CC-226 Aula 05 - Inferência Bayesiana

Carlos Henrique Q. Forster - Instituto Tecnológico de Aeronáutica

2008

1 Probabilidade Condicional

1.1 Exemplo

(Retirado do Livro do Parzen)

4 bolas brancas

2 bolas vermelhas

Sorteio uma e depois outra sem reposição

	WR	WR	WW	RW	WW	WW
	WW	RW	WW	WW	WR	RR
Ocorrências:	RW	WW	RW	RR	WR	RW
	WW	RW	WW	WR	WR	RW
	WW	RW	WR	RW	RW	WW

$$N = \text{número de jogadas} = 30$$

$$N_A = \text{número de vezes que ocorreu W no primeiro sorteio} = 18$$

$$N_B = \text{número de vezes que ocorreu W no segundo sorteio} = 21$$

$$N_{AB} = \text{número de vezes que ocorreu WW} = 11$$

$$P(A) = \frac{N_A}{N}$$

$$P(B) = \frac{N_B}{N}$$

$$P(AB) = \frac{N_{AB}}{N}$$

Probabilidade condicional do evento B, dado o evento A:

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{N_{AB}}{N_A}$$

... definido apenas para $N_A > 0$.

Probabilidade incondicional ou marginal, obtida da probabilidade condicional:

$$P(B) = P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})$$

Probabilidade conjunta:

$$P(AB) = P(B|A) \cdot P(A)$$

1.2 Exemplo

Considerar uma família com duas crianças. Assumir que haja a mesma chance de cada criança ser menino ou menina.

Qual a probabilidade de ambas serem meninos?

a) dado que a mais velha é menino.

b) dado que pelo menos um deles é menino.

Solução:

a)

$$P(AB|A) = \frac{P(AB)}{P(A)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

b)

$$P(AB|A \cup B) = \frac{P(AB)}{P(A \cup B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Probabilidade conjunta e probabilidade condicional:

$$P(AB) = P(B|A)P(A) = P(A|B)P(B)$$

Reescrevendo, obtemos a regra de Bayes:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Lembrando independência estatística, se A, B e C são independentes, então:

$$P(AB) = P(A)P(B)$$

$$P(ABC) = P(A)P(B)P(C)$$

Supondo A_j independente de todos anteriores, temos:

$$P(A_j|A_1, A_2, \ldots, A_{j-1}) = P(A_j)$$

A probabilidade condicional para uma seqüência de eventos é dada por:

$$P(A_n|A_1, A_2, \ldots, A_{n-1}) = \frac{P(A_1, \ldots, A_n)}{P(A_1, \ldots, A_{n-1})}$$

1.3 Exemplo

Considere duas urnas:

Urna I contém 5 bolas brancas e 3 bolas pretas.
Urna II contém 3 bolas brancas e 7 bolas pretas.
Uma urna é escolhida aleatoriamente e uma bola é tirada dela.
Qual a probabilidade da bola ser branca?
Solução:
Supondo $P(C_1) = P(C_2) = \frac{1}{2}$ as probabilidades de escolher uma das urnas.
$\frac{1}{2}$ de pegar urna I

- $5/8$ de pegar B
- $3/8$ de pegar P

$\frac{1}{2}$ de pegar urna II

- $3/10$ de pegar B
- $7/10$ de pegar P

$$P(B) = P(B|C_1)P(C_1) + P(B|C_2)P(C_2) = \frac{5}{8}\frac{1}{2} + \frac{3}{10}\frac{1}{2} = \frac{5}{16} + \frac{3}{20} = \frac{35}{80}$$

Interpretação

Sejam A e B eventos definidos no mesmo espaço de probabilidades S.

Assumir $P(A) > 0$
B é independente de A se $P(B|A) = P(B)$
$P(B|A)$ é a probabilidade de ocorrer B uma vez que se assuma que A ocorreu.
$P(B|A)$ é uma reavaliação da probabilidade de B dado o fato de que A ocorreu.
Se $A = A_1A_2A_3 \dots A_n$ (intersecção),

$$P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\dots P(A_n|A_1A_2\dots A_{n-1})$$

Exemplo

Urna com M bolas, das quais N são brancas.

