Inteligencia Artificial

Inferencia Aproximada en Redes Bayesianas

Jorge Luis Guevara Diaz

www.jorge.sistemasyservidores.com

Que veremos?

- Inferencia exacta por ennumeracion
- Inferencia exacta por eliminacion de variables
- Inferencia aproximada por simulación estocástica
- Inferencia aproximada por método de Monte Carlo y cadenas de Markov

Inferencia aproximada por simulación estocástica

- Idea básica:
  - Realizar N muestras de una distribución de muestras S
  - Calcular la probabilidad a posteriori aproximada $\overline{P}$
  - Mostrar su convergencia a la probabilidad verdadera P
- Algoritmos
  - Muestreo de una red bayesiana
  - Rejection sampling (RS) : rechazar las muestras que no están de acuerdo con la evidencia (rechazo de evidencia)
  - Likelihood weighting (LW) : usa la evidencia para ponderar las muestras (Ponderacion)
  - Markov chain Monte Carlo (MCMC) : Método de Monte Carlo con Cadenas de Markov

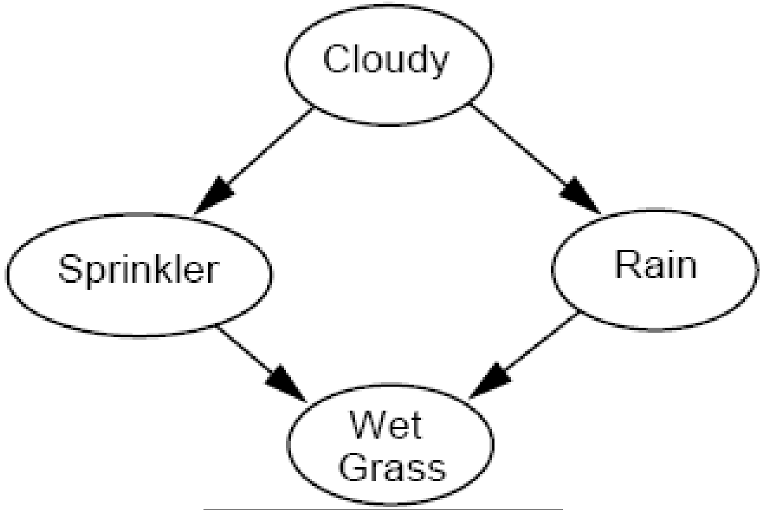
Muestreo de una red bayesiana

function PRIOR-SAMPLE(bn) **returns** an event sampled from bn
  inputs: bn, a belief network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $\mathbf{x} \leftarrow$ an event with n elements
  for $i = 1$ **to** n **do**
    $x_i \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
      given the values of $Parents(X_i)$ in $\mathbf{x}$
  return $\mathbf{x}$

