by maximizing its likelihood
- Gradient Descent computed by finite difference
- Trained on a few iterations upon convergence
- Yet, too many edges are drawn

Conclusion

Implemented belief propagation on two types of graphs:

- Bayesian networks → case study with the wet grass example
- Factor graphs → taxonomy induction