Tiramos n bolas: qual a probabilidade de todas serem brancas?

$$P(A_1) = \frac{N}{M}$$

$$P(A_2|A_1) = \frac{N-1}{M-1}$$

$$P(A_3|A_1A_2) = \frac{N-2}{M-2}$$

$$\ldots$$

$$P(A_n|A_1A_2\ldots A_{n-1}) = \frac{N-(n-1)}{M-(n-1)}$$

$$P(A_1\ldots A_n) = \frac{N(N-1)(N-2)\ldots(N-(n-1))}{M(M-1)(M-2)\ldots(M-(n-1))}$$

1.4 Regra de Bayes

$C_1 \ldots C_n$ são eventos mutuamente exclusivos e exaustivos com probabilidades $P(C_i)$ conhecidas.

B é um evento para o qual se conhece $P(B|C_i)$.

Interpretar C_i como possíveis causas para o evento B.

Computar $P(C_i|B)$

$$P(C_i|B) = \frac{P(BC_i)}{P(B)} = \frac{P(B|C_i)P(C_i)}{\sum_{j=1}^{n} P(B|C_j)P(C_j)}$$

Exemplo

O evento C significa que o paciente tem câncer.

O evento A significa que o teste deu positivo.

Vamos supor que

$$P(A|C) = 0,95$$

$$P(\bar{A}|\bar{C}) = 0,95$$

(é um bom teste?)

Computar $P(C|A)$, probabilidade de ter câncer dado que o teste deu positivo.

$$P(C|A) = \frac{P(AC)}{P(A)} = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|\bar{C})P(\bar{C})}$$

Assumindo que a probabilidade de ter câncer na população é de $P(C) = 0,005$.

$$P(C|A) = \frac{0,95 \cdot 0,005}{0,95 \cdot 0,005 + 0,005 \cdot 0,995} = 0,087$$

(é um bom teste?)

Vamos descobrir qual deve ser a probabilidade $P(A|C) = P(\bar{A}|\bar{C}) = R$ para que a taxa de acerto do teste seja $P(C|A) = 0,95$.

$$P(C|A) = \frac{R \cdot 0,005}{R \cdot 0,005 + (1-R) \cdot 0,995} = 0,95$$

Resolvendo para R, obtemos:

$$R = 0,999736$$

Exemplo

Urna com bolas: 2 brancas e 3 vermelhas.

X_1 e X_2 são variáveis aleatórias binárias correspondentes ao primeiro e ao segundo sorteio resultarem em bola branca.

Primeiro caso: Com reposição.

Neste caso, as variáveis X_1 e X_2 são independentes.

$3/5$ de chance de sortear não-branca na primeira

- $3/5$ de chance de sortear não-branca na segunda
- $2/5$ de chance de sortear branca na segunda

$2/5$ de chance de sortear branca na primeira

- $3/5$ de chance de sortear não-branca na segunda
- $2/5$ de chance de sortear branca na segunda

	$p_{X_1X_2}(x_1,x_2)$		
	0	1	$p_{X_2}(x_2)$
0	$\frac{3}{5}\frac{3}{5}$	$\frac{3}{5}\frac{2}{5}$	$\frac{3}{5}$
1	$\frac{2}{5}\frac{3}{5}$	$\frac{2}{5}\frac{2}{5}$	$\frac{2}{5}$
$p_{X_1}(x_1)$	$\frac{3}{5}$	$\frac{2}{5}$	

Segundo caso: Sem reposição

Neste caso, há dependência das variáveis X_1 e X_2.

$3/5$ de chance de sortear não-branca na primeira

- $2/4$ de chance de sortear não-branca na segunda
- $2/4$ de chance de sortear branca na segunda

$2/5$ de chance de sortear branca na primeira

- $3/4$ de chance de sortear não-branca na segunda
- $1/4$ de chance de sortear branca na segunda

	$p_{X_1X_2}(x_1,x_2)$		
	0	1	$p_{X_2}(x_2)$
0	$\frac{3}{5}\frac{2}{4}$	$\frac{2}{5}\frac{3}{4}$	$\frac{3}{5}$
1	$\frac{3}{5}\frac{2}{4}$	$\frac{2}{5}\frac{1}{4}$	$\frac{2}{5}$
$p_{X_1}(x_1)$	$\frac{3}{5}$	$\frac{2}{5}$	

Notar que a probabilidade conjunta não é mais o produto das probabilidades marginais.