Ejemplo

P(C)
.50

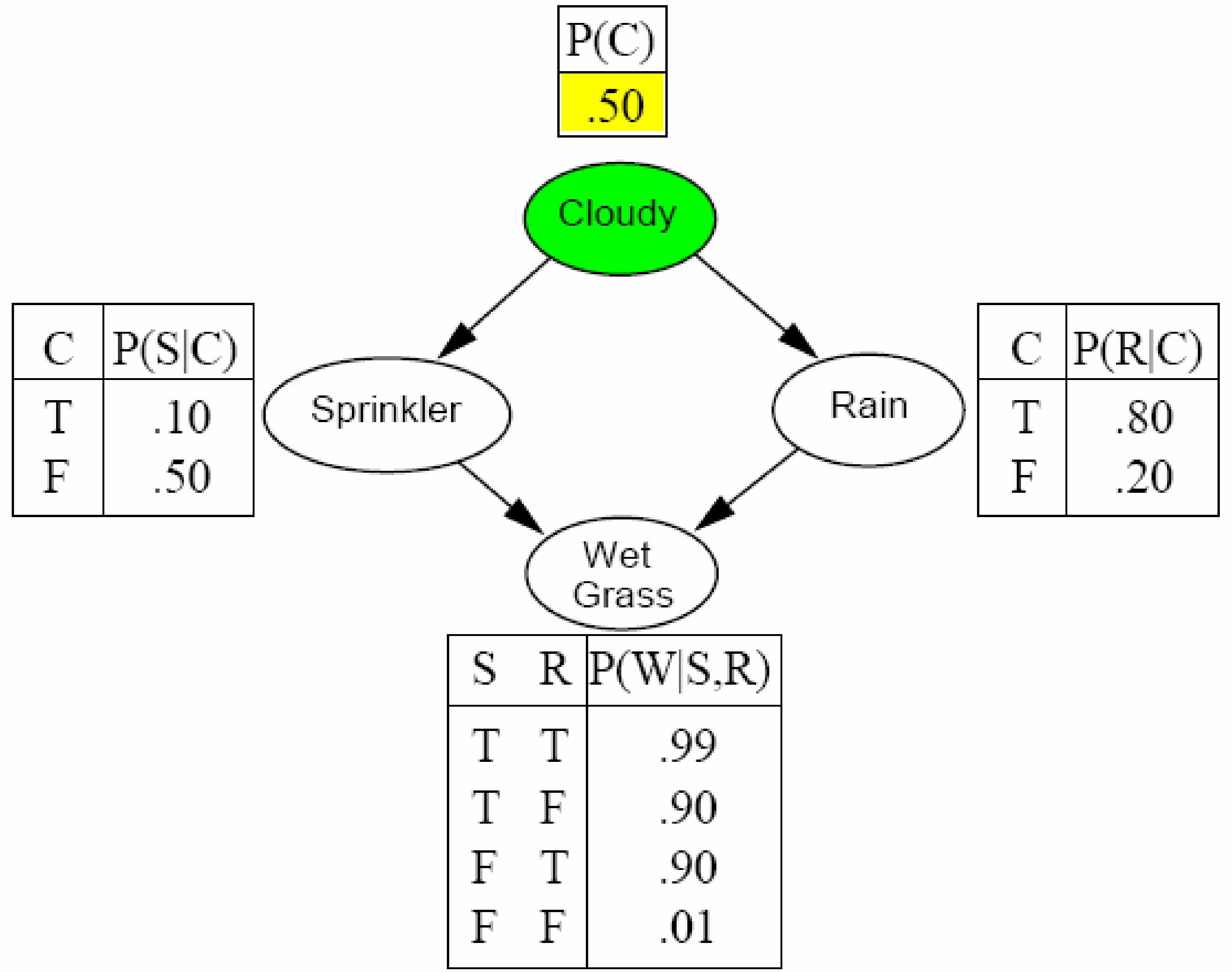


C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

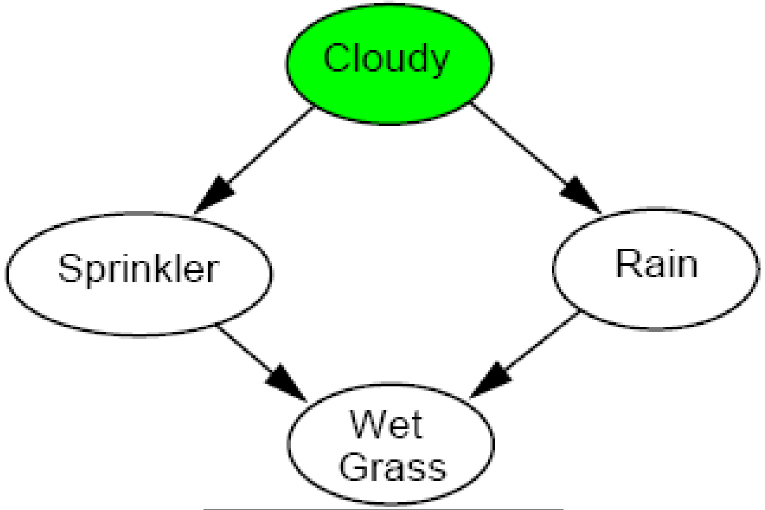
S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Ejemplo


P(C)
.50
Cloudy
C P(S|C)
T .10
F .50
Sprinkler
Rain
C P(R|C)
T .80
F .20
Wet Grass
S R P(W|S,R)
T T .99
T F .90
F T .90
F F .01

Ejemplo

P(C)
.50



C	P(S|C)
T	.10
F	.50

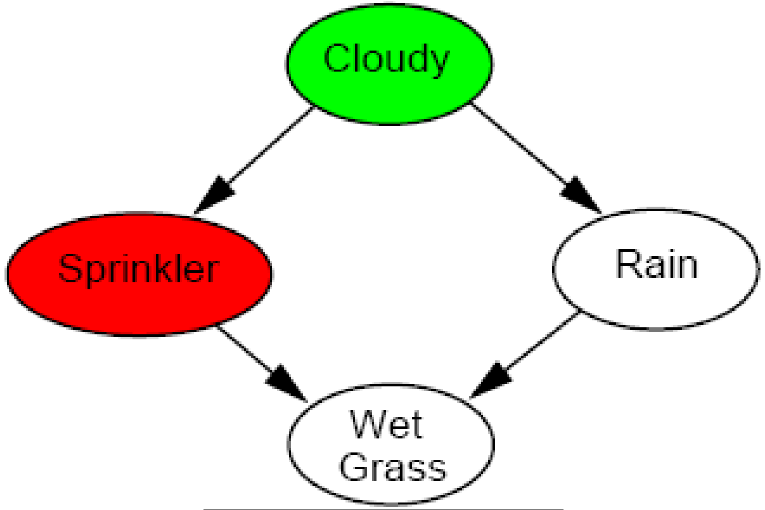
C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Ejemplo

