Raciocínio Probabilístico

Tópico 10 – EA072

Inteligência Artificial em Aplicações Industriais

Prof. Fernando J. Von Zuben
(DCA/FEEC/Unicamp)
Pablo A. D. de Castro
(DCA/FEEC/Unicamp)

Nov/2008

Roteiro

- Parte I
  – Introdução
- Parte II
  – Fundamentos da teoria da probabilidade
- Parte III
  – Redes Bayesianas
- Parte IV
  – Aprendizado em redes Bayesianas

PARTE I - Introdução

Sistemas Inteligentes

- A construção de sistemas inteligentes artificiais está fundamentada na disponibilidade de recursos computacionais para sua implementação.

- Neste contexto, uma das tarefas mais difíceis é simular em computadores digitais a forma como os seres humanos raciocinam.

- Diante de informações parciais ou imprecisas, apenas soluções aproximadas podem ser obtidas, isto é, soluções com incerteza.

- Deste modo, é preciso dispor de técnicas para lidar com a incerteza.

Incerteza

"Incerteza se origina de alguma deficiência de informação. A informação pode estar incompleta, ser vaga, imprecisa ou contraditória."

Klir,G.J. and Folger,T.A. "*Fuzzy Sets, Uncertainty and Information*", Prentice Hall, 1998.

Incerteza

- Algumas técnicas que podem ser utilizadas para tratar incertezas
  - Raciocínio lógico: acredite em alguma coisa até que se encontre uma evidência que prove o contrário (ignora as incertezas)
  - Lógica nebulosa: informações possuem grau de verdade
  - **Raciocínio Probabilístico**

Raciocínio Probabilístico

- Teoria da probabilidade oferece uma maneira quantitativa de codificar incertezas.
- Possui uma semântica clara
- Probabilidades podem ser obtidas a partir de dados.
- Permite incorporar novas evidências de forma direta.
- Frequencistas $\times$ Bayesianos

Objetivos deste tópico

Após estudarmos este tópico, seremos capazes de responder a duas perguntas:

1) Como lidar com incerteza usando probabilidades?
2) Como representar eficientemente a base de conhecimento de um sistema inteligente que utiliza raciocínio probabilístico?

PARTE II – Fundamentos da Teoria de Probabilidade

Experimento Aleatório

- Um experimento aleatório, **E**, é um experimento em que o resultado não pode ser predito, mesmo que seja repetido várias vezes e nas mesmas condições.

- Experimentos:

E1: “Arremesso de moeda”.

E2: “Arremesso de dado”.

E3: “Retirar uma bola de uma urna contendo bolas numeradas de 1 a 100”.

- Variáveis aleatórias:

E1: “Face da moeda para cima”.

E2: “Face do dado para cima”.

E3: “Número da bola”.

Espaço amostral

- É o conjunto de todos os valores que a variável aleatória pode assumir.

- Exemplos de espaços amostrais:

E1: {cara, coroa}

E2: {1, 2, 3, 4, 5, 6}

E3: {1, 2, ..., 100}

Eventos

- Um **evento** é um subconjunto do espaço amostral.

Exemplos:

E1: “Jogar uma moeda três vezes e observar quantas vezes aparece cara e quantas vezes aparece coroa”.

Evento: “Aparecer 2 caras (k) e uma coroa (c) em qualquer ordem”.

Subconjunto: {(kkc), (kck), (ckk)}

Eventos

E2: “Jogar um dado e observar o número na face de cima”.

Evento: “Número é par”.

Subconjunto: {2,4,6}

E3: “Uma urna contém bolas numeradas de 1 a 100. Retirar uma bola e observar seu número”.

Evento: “bola com número maior que 80”

Subconjunto: {81, 82, ..., 98, 99, 100}

Variável Aleatória

- É aquela que assume valores num espaço amostral e para a qual está determinada a probabilidade de ocorrência de cada um dos elementos do espaço amostral.
- Assim, se o espaço amostral for contínuo, é necessário conhecer a função distribuição de probabilidade para caracterizar a variável aleatória.
- Se o espaço amostral for discreto, é necessário conhecer a função massa de probabilidade para caracterizar a variável aleatória.

Probabilidade

- Para cada evento é atribuído um número real no intervalo [0,1], dizendo qual a probabilidade de ocorrer tal evento.
- Se o **espaço amostral** consiste de **N** elementos igualmente prováveis e o evento A corresponde a um subconjunto de **r** elementos do espaço amostral, então a probabilidade de ocorrer A é dada por:

$$\mathbf{P(A) = r / N}$$

Probabilidade

Exemplo: Uma urna contém 10 bolas numeradas de 0 a 9. Num experimento, é preciso selecionar uma bola da urna e anotar seu número. Deseja-se encontrar a probabilidade dos eventos:

A = “número da bola é 5”

B = “número da bola é ímpar”

C = “número da bola é múltiplo de 3”

O espaço amostral é S={0,1,...,9} e os resultados correspondentes aos eventos acima são:

A={5}　　　B={1,3,5,7,9}　　　C={3,6,9}

Se for suposto que os resultados são equiprováveis, então:

P(A)= P(5) = 1/10

P(B)=P(1)+P(3)+P(5)+P(7)+P(9)
= 1/10 + 1/10 + 1/10 + 1/10 + 1/10 = 5/10.

P(C)=P(3)+P(6)+P(9)
= 1/10 + 1/10 + 1/10 = 3/10.

Probabilidade Conjunta

Fornece a probabilidade de dois ou mais eventos ocorrerem **simultaneamente**.

Ao se jogar uma moeda duas vezes, temos os eventos:

A={Obter cara (k) na primeira jogada} -> P(A) = 1/2

B={Obter cara (k) na segunda jogada também} -> P(B)=1/2

Logo,

$$P(A \wedge B) = P(A)P(B) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

Probabilidade Conjunta

Atenção!!!

$$P(A \wedge B)=P(A)P(B)$$

Só é válida se os eventos forem **independentes,** como no exemplo anterior.

Dois eventos são independentes se a ocorrência de um não altera a probabilidade de ocorrência do outro. Caso contrário, eles são chamados de eventos **não independentes** ou eventos dependentes.

Probabilidade Conjunta

A probabilidade conjunta de dois eventos **não independentes** pode ser calculada pela intersecção dos resultados de cada evento. Exemplo:

Possíveis eventos num experimento de jogar um dado uma vez:

A: “Número da face igual a 6”   Resultado: {6}

B: “Número da face maior que 4”   Resultado: {5,6}

P(A $\wedge$ B)= Intersecção de {6} e {5,6}, o que resulta em {6}

Logo, P(AB)= 1/6

Número da face maior que 4 e número da face igual a 6. Só pode haver um resultado que atenda este evento conjunto.