Mais uma interpretação

Seja H o evento que represente dor-de-cabeça.

Seja F o evento que representa gripe.

Sabemos que $P(H) = 1/10$ (tem uma dor de cabeça a cada dez semanas)

Sabemos que $P(F) = 1/40$ (tem uma gripe a cada 40 semanas)

Porém $P(H|F) = 1/2$.

Dor de cabeça é raro e também é gripe, mas, se estiver com gripe, a chance de ter dor de cabeça passa a 50%.

$$P(H|F) = \frac{P(H \cap F)}{P(F)} = \frac{\text{situações com gripe e dor de cabeça}}{\text{situações com gripe}}$$

Corolário

$$P(HF) = P(H|F)P(F)$$

Regra de Bayes

$$P(F|H) = \frac{P(H|F)P(F)}{P(H)} = \frac{0,5 \cdot 1/40}{1/10} = \frac{1}{8}$$

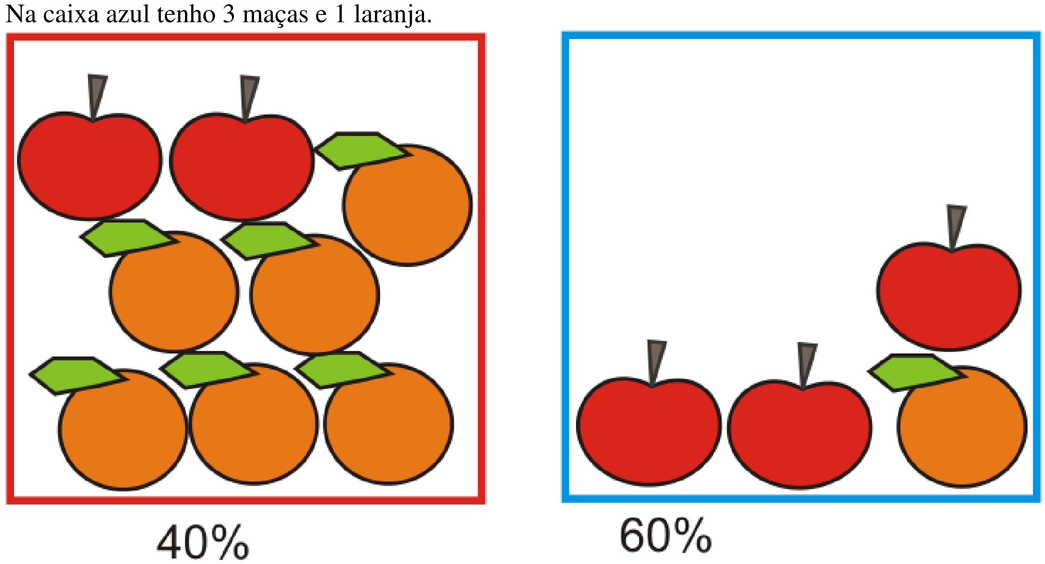
2 Inferência

Exemplo do Christopher Bishop.

Considere que tenho duas caixas de frutas, uma vermelha (vm) e uma azul (az).

Na caixa vermelha tenho 2 maças e 6 laranjas.

Na caixa azul tenho 3 maças e 1 laranja.



Cada Experimento consiste em pegar uma fruta de uma caixa e colocar de volta.

Supor que 40% das vezes peguei da caixa vermelha e que 60% das vezes peguei da caixa azul.

B é a variável aleatória que representa a caixa de onde peguei.