P(C)
.50

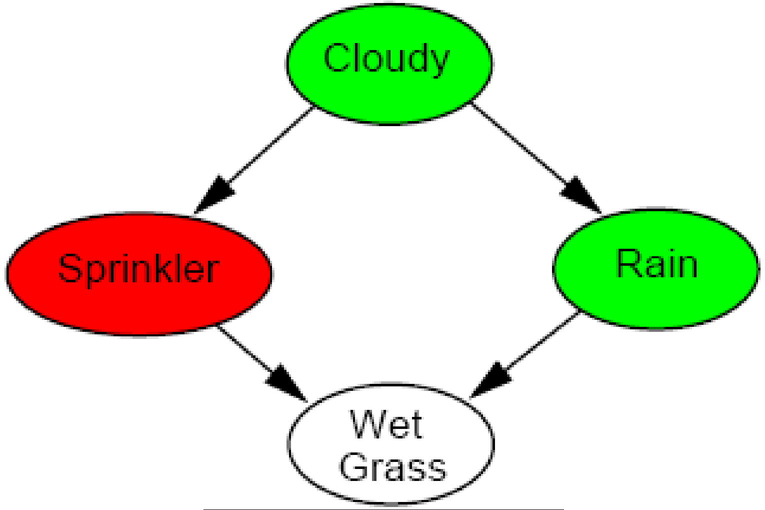
C	P(S|C)
T	.10
F	.50



C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Ejemplo



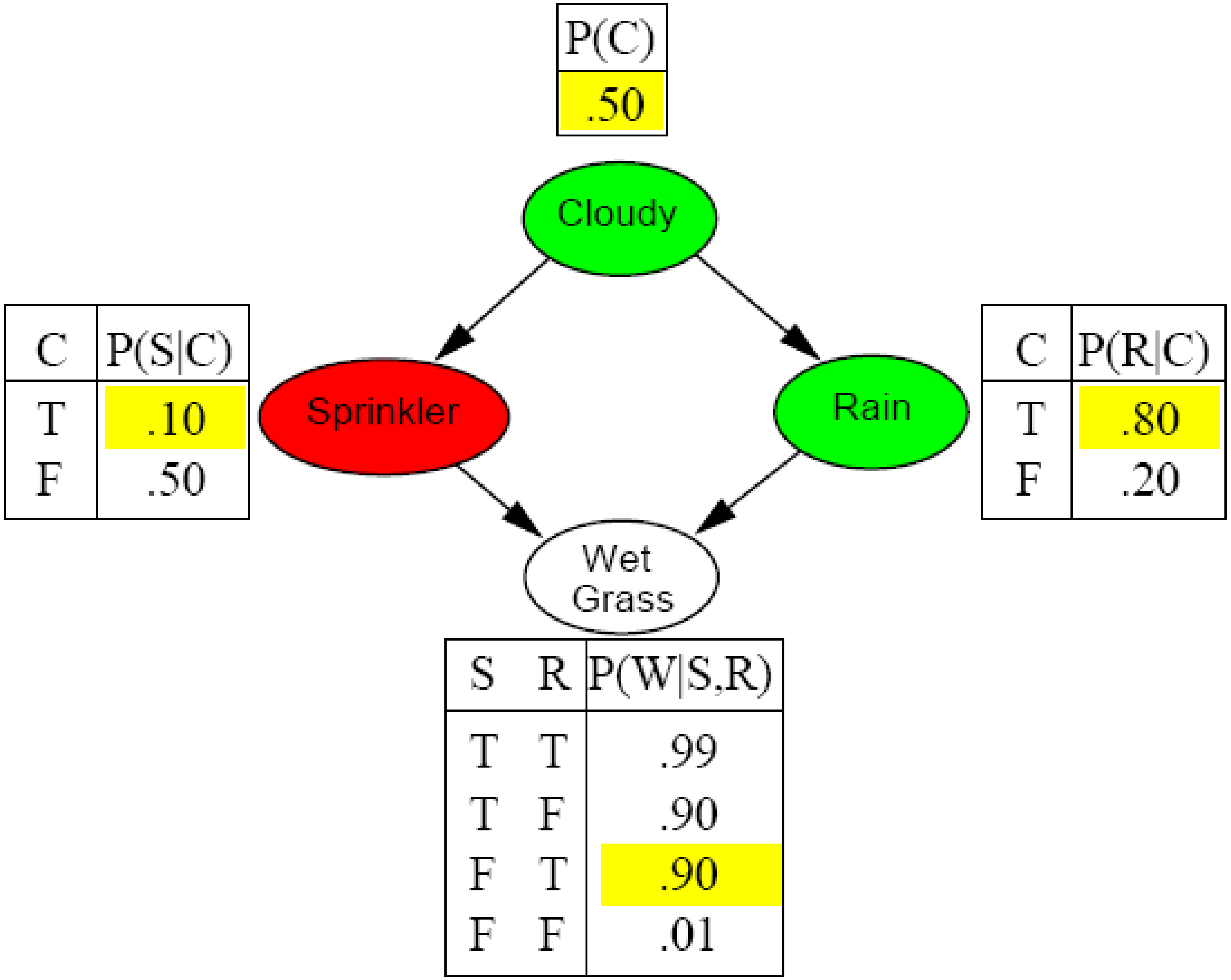
P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Ejemplo