Probabilidade Condicional

A probabilidade condicional responde questões do tipo:

“Dado que ocorreu o evento A, qual a probabilidade de ocorrer o evento B?”

Por exemplo:

Qual a probabilidade de um paciente estar com cárie, dado que ele está com dor de dente?

Qual a probabilidade de chover, dado que o céu está completamente nublado?

Probabilidade Condicional

Um experimento consiste em jogar um dado uma vez. Sejam os eventos:

A: “Número na face é 6”

B: “Número na face é maior que 4”

O dado foi jogado. Qual a probabilidade de ocorrer o evento A (sair 6), sabendo que o evento B ocorreu?

$$
\begin{aligned}
P(A|B) &= P(A \wedge B) \;/\; P(B) \\
&= 1/6 \quad / \quad 1/3 \\
&= 1/2
\end{aligned}
$$

Probabilidade Condicional

Quantidade de pessoas:

	Não Fuma	Fuma	Total
Sem Câncer	40	10	50
Câncer	7	3	10
Total	47	13	60

Qual a probabilidade de uma pessoa ter câncer, dado que ela fuma?

P(C="S"|F="S") = P(C="S" $\wedge$ F="S") / P(F="S")

= 3/60 / 13/60

= 0,23 ou 23%

Marginalização

P(Raining, Windy)

Raining	Windy = False	Windy = True	Sum
False	0.64	0.16	0.8
True	0.1	0.1	0.2

P(Raining)

Raining = False	Raining = True
0.8	0.2

Fonte: https://www.bayesserver.com/docs/introduction/bayesian-networks

Marginalização

P(A,B,C,D)

B	C	D	A = True	A = False
True	True	True	0.0036	0.0054
True	True	False	0.0098	0.0252
True	False	True	0.0024	0.0486
True	False	False	0.0042	0.1008
False	True	True	0.0256	0.0864
False	True	False	0.0432	0.1728
False	False	True	0.0064	0.2016
False	False	False	0.0048	0.2592

P(A,C)

C	A = True	A = False
True	0.0822	0.2898
False	0.0178	0.6102

Fonte: https://www.bayesserver.com/docs/introduction/bayesian-networks

Lei da Probabilidade Total

A Lei da Probabilidade Total diz que, se um evento A ocorre em *m* condições diferentes, todas mutuamente exclusivas, então a probabilidade de ocorrer o evento A é a soma das probabilidades dele ocorrer nas *m* condições diferentes.

Denominando de c_i a *i*-ésima condição em que ocorre o evento A, formalmente tem-se:

$$P(A) = P(A \mid c_1)\, P(c_1) + P(A \mid c_2)\, P(c_2) + \ldots + P(A \mid c_m)\, P(c_m)$$

Lei da Probabilidade Total

Exemplo: Controle de Qualidade

Uma companhia multinacional tem três fábricas que produzem o mesmo tipo de produto. A fábrica I é responsável por 30% do total produzido, a fábrica II produz 45% do total, e o restante vem da fábrica III. Cada uma das fábricas, no entanto, produz uma propprção de produtos que não atendem aos padrões estabelecidos pelas normas internacionais. Tais produtos são considerados "defeituosos" e correspondem a 1%, 2% e 1,5%, respectivamente, dos totais produzidos por fábrica.

No centro de distribuição, é feito o controle de qualidade da produção combinada das fábricas.

Lei da Probabilidade Total

Exemplo: Controle de Qualidade

Qual é a probabilidade de encontrar um produto defeituoso durante a inspeção de qualidade?

Resp.: Seja o evento $A = \{\text{Produto Defeituoso}\}$ e $F_i = \{\text{Produto da Fábrica } i\}$. Sabemos, pelo enunciado, que $P(F_1) = 0{,}3$, $P(F_2) = 0{,}45$ e $P(F_3) = 0{,}25$. Além disso, sabemos que $P(A|F_1) = 0{,}01$, $P(A|F_2) = 0{,}02$ e $P(A|F_3) = 0{,}015$.

Então, pela lei da probabilidade total,

$$P(A) = P(A|F_1)P(F_1) + P(A|F_2)P(F_2) + P(A|F_3)P(F_3) =$$

$$= 0{,}3 * 0{,}01 + 0{,}45 * 0{,}02 + 0{,}25 * 0{,}015 = 0{,}01575$$

Teorema de Bayes

Thomas Bayes (1702-1761) formulou um teorema capaz de lidar com incertezas e atualizar nossa crença em um determinado evento à medida que novas informações chegam.

Esse teorema simples é a base de todos os sistemas inteligentes modernos que utilizam a inferência probabilística.

Teorema de Bayes

Equação do Teorema de Bayes:

$$\mathbf{P(A|B) = P(B|A)P(A) / P(B)}$$

Esta equação simples é uma ferramenta útil para inferir a probabilidade a *posteriori* de um evento baseado na evidência e num conhecimento a *priori* de outro(s) evento(s).

Em outras palavras, qual é a probabilidade de um evento anterior ter ocorrido sabendo-se que um evento posterior ocorreu?

Teorema de Bayes

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

A: representa a hipótese

B: representa a evidência

P(A): probabilidade a priori de ocorrer **A**

P(B|A): probabilidade de ocorrer a evidência **B** dado que a hipótese **A** é verdadeira

P(B): probabilidade marginal de ocorrer **B**. É também interpretada como a probabilidade de ocorrência de **B** sob todas as hipóteses mutuamente exclusivas: $\mathbf{P(B)=\sum_i P(B|A_i)P(A_i)}$

P(A|B): probabilidade a posteriori de **A** dado que ocorreu **B**.

Teorema de Bayes

- A inferência bayesiana indica a probabilidade de cada hipótese, a partir dos dados já observados, e é o resultado do uso de todas as hipóteses, devidamente ponderadas.

- Repare que não se adota aqui a escolha da hipótese mais provável para se executar a tarefa de predição. Logo, a predição se transforma em um problema de inferência probabilística.

- Seja $\mathbf{d} = [d_1 \quad d_2 \quad \cdots \quad d_N]^T$ o vetor de dados já observados. Pela regra de Bayes, a probabilidade de cada hipótese é dada por:

$$P(h_i|\mathbf{d}) = P(\mathbf{d}|h_i)\, P(h_i) \,/\, P(\mathbf{d})$$

Teorema de Bayes

Duas quantidades-chaves na abordagem bayesiana são:

- probabilidade a priori de cada hipótese: $P(h_i)$
- probabilidade dos dados, condicionada à hipótese: $P(\mathbf{d}|h_i)$.

$P(h_i)$ indica o grau inicial de verossimilhança da hipótese.