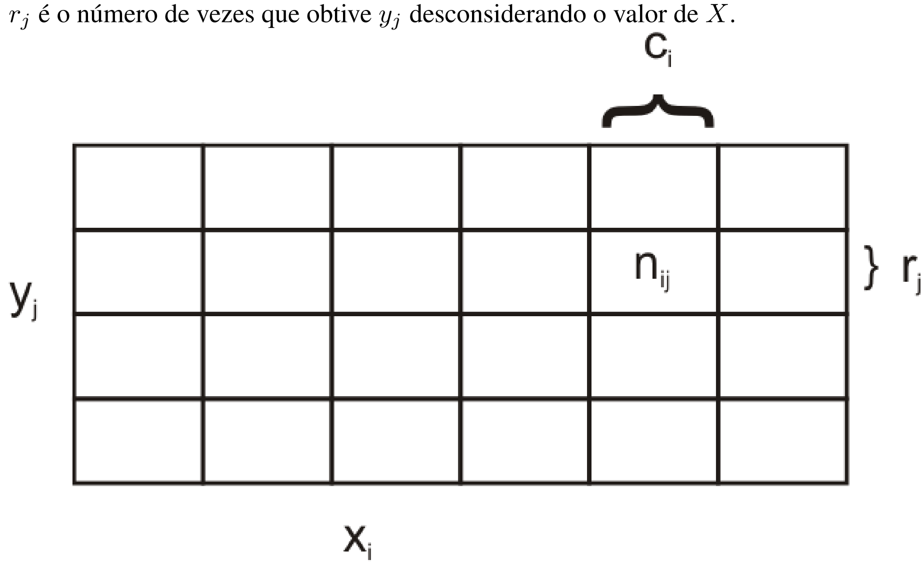
$$p(B = vm) = 4/10$$

$$p(B = az) = 6/10$$

A variável F representa a fruta que pode ser mc ou lj.
Qual a probabilidade de se pegar uma maça?
Dado que peguei uma laranja, qual a probabilidade de tê-la pego da caixa azul?

Generalizando

X é uma variável aleatória com valores x_i para $i = 1 \dots M$
Y é uma variável aleatória com valores y_j com $j = 1 \dots L$
Após N jogadas, amostrando X e Y, temos:
n_{ij} é o número de vezes que obtive x_i e y_j simultaneamente.
c_i é o número de vezes que obtive x_i desconsiderando o valor de Y.
r_j é o número de vezes que obtive y_j desconsiderando o valor de X.



$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

é a probabilidade conjunta de x_i e y_j.

A probabilidade de que X tire o valor x_i, desconsiderando o valor obtido para Y é escrita como $p(X = x_i)$ e é dada pela fração do número total de pontos que cai na coluna i, então

$$p(X = x_i) = \frac{c_i}{N}$$

A regra da soma ou marginalização serve para computar essa probabilidade chamada marginal.

$$p(X = x_i) = \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

No caso da nossa tabela:

$$c_i = \sum_j n_{ij}$$

Se considerarmos apenas as instâncias para a qual $X = x_i$, ou seja, as c_i instâncias da coluna i da tabela, escrevemos a probabilidade de $Y = y_j$ como

$$p(Y = y_j | X = x_i) = \frac{P(Y = y_j \wedge X = x_i)}{P(X = x_i)} = fracn_{ij}c_i$$

que é a probabilidade condicional de $Y = y_j$ dado $X = x_i$.

A probabilidade conjunta é dada por:

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N}$$

$$= p(Y = y_j | X = x_i) \cdot p(X = x_i)$$

Esta é a regra do produto, enunciada como

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

Utilizando a simetria

$$p(X, Y) = p(Y, X)$$

Expandindo,

$$p(X|Y)p(Y) = p(Y|X)p(X)$$

Obtemos a regra de Bayes

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

onde, em geral se utiliza

$$p(X) = \sum_{Y'} p(X|Y')p(Y')$$

Voltando à caixa de frutas

$$p(B = vm) = 4/10$$

$$p(B = az) = 6/10$$

$$p(B = vm) + p(B = az) = 1$$

Pegando da caixa azul $B = az$, a probabilidade de ser maça é $p(F = mc|B = az) = 3/4$

Listando todas probabilidades condicionais:

$$\begin{aligned} p(F = mc|B = vm) &= 1/4 \quad (1) \\ p(F = lj|B = vm) &= 3/4 \quad (2) \\ p(F = mc|B = az) &= 3/4 \quad (3) \\ p(F = lj|B = az) &= 1/4 \quad (4) \end{aligned}$$

Probabilidade de obter uma maça é dada por

$$p(F = mc) = p(F = mc|B = vm)p(B = vm) + p(F = mc|B = az)P(B = az)$$

$$p(F = mc) = 1/4 \cdot 4/10 + 3/4 \cdot 6/10 = \frac{11}{20}$$

Obviamente

$$p(F = lj) = \frac{9}{20}$$

Uma fruta foi escolhida e é uma laranja, qual a probabilidade da caixa?