P(C)
.50

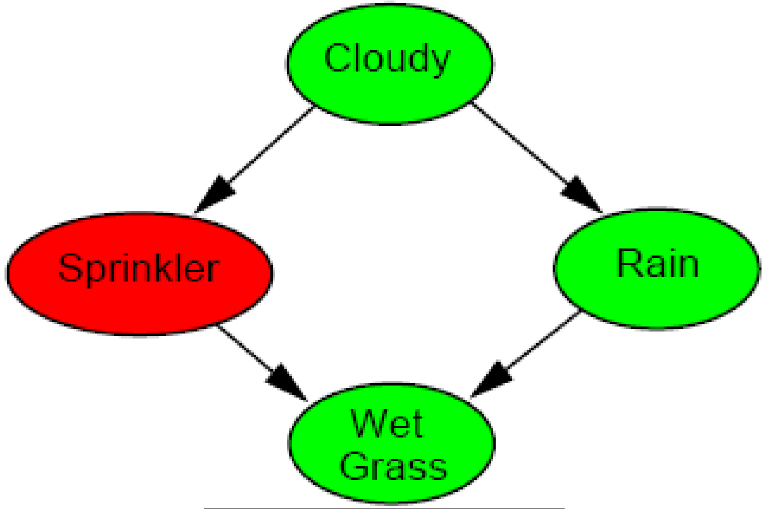
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Ejemplo

P(C)
.50



C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Muestreo de una red bayesiana

- La probabilidad de que PRIORSAMPLE- genera un evento particular
  $S_{PS}(x_1 \dots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i)) = P(x_1 \dots x_n)$

  Es la verdadera probabilidad a priori

  Ejemplo:

  $S_{PS}(t, f, t, t) = 0.5 \times 0.9 \times 0.8 \times 0.9 = 0.324 = P(t, f, t, t)$

  Sea $N_{PS}(x_1 \dots x_n)$ el número de muestras generadas para el evento $x_1, \dots, x_n$

  Entonces:
  $$\begin{aligned} \lim_{N\to\infty} \hat{P}(x_1, \dots, x_n) &= \lim_{N\to\infty} N_{PS}(x_1, \dots, x_n)/N \\ &= S_{PS}(x_1, \dots, x_n) \\ &= P(x_1 \dots x_n) \end{aligned}$$

  Las estimaciones derivadas de PRIOR-SAMPLE son consistentes

  $$\hat{P}(x_1, \dots, x_n) \approx P(x_1 \dots x_n)$$

Muestreo de una red bayesiana

- En

$$S_{PS}(t, f, t, t) = 0.5 \times 0.9 \times 0.8 \times 0.9 = 0.324 = P(t, f, t, t)$$

- Se espera para un N "grande" que el 32.4% de las muestras pertenezcan a este evento

- Ejemplo si se generan 1000 muestras de la red anterior, y 511 tienen rain=true, entonces

$$P_{aprox}(Rain{=}true) = 0.511$$

Algoritmo RS (Rejection sampling) – Muestreo por rechazo de evidencia

- $\hat{\mathbf{P}}(X|\mathbf{e})$ Se calcula a partir de las muestras que están de acuerdo con la evidencia

```
function REJECTION-SAMPLING(X, e, bn, N) returns an estimate of P(X|e)
   local variables: N, a vector of counts over X, initially zero

   for j = 1 to N do
       x ← PRIOR-SAMPLE(bn)
       if x is consistent with e then
           N[x] ← N[x]+1 where x is the value of X in x
   return NORMALIZE(N[X])
```

- Ejemplo: Estimar P(Rain\Sprinkler=true) usando 100 muestras

  27 tienen Sprinkler = true

  8 tienen Rain = true y 19 tienen Rain = false

  P(Rain\Sprinkler=true) = Normalizar (<8,19>) =<0.296, 0.704>

Análisis del Algoritmo RS (Rejection sampling)