$P(\mathbf{d}|h_i)$ indica o quão bem os dados são explicados pela hipótese.

$P(h_i|\mathbf{d})$ indica o novo grau de verossimilhança da hipótese, condicionada aos dados, ou seja, o quão bem a hipótese é explicada pelos dados.

Derivando o Teorema de Bayes

Da expressão da probabilidade conjunta:

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

Resulta:

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

Da expressão da probabilidade conjunta:

$$P(B|A) = P(BA)/P(A)$$

Resulta:

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

Como $P(AB) = P(BA)$, então:

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

E, assim, obtém-se:

$$P(A|B) = P(B|A)P(A) \,/\, P(B)$$

Teorema de Bayes

<u>Exemplo</u>:

Um médico sabe que meningite causa dor no pescoço em 50% dos casos. Ele sabe que a probabilidade a priori de um paciente ter meningite (M) é 1/50000 e a probabilidade a priori de qualquer paciente ter uma dor no pescoço (S) é 1/20.

Tem-se que:

$P(S|M) = 1/2$

$P(M) = 1/50000$

$P(S) = 1/20$

Um paciente chega ao consultório com dor no pescoço. Qual a probabilidade dele estar com meningite - $P(M|S)$?

$$P(M|S) = P(S|M)P(M) / P(S) = 1/2 * 1/50000 / 1/20 = 1 / 5000 = 0,0002$$

Teorema de Bayes

- Como já antecipado, a Lei da Probabilidade Total pode ser empregada na formulação do Teorema de Bayes, particularmente na definição de P(B) quando o evento B é formado por partições mutuamente exclusivas, produzindo:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_j P(B \mid A_j)P(A_j)}$$

Teorema de Bayes

Voltando ao exemplo do Controle de Qualidade:

Se durante a inspeção, encontramos um produto defeituoso, qual é a probabilidade que ele tenha sido produzido na fábrica II?

Resp.: Aqui, aplicaremos o Teorema de Bayes usando o item anterior para encontrar $P(A)$:

$$P(F_2|A) = \frac{P(A|F_2)P(F_2)}{P(A)} = \frac{0{,}02 * 0{,}45}{0{,}01575} = 0{,}5714$$

PARTE III – Redes Bayesianas

Representação do Conhecimento

Como representar o conhecimento em sistemas inteligentes que utilizam o raciocínio probabilístico?

Uma alternativa é usar **Redes Bayesianas.**

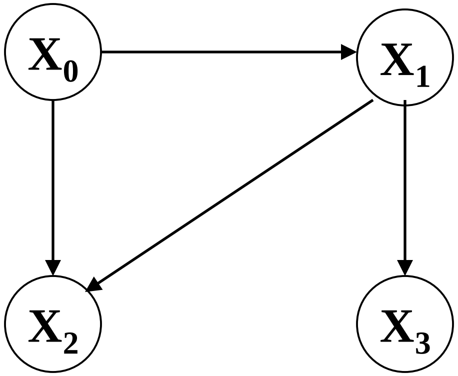
Redes Bayesianas

Rede Bayesiana é uma ferramenta gráfica para raciocínio e representação de conhecimento frente a incertezas.

Ela é uma representação compacta da distribuição de probabilidades conjuntas do universo do problema.

Redes Bayesianas

Formalmente, é um grafo acíclico direcionado



Os nós (também chamados de variáveis) são os eventos que queremos modelar.

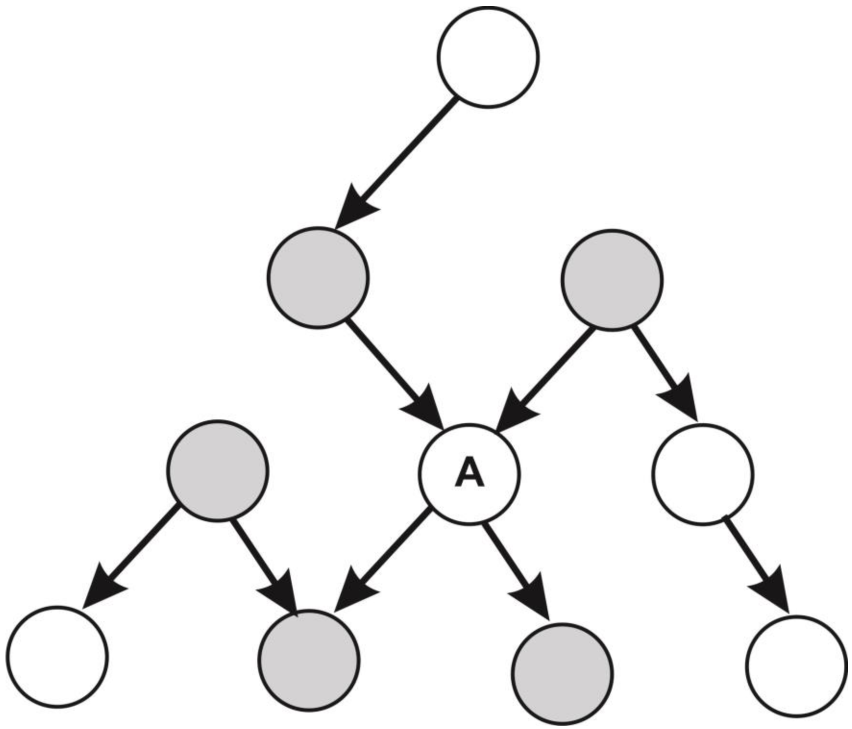
Os arcos ligando os nós indicam a presença de um dependência condicional entre eles.

Cada nó possui uma tabela de probabilidades, dizendo as chances de ocorrer o evento representado por ele (inclusive probabilidade condicional).

Markov Blanket

O <u>Markov Blanket</u> de um nó A é o conjunto de nós formados pelos pais do nó A, pelos filhos de A e pais dos filhos de A.

Um nó é independente de todos os outros nós da rede, dado seu Markov Blanket.