Aplicamos a regra de Bayes:

$$p(B = vm|F = lj) = \frac{p(F = lj|B = vm)p(B = vm)}{p(F = lj|B = vm)p(B = vm) + p(F = lj|B = az)p(B = az)}$$

$$= \frac{\frac{3}{4} \cdot \frac{4}{10}}{\frac{9}{20}} = \frac{2}{3}$$

Obviamente,

$$p(B = az|F = lj) = \frac{1}{3}$$

Interpretação

A regra de Bayes pode ser vista como um incremento ou atualização da probabilidade dada uma nova informação.

Antes de utilizar a informação, temos uma distribuição de probabilidades a priori $p(X)$.

Após utilizar a informação Y, atualizamos a distribuição de probabilidades, obtendo uma distribuição a posteriori $p(X|Y)$.

O termo $p(Y|X)$ é a chamada verossimilhança e não é necessariamente uma distribuição de probabilidades. O denominador $p(Y)$ fornece uma normalização. Em geral se utiliza um somatório sobre todo possível valor para X para obter a distribuição marginal $p(Y)$. Assim:

$$p(Y) = \sum_x P(Y|X = x)P(X = x)$$

No caso contínuo, a marginalização será uma integral:

$$p(y) = \int p(x,y)dx = \int p(y|x)p(x)dx$$

Podemos definir a esperança condicional de uma função f como

$$\mathcal{E}_x[f|y] = \sum_x p(x|y)f(x)$$

3 Redes bayesianas

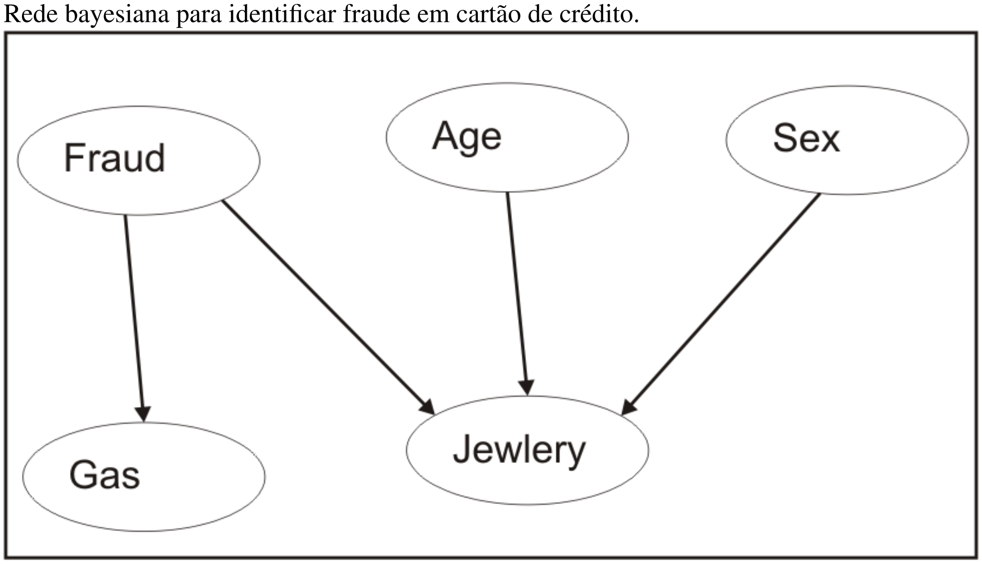
Utilizando a regra da cadeia da probabilidade:

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i|x_1 \ldots x_{i-1})$$

As setas indicam dependência estatística (e também causalidade).