- Similar al procedimientos empíricos de estimación en el mundo real

$$\begin{aligned}
\hat{\mathbf{P}}(X|\mathbf{e}) &= \alpha \mathbf{N}_{PS}(X,\mathbf{e}) && \text{(algorithm defn.)} \\
&= \mathbf{N}_{PS}(X,\mathbf{e})/N_{PS}(\mathbf{e}) && \text{(normalized by } N_{PS}(\mathbf{e})\text{)} \\
&\approx \mathbf{P}(X,\mathbf{e})/P(\mathbf{e}) && \text{(property of } \textsc{PriorSample}\text{)} \\
&= \mathbf{P}(X|\mathbf{e}) && \text{(defn. of conditional probability)}
\end{aligned}$$

- Retorna estimaciones posteriores consistentes
- Problema: Muy costoso si P(e) es muy pequeño
- P(e) decrece exponencialmente con el número de variables de evidencia

Algoritmo LW (Likelihood weighting)

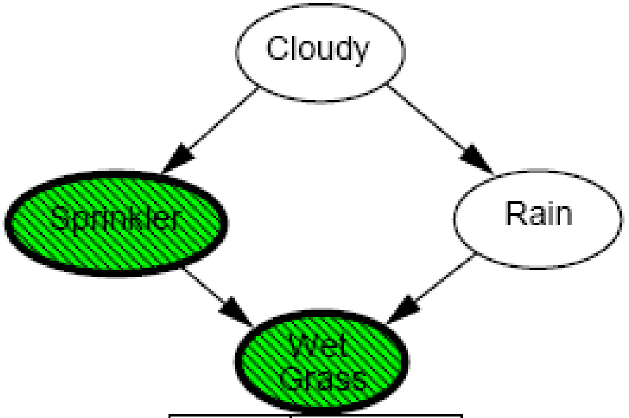
- Idea: Mantener fijas las variables e evidencia y ponderar cada muestra de acuerdo al likelihood de acuerdo a la evidencia

```
function LIKELIHOOD-WEIGHTING(X, e, bn, N) returns an estimate of P(X|e)
   local variables: W, a vector of weighted counts over X, initially zero

   for j = 1 to N do
       x, w ← WEIGHTED-SAMPLE(bn)
       W[x] ← W[x] + w where x is the value of X in x
   return NORMALIZE(W[X])
```

```
function WEIGHTED-SAMPLE(bn, e) returns an event and a weight

   x ← an event with n elements; w ← 1
   for i = 1 to n do
       if X_i has a value x_i in e
           then w ← w × P(X_i = x_i | parents(X_i))
           else x_i ← a random sample from P(X_i | parents(X_i))
   return x, w
```

Algoritmo LW (Likelihood weighting)

- P(Rain\Sprinkler=true, Wet Grass=true)
  - Muestrear P(Cloudy) suponer true
  - Sprinkler es evidencia
    - w = w x P(Sprinkler =true\Cloudy =true) =0.1
  - Muestrear P(Rain \ cloudy =true) suponer retorna true
  - Wet Grass es evidencia
    - w = w x P(Wet Grass=true \Sprinkler = true, Rain =true) = 0.099