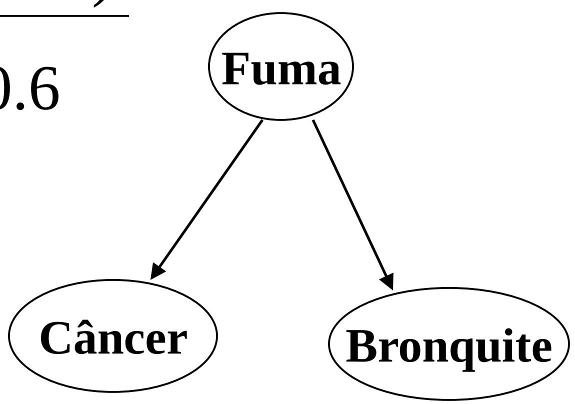


Aplicações

- Diagnóstico
  - Medicina, falhas em computadores
- Inferência
  - Economia, bolsa de valores
- Classificação
  - Reconhecimento de fala e de caracteres
- Mineração de dados
  - Bioinformática

Redes Bayesianas

P(F=s)	P(F=n)
0.4	0.6



F	P(C=s)	P(C=n)
s	0.7	0.3
n	0.1	0.9

F	P(B=s)	P(B=n)
s	0.8	0.2
n	0.3	0.7

Redes Bayesianas

Pela rede bayesiana do slide anterior, podemos responder perguntas do tipo:

Qual a probabilidade de uma pessoa ter câncer, sabendo que ela fuma?

$$P(C|F) = 0.7$$

Qual a probabilidade de uma pessoa ter bronquite, dado que ela fuma?

$$P(B|F) = 0.8$$

Qual a porcentagem de pessoas que fumam?

$$P(F) = 0.4$$

Inferência em Redes Bayesianas

Anteriormente, perguntamos: Qual a probabilidade de uma pessoa ter câncer, sabendo que ela fuma?

Agora faremos o contrário: Sabemos que uma pessoa está com câncer. Qual a probabilidade dessa pessoa ter fumado?

Reparem na diferença!!!

Inferência em Redes Bayesianas

Queremos encontrar:

$$P(F=\text{“s”}|C=\text{“s”}) = P(F=\text{“s”} \wedge C=\text{“s”}) / P(C=\text{“s”})$$

1ª parte:

$P(F=\text{“s”} \wedge C=\text{“s”}) = P(C=\text{“s”}|F=\text{“s”})\, P(F=\text{“s”}) = 0.4*0.7 = 0.28$

Inferência em Redes Bayesianas

2ª parte:

$$P(C=\text{“s”}) = \sum_i P(C=\text{“s”} \mid F=i)\, P(F=i)$$
$$= P(C=\text{“s”}|F=\text{“s”})\, P(F=\text{“s”}) + P(C=\text{“s”}|F=\text{“n”})\, P(F=\text{“n”})$$
$$= (0.7 * 0.4) + (0.1 * 0.6) = 0.34$$

Portanto, $P(F=\text{“s”} \mid C=\text{“s”}) = P(F=\text{“s”} \wedge C=\text{“s”}) / P(C=\text{“s”})$

$$= 0.28 / 0.34 = 0.82$$

Para uma pessoa que tem câncer, podemos inferir com 82% de certeza que ela fumava (**com base na rede bayesiana fornecida**).

Exercícios 1 e 2

Você instalou um alarme contra roubos na sua casa, que dispara em caso de invasão.

Infelizmente, o alarme é sensível a terremotos.

Quando o alarme disparar, seus 2 vizinhos, João e Maria, disseram que vão te ligar.

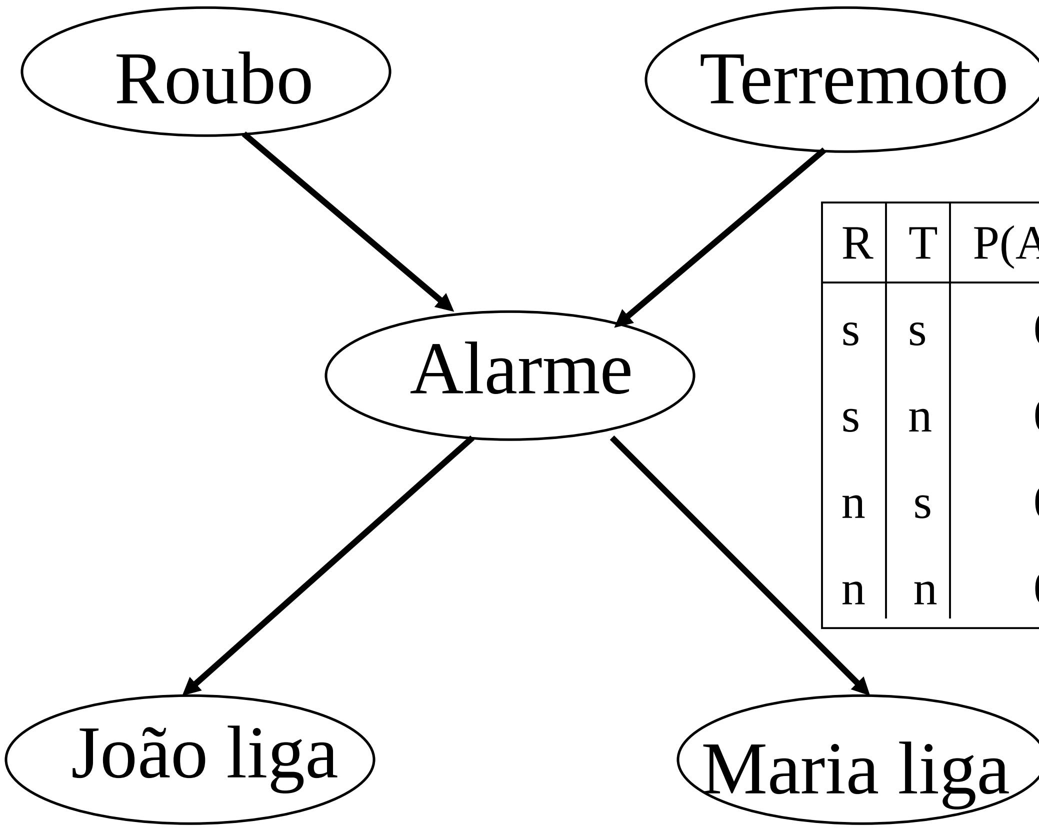
João, às vezes, confunde o alarme com a sirene do carro de bombeiro.

Maria ouve música num volume alto e nem sempre escuta o alarme.

Exercícios 1 e 2

P(R=s)
0.001

P(T=s)
0.002



R	T	P(A=s)
s	s	0.95
s	n	0.94
n	s	0.29
n	n	0.001

A	P(J=s)
s	0.90
n	0.05

A	P(M=s)
s	0.70
n	0.01

Exercício 1

Qual é a probabilidade de não haver roubo, nem terremoto, o alarme tocar, João ligar e Maria ligar?

- $P(\neg R \wedge \neg T \wedge A \wedge J \wedge M) =$

$= P(J \mid A)*P(M \mid A)*P(A \mid \neg R \wedge \neg T)*P(\neg R)*P(\neg T)$

$= 0.9 * 0.7 * 0.001 * 0.999 * 0.998$

$= 0.00062$ ou $0.062\ \%$

Exercício 2

João te ligou. Qual a probabilidade de estarem roubando sua casa?