Exemplo do tutorial do David Heckerman (A tutorial on learning with bayesian networks)

Rede bayesiana para identificar fraude em cartão de crédito.



Fraud indica o evento de ser um caso de fraude. Gas indica uma compra de combustível dentro de 24 horas. Jewelry indica uma compra de jóias dentro de 24 horas. Age indica a idade do cliente. Sex indica o sexo do cliente.

$$\begin{aligned} p(f = yes) &= 0,00001 \\ p(a \leq 30) &= 0,25 \\ p(a = 30..50) &= 0,40 \end{aligned}$$

$$
\begin{aligned}
p(s = male) &= 0,5 \\
p(g = yes|f = yes) &= 0,2 \\
p(g = yes|f = no) &= 0,01 \\
p(j = yes|f = yes, a = *, s = *) &= 0,05 \\
p(j = yes|f = no, a \leq 30, s = male) &= 0,0001 \\
p(j = yes|f = no, a = 30..50, s = male) &= 0,0004 \\
p(j = yes|f = no, a \geq 50, s = male) &= 0,0002 \\
p(j = yes|f = no, a \leq 30, s = female) &= 0,0005 \\
p(j = yes|f = no, a = 30..50, s = female) &= 0,002 \\
p(j = yes|f = no, a \geq 50, s = female) &= 0,001
\end{aligned}
$$

Percorrer na ordem topológica.
Exemplo: FASGJ

$$
\begin{aligned}
p(a|f) &= p(a) \\
p(s|f, a) &= p(s) \\
p(g|f, a, s) &= p(g|f) \\
p(j|f, a, s, g) &= p(j|f, a, s)
\end{aligned}
$$

Computar a probabilidade de fraude, dada informação de idade, sexo, compra de combustível e de jóia.

$$
p(f|a, s, g, j) = \frac{p(f, a, s, g, j)}{p(a, s, g, j)} = \frac{p(f, a, s, g, j)}{p(f, a, s, g, j) + p(\bar{f}, a, s, g, j)}
$$

Usando as independências conhecidas:

$$
p(f|a, s, g, j) = \frac{p(f)p(g|f)p(j|f, a, s)}{\sum_{f'} p(f')p(g|f')p(j|f', a, s)}
$$

Observação

Independência condicional.

$$
p(x_3|x_2, x_1) = p(x_3|x_2)
$$

quer dizer que x_3 é independente de x_1 quando x_2 é conhecido, o que não garante que

$$
p(x_3|x_1) = p(x_3)
$$

x_3 e x_1 só são independentes quando conheço x_2.

Explain-away

Exemplo do Alpaydin.

Diagnóstico: o sintoma é a grama estar molhada, vamos encontrar a probabilidade de ter chovido.

Antes de sabermos que a grama estava molhada, a probabilidade de ter chovido era:

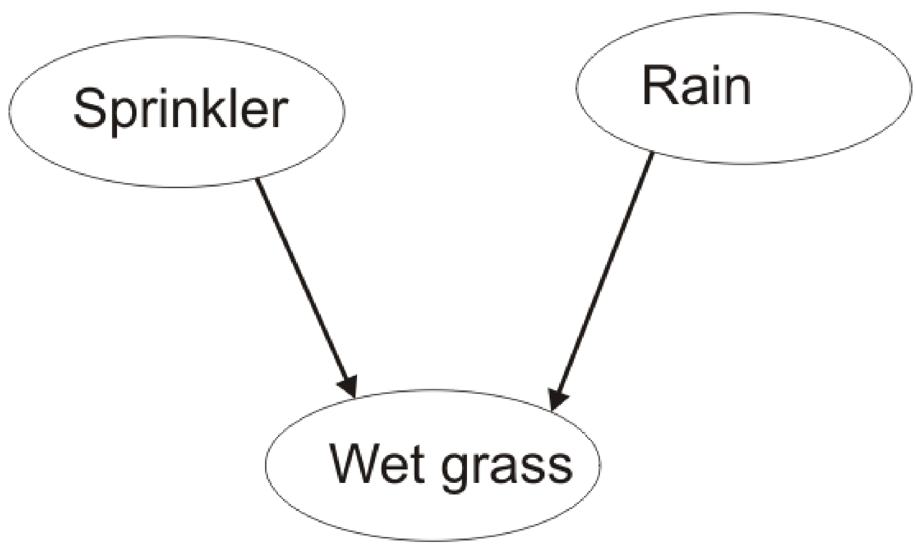
$$P(R) = 0,4$$

Com a nova informação, obtemos

$$P(R|W) = \frac{P(W|R)P(R)}{P(W)} = \frac{P(W|R)P(R)}{P(W|R)P(R) + P(W|\bar{R})P(\bar{R})}$$