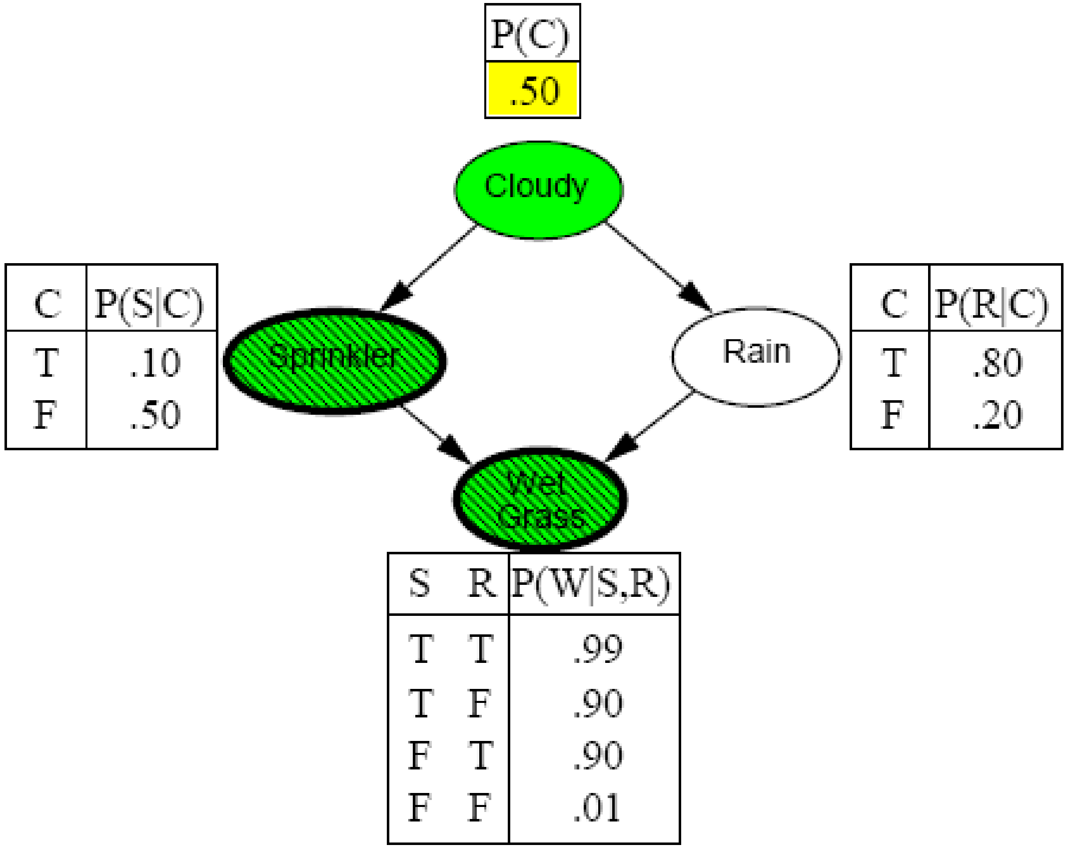
P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$w = 1.0$



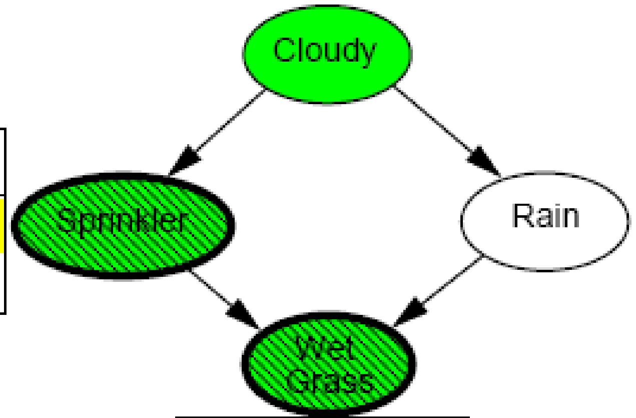
P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$w = 1.0$



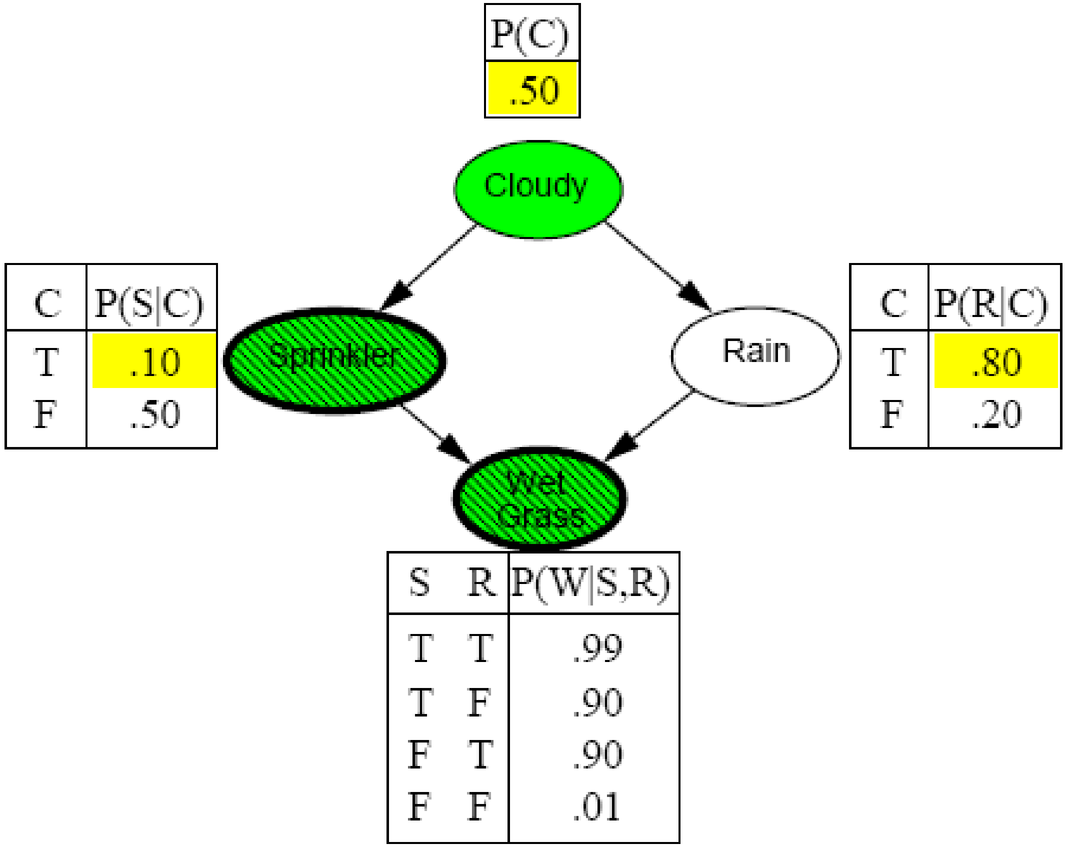
P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$w = 1.0$