$P(R \mid J) = P(R \wedge J) \,/\, P(J)$

Cálculo de $P(R \wedge J)$:

$P(R \wedge J) = P(R \wedge J \wedge A \wedge T) + P(R \wedge J \wedge A \wedge \neg T) +$

$P(R \wedge J \wedge \neg A \wedge T) + P(R \wedge J \wedge \neg A \wedge \neg T)$

$= 0.001*0.002*0.95*0.9+0.001*0.998*0.94*0.9+$

$0.001*0.002*0.05*0.05+0.001*0.998*0.06*0.05$

$= 0.00084902$

Exercício 2

Cálculo de P(J):

$P(J) = P(R \wedge T \wedge A \wedge J) + P(R \wedge T \wedge \neg A \wedge J) +$
$P(R \wedge \neg T \wedge A \wedge J) + P(R \wedge \neg T \wedge \neg A \wedge J) +$
$P(\neg R \wedge T \wedge A \wedge J) + P(\neg R \wedge T \wedge \neg A \wedge J) +$
$P(\neg R \wedge \neg T \wedge A \wedge J) + P(\neg R \wedge \neg T \wedge \neg A \wedge J)$

= 0.001*0.002*0.95*0.9+0.001*0.002*0.05*0.05+

0.001*0.998*0.94*0.9+0.001*0.998*0.06*0.05+

0.999*0.002*0.29*0.9+0.999*0.002*0.71*0.05+

0.999*0.998*0.001*0.9+0.999*0.998*0.999*0.05

= 0.0521

Exercício 2

João te ligou. Qual a probabilidade de estarem roubando sua casa?

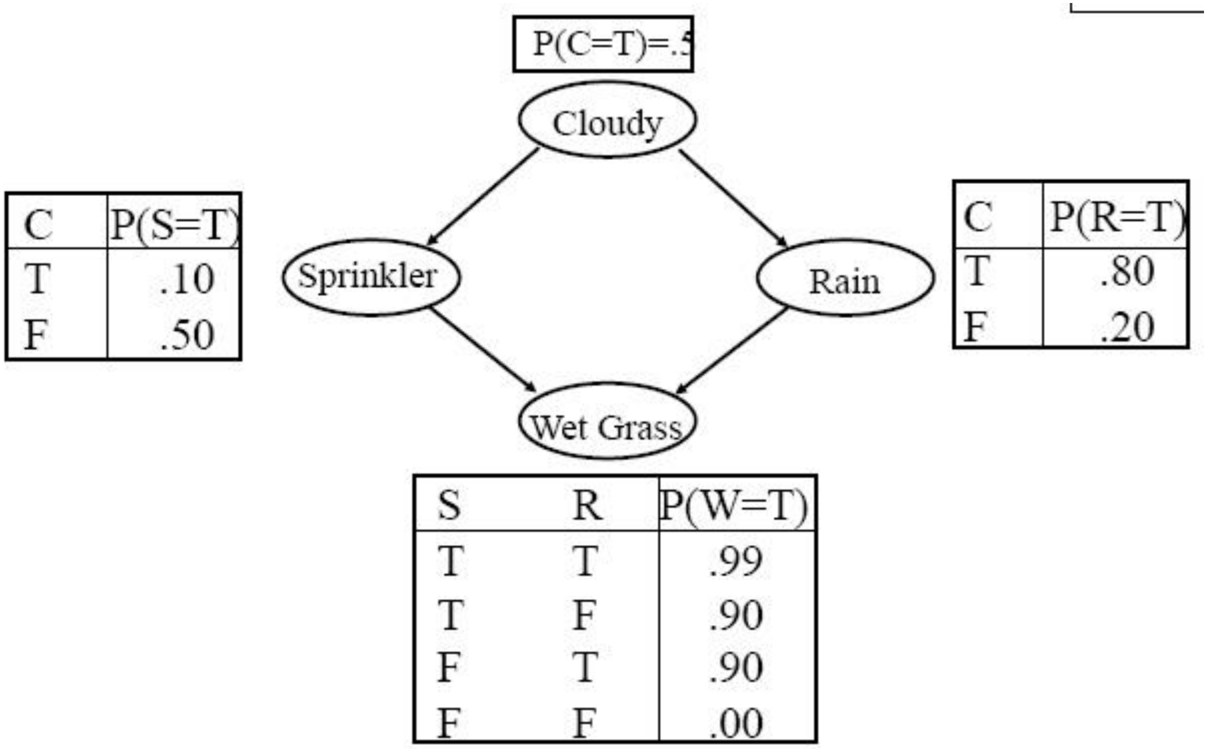
$P(R / J) = P(R \wedge J) / P(J)$

$= 0.00084902 / 0.0521 = 0.01628373 = 1,628373\%$

- Esta resposta vai contra a intuição, por ser um valor surpreendentemente baixo.
- No entanto, se lembrarmos que, em 1000 dias vai ocorrer em média um roubo e João vai ligar em média 51 vezes, então a chance de haver roubo quando ele liga já caiu para 2%. Considerando a possibilidade de terremoto e do alarme disparar sem motivo, chega-se ao resultado obtido.

Exercício 3

Considere uma rede bayesiana formada pelas seguintes variáveis: Cloudy (Tempo Nublado), Sprinklet (Regador), Rain (Chuva) e Wet Grass (Grama Molhada). Dado que observamos (temos a evidência de) que a grama está molhada, qual a probabilidade de ter chovido?



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

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

S	R	P(W=T)
T	T	.99
T	F	.90
F	T	.90
F	F	.00

Exercício 3

- Considerações: True=1 e False=0.
  - Cloudy(C)
  - Sprinkler (S)
  - Rain(R)
  - WetGrass(W)

- Queremos calcular: $P(R=1|W=1)$

Sabemos que a probabilidade condicional pode ser escrita como $P(A \mid B) = P(A \wedge B) / P(B)$. Por isso, usaremos a seguinte fórmula:

$$P(R=1 \mid W=1) = P(R=1 \wedge W=1) / P(W=1)$$

Exercício 3

$P(R=1 \wedge W=1) = \Sigma_{i,j}\; P(C = i, S = j, R = 1, W=1)$

$= P(C=1, S=1, R=1, W=1) + P(C=0, S=1, R=1, W=1) +$
$P(C=1, S=0, R=1, W=1) + P(C=0, S=0, R=1, W=1)$

$= (0.5 * 0.1 * 0.8 * 0.99) + (0.5 * 0.5 * 0.2 * 0.99) +$
$(0.5 * 0.9 * 0.8 * 0.9) + (0.5 * 0.5 * 0.2 * 0.9)$