Supondo que $P(W|R) = 0,9$ e que $P(W|\bar{R}) = 0,2$:

$$P(R|W) = \frac{0.9 \times 0.4}{0.9 \times 0.4 + 0.2 \times 0.6} = 0.75$$



Duas causas para encontrarmos a grama molhada (wet grass): pode ter sido a chuva (rain) ou o irrigador (sprinkler).

$$\begin{aligned}
P(S) &= 0,2 && (5)\\
P(R) &= 0,4 && (6)\\
P(W|R,S) &= 0,95 && (7)\\
P(W|R,\bar{S}) &= 0,90 && (8)\\
P(W|\bar{R},S) &= 0,90 && (9)\\
P(W|\bar{R},\bar{S}) &= 0,10 && (10)
\end{aligned}$$

Qual a probabilidade da grama estar molhada se ligarmos o irrigador, chova ou faça sol?

Calculamos a probabilidade marginal, marginalizando em função de R

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

Obtemos as relações de independência da rede bayesiana, por exemplo, sabemos que

$$P(R|S) = P(R)$$

porque não há setas entre R e S.

Assim,

$$P(W|S) = P(W|R, S)P(R) + P(W|\bar{R}, S)P(\bar{R}) = 0.95 \times 0.4 + 0.9 \times 0.6 = 0.92$$

Vamos estimar agora a chance do irrigador estar ligado se a grama estava molhada

$$P(S|W) = \frac{P(W|S)P(S)}{P(W)} = \frac{0.92 \times 0.2}{0.52} = 0.35$$

onde

$$P(W) = P(W|R, S)P(R, S) + P(W|\bar{R}, S)P(\bar{R}, S) + P(W|R, \bar{S})P(R, \bar{S}) + P(W|\bar{R}, \bar{S})P(\bar{R}, \bar{S}) = 0.52$$

Agora vamos supor que sabemos que choveu. Assim, temos

$$P(S|R, W) = \frac{P(W|R, S)P(S|R)}{P(W|R)} = \frac{P(W|R, S)P(S)}{P(W|R)} = 0,21$$

Observe que por sabermos da chuva, a probabilidade do irrigador estar ligado diminuiu. Isso é chamado explaining-away. Uma vez que sabemos da chuva, R e S ficam dependentes.

4 Classificador Naïve Bayes

Supomos os atributos x_j estatisticamente independentes.

$$p(\mathbf{x}|C_k) = \prod_{j=1}^{M} p(x_j|C_k), k = 1, \ldots, L$$

Associar $\mathbf{x}$ à classe C_m que:

$$C_m = arg \max_{C_k} \prod_{j=1}^{M} p(x_j|C_k), k = 1, \ldots, L$$

O classificador Naive Bayes é um caso particular de rede bayesiana.

5 Regras de associação

Uma regra de associação é uma implicação da forma $X \rightarrow Y$. Um exemplo típico é a análise de carrinho de compras. Quem compra X também compra Y.

Duas medidas são calculadas para essa análise:

Confiança da regra de associação deve ser próxima de 1 e maior que $p(Y)$ e fornece uma medida da importância ou intensidade da regra

$$conf(x \rightarrow y) = p(y|x) = \frac{p(x,y)}{p(x)} = \frac{\text{quantos compraram X e Y}}{\text{quantos compraram X}}$$

O suporte da regra de associação informa a significância estatística da regra:

$$support(x \rightarrow y) = p(x,y) = \frac{\text{quantos compraram X e Y}}{\text{compradores no total}}$$

O algoritmo Apriori busca essas regras numa base de dados de forma eficiente.