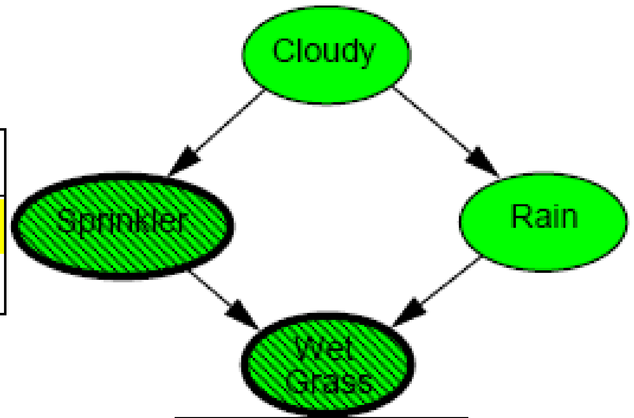
P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$w = 1.0 \times 0.1$



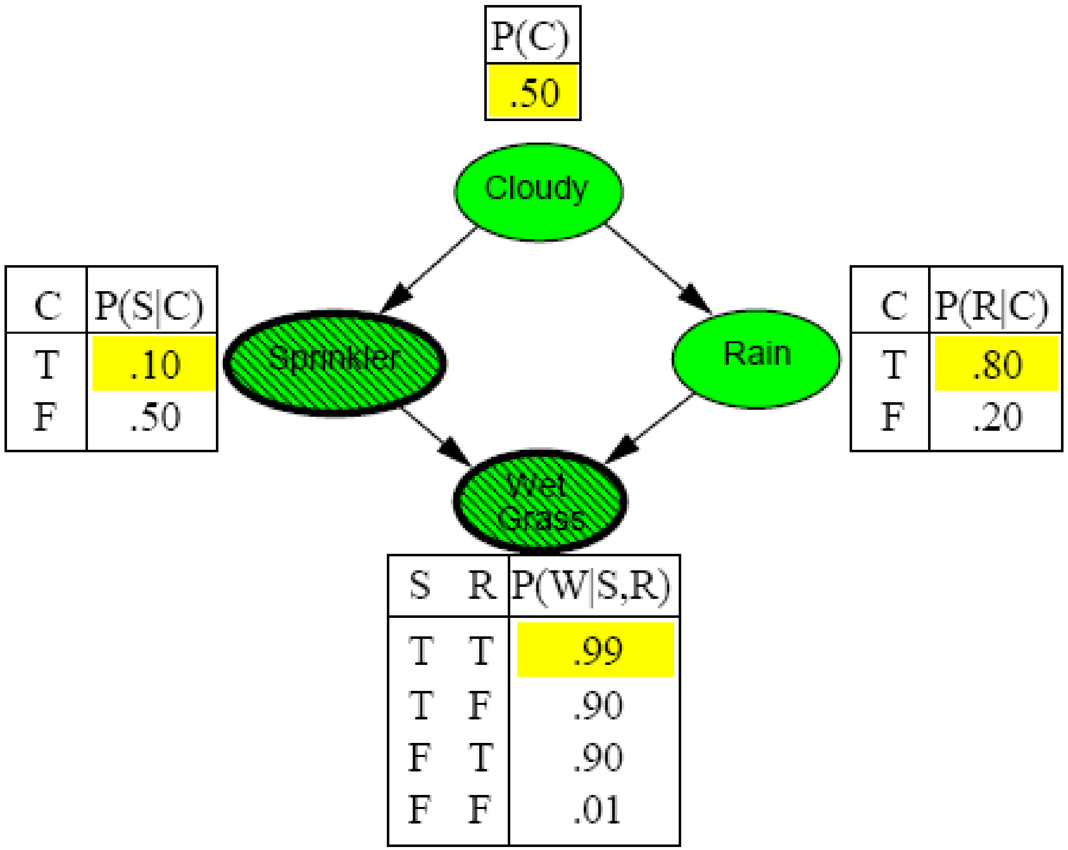
P(C)
.50

C	P(S|C)
T	.10
F	.50

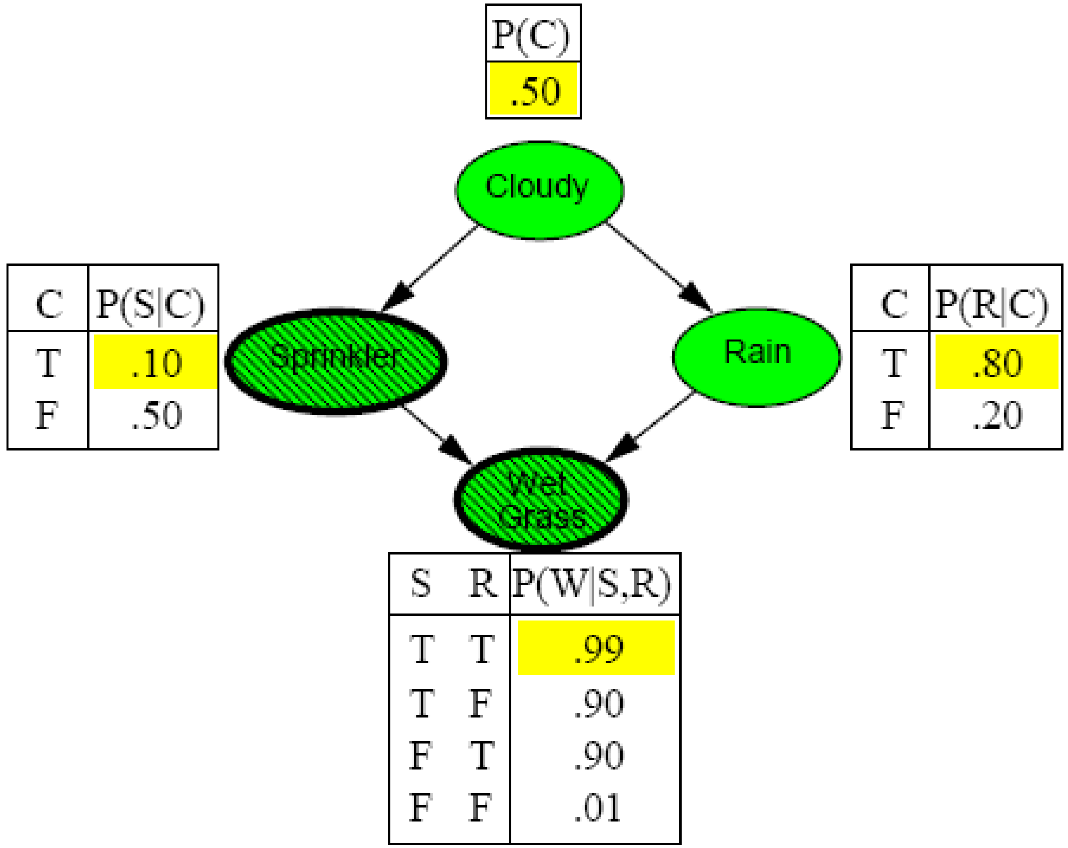
C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$$w = 1.0 \times 0.1$$


P(C)
.50
Cloudy
C P(S|C)
T .10
F .50
Sprinkler
Rain
C P(R|C)
T .80
F .20
Wet Grass
S R P(W|S,R)
T T .99
T F .90
F T .90
F F .01

$w = 1.0 \times 0.1$



$w = 1.0 \times 0.1 \times 0.99 = 0.099$

$$
\begin{aligned}
\hat{P}(x|\mathbf{e}) &= \alpha \sum_{\mathbf{Y}} N_{WS}(x,\mathbf{y},\mathbf{e})w(x,\mathbf{y},e) \\
&\approx \alpha' \sum_{\mathbf{Y}} S_{WS}(x,\mathbf{y},\mathbf{e})w(x,\mathbf{y},\mathbf{e}) \\
&= \alpha' \sum_{\mathbf{y}} P(x,\mathbf{y},\mathbf{e}) \\
&= \alpha' P(x,\mathbf{e}) = P(x|\mathbf{e})
\end{aligned}
$$

Referencias Bibliográficas

- Capitulo 14 Artificial Intelligence: A Modern Approach, Russell and Norvig
- http://www.cs.iastate.edu/~cs572/studyguide.html
- http://www.aaai.org/AITopics/pmwiki/pmwiki.php/AITopics/Uncertainty