$= (0.0396) + (0.0495) + (0.324) + (0.045)$

$=0.4581$

Exercício 3

$P(W=1) = \Sigma_{i,j,k} P(C = i, S = j, R = k, W=1)$

$= P(C=1, S=1, R=1, W=1) + P(C=1, S=1, R=0, W=1) +$

$P(C=1, S=0, R=1, W=1) + P(C=0, S=0, R=0, W=1) +$

$P(C=0, S=1, R=1, W=1) + P(C=0, S=1, R=0, W=1) +$

$P(C=1, S=0, R=0, W=1) + P(C=0, S=0, R=1, W=1)$

$= (0.5 * 0.1 * 0.8 * 0.99) + (0.5 * 0.1 * 0.2 * 0.9) +$

$(0.5 * 0.9 * 0.8 * 0.9) + (0.5 * 0.5 * 0.8 * 0) +$

$(0.5 * 0.5 * 0.2 * 0.99) + (0.5 * 0.5 * 0.8 * 0.9) +$

$(0.5 * 0.9 * 0.2 * 0) + (0.5 * 0.5 * 0.2 * 0.9)$

$= (0.0396) + (0.009) + (0.324) + (0) + (0.0495) + (0.18) + (0) + (0.045)$

$= 0.6471$

Exercício 3

$P(R=1 \mid W=1) = P(R=1 \wedge W=1) / P(W=1)$

$= 0.4581 / 0.6471$

$= 0.7079$

- Portanto, dado que observamos a grama molhada (evidência), a probabilidade de ter chovido é de 70.79%.

Exercício 4

- Construa uma rede bayesiana capaz de modelar o cenário de incertezas descrito a seguir (não é necessário atribuir valores para as probabilidades envolvidas:

Eu quero prever se minha esposa está em casa antes de eu abrir a porta. Eu sei que minha esposa liga a luz quando chega em casa, mas às vezes ela também liga ao sair se ela for retornar com alguma visita. Quando não tem ninguém em casa, ela solta o cachorro no quintal, mas às vezes ela solta o cachorro quando ele está molhado. Quando o cachorro está solto, consigo ouvir o seu latido da rua, mas às vezes o confundo com o cachorro da vizinha.

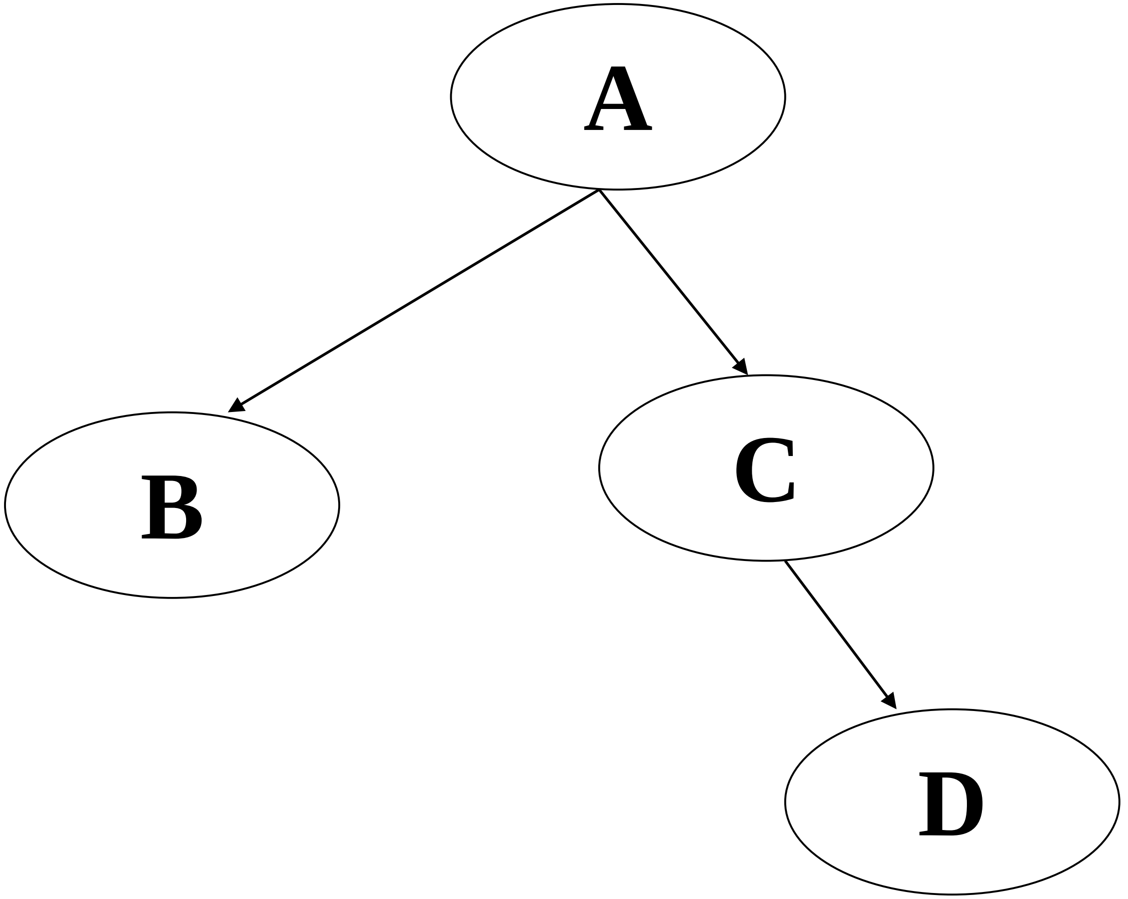
Exercício 4

- É necessário definir os eventos que têm alguma probabilidade de ocorrência:

- **A → Minha esposa em casa**
- **B → Luz de minha casa acesa**
- **C → Meu cachorro solto**
- **D → Meu cachorro latindo**

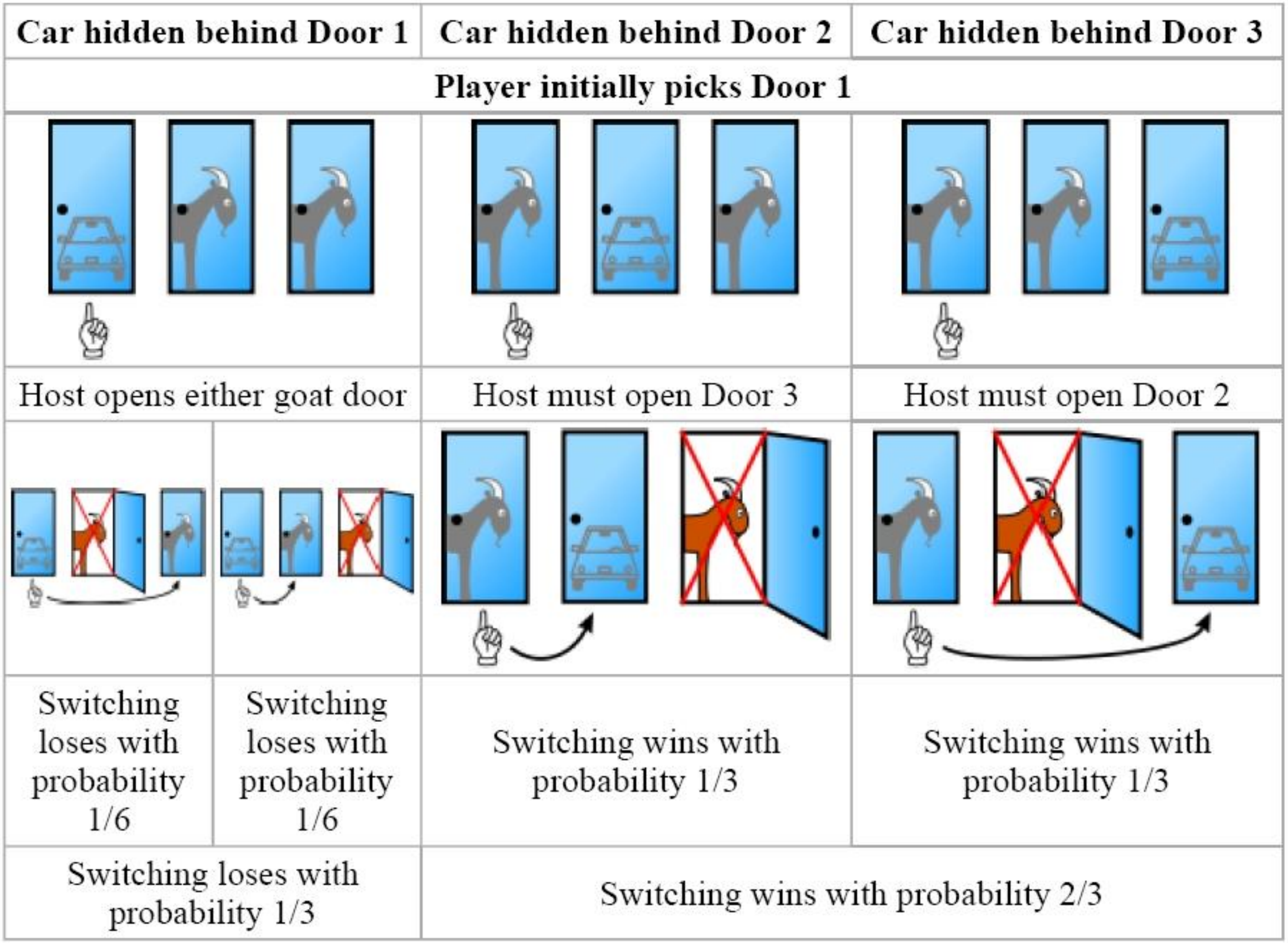
- Em seguida, propõe-se a topologia de rede bayesiana capaz de explicitar os relacionamentos condicionais existentes entre esses eventos.

Exercício 4



Exercício 5

Monty Hall Problem



Vantagens

- **A maior vantagem do raciocínio probabilístico em relação ao raciocínio fundamentado em lógica dedutiva é permitir chegar a decisões racionais mesmo quando não há informação suficiente para provar qualquer das hipóteses.**
- Raciocínio probabilístico é capaz de tratar o grau de incerteza de domínios bem variados, além de propor representações parcimoniosas.
- Redes bayesianas podem ser diretamente geradas a partir de dados coletados de processos do mundo real.

Limitações

Algumas dificuldades para representar o conhecimento usando redes bayesianas:

- Requer conhecimento a priori.
- Número de variáveis do problema (nós) é alto.
- A relação entre as variáveis pode ser muito complexa.
- Redes bayesianas não modelam comportamentos dinâmicos, o que requer adaptações do modelo (Redes bayesianas dinâmicas).

Naïve Bayes Classifier

- Suponha uma função de classificação f: X $\rightarrow$ V, onde cada instância x é descrita pelos atributos $\{a_1, \ldots, a_n\}$
- O valor mais provável de f(x) é

$$
\begin{aligned}
v_{MAP} &= \underset{v_j \in V}{\mathbf{arg\,max}}\, P(v_j / a_1, \cdots, a_n) \\
&= \underset{v_j \in V}{\mathbf{arg\,max}}\, \frac{P(a_1, \cdots, a_n / v_j) P(v_j)}{P(a_1, \cdots, a_n)} \\
&= \underset{v_j \in V}{\mathbf{arg\,max}}\, P(a_1, \cdots, a_n / v_j) P(v_j)
\end{aligned}
$$

- Suposição Bayesiana Ingênua $\quad P(a_1, \cdots, a_n / v_j) = \prod_i P(a_i / v_j)$

- Classificador Bayesiano Ingênuo

$$v_{NB} = \underset{v_j \in V}{\mathbf{arg\,max}}\, P(v_j) \prod_i P(a_i / v_j)$$

PARTE IV – Aprendizado em Redes Bayesianas

Construção de Redes Bayesianas

Para se projetar uma rede bayesiana, geralmente utiliza-se um especialista no domínio do problema em questão.

Porém:

Especialista no problema nem sempre está disponível.

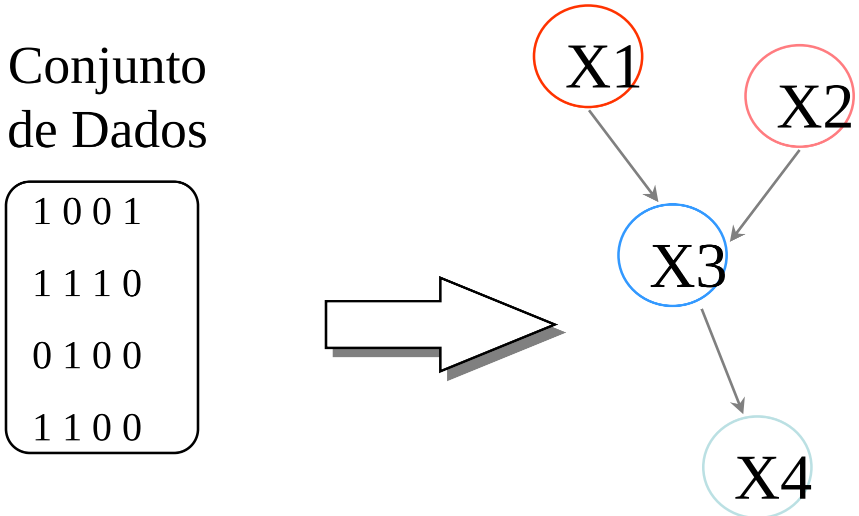
Ele conhece parcialmente o problema.

Tempo para projetar a rede é demorado.

Construção de Redes Bayesianas

Solução:

Modelar uma RB a partir de um conjunto de dados que representam amostras do problema.



Aprendizado em Redes Bayesianas

Abordagem baseada em busca e pontuação

A aprendizagem se dá por meio da **busca** de uma rede que seja capaz de **representar** o conjunto de dados da melhor forma.

É preciso, então:

1) Um mecanismo de busca

2) Uma forma de avaliar cada solução candidata

Aprendizado em Redes Bayesianas

Abordagem baseada em busca e pontuação

1) Mecanismo de busca

Consiste em encontrar diferentes tipos de redes para o problema em questão. A busca utiliza meta-heurísticas, uma vez que o problema de otimização correspondente é NP-completo.

Exemplos de meta-heurísticas que podem ser utilizadas como mecanismos de busca: hill climbing, algoritmos genéticos, sistemas imunológicos artificiais, colônia de formigas.

Aprendizado em Redes Bayesianas

Abordagem baseada em busca e pontuação

2) Forma de avaliar a rede

São métodos que calculam a verossimilhança da rede a partir de dados amostrados, podendo incorporar algum termo penalizando modelos complexos. São métodos para **seleção de modelos**.

Bayesian Information Criterion (BIC);

Akaike Information Criterion (AIC);

Minimum Description Length (MDL).

Aprendizado em Redes Bayesianas

Bayesian Information Criterion (BIC)

$$\mathrm{BIC}(B) = \log\ P(D|B) - \frac{1}{2}\log N * \dim(B)$$

sendo,

$$P(D|B) = \prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{(r_i-1)!}{(N_{ij}+r_i-1)!}\prod_{k=1}^{r_i}N_{ijk}!$$

B representa um rede bayesiana, dim(*B*) é o número de parâmetros da rede, *N* é o número de amostras, *n* é o número de variáveis, q_i denota o número de possíveis valores que os pais de x_i podem assumir, r_i é o número de possíveis valores de x_i , N_{ijk} é o número de casos em que x_i assume o *k*-ésimo valor com seus pais assumindo o *j*-ésimo valor *e* $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$

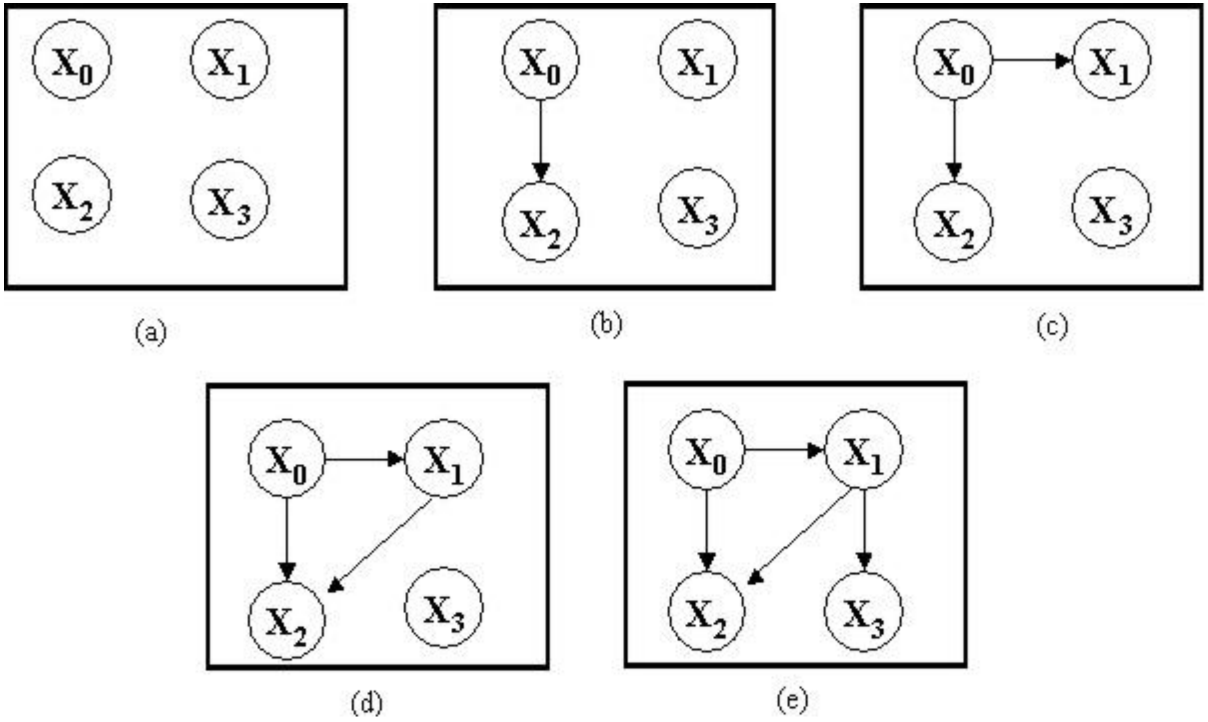
Aprendizado em Redes Bayesianas

Abordagem baseada em busca e pontuação

O aprendizado começa, geralmente, com um grafo sem arcos. Em seguida, ele é avaliado. O próximo passo é perturbar a estrutura previamente encontrada e avaliá-la novamente. Este processo continua até que nenhuma outra estrutura melhor que a anterior seja encontrada.

Aprendizado em Redes Bayesianas

Exemplo de busca:



Redes Bayesianas na Prática

PATHFINDER: diagnóstico de doenças que atacam os nodos linfáticos. (Russel&Norvig, 2009)

PAINULIM: diagnóstico de doenças neuro-musculares:
http://www.cis.uoguelph.ca/~yxiang/paper/ismm91.pdf

Tutores inteligentes: http://www.andestutor.org/

Confira mais links em:
Berger, J.O. “Bayesian Analysis: A Look at Today and Thoughts of Tomorrow”, 1999.

Softwares

http://www.bayesia.com/

http://www.cs.waikato.ac.nz/ml/weka/

http://www.norsys.com/netica.html

http://www.cs.ubc.ca/~murphyk/Software/bnsoft.html

http://www.stat.duke.edu/~mw/bats.html (Análise de séries temporais)

Referências

- Darwiche, A. (2009). Modeling and Reasoning with Bayesian Networks, Cambridge University Press.
- Koller, D. & Friedman, N. (2009). Probabilistic Graphical Models: Principles and Techniques, The MIT Press.
- Mitchell, T. (1997). Machine Learning, McGraw-Hill.
- Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann.
- Russel, S, & Norvig, P. (2009). Artificial Intelligence: A Modern Approach, Prentice-Hall, 3rd. Edition.