Fundamentos de Aprendizagem de Máquina

Hélio Pio

Programação das Aulas

Tópico 1: Introdução a Inteligência Artificial
Tópico 2: Agentes Inteligentes
Tópico 3: Fundamentos de Aprendizagem de Máquina
Tópico 4: Redes Neurais Artificiais
Tópico 5: Atividade em Aula – Primeira Avaliação
Tópico 6: Representação da Incerteza e Lógica Fuzzy
Tópico 7: Redes Bayesianas
Tópico 8: Support Vector Machines
Tópico 9: Atividade em Aula – Segunda Avaliação
Tópico 10: Resolução de Problemas por Meio de Busca e Otimização
Tópico 11: Técnicas de Ensemble
Tópico 12: Atividade em Aula – Terceira Avaliação

O que são redes Bayesianas?

Redes Bayesianas

O que são redes Bayesianas?

Problema de Monte Hall
Let's Make a Deal, 1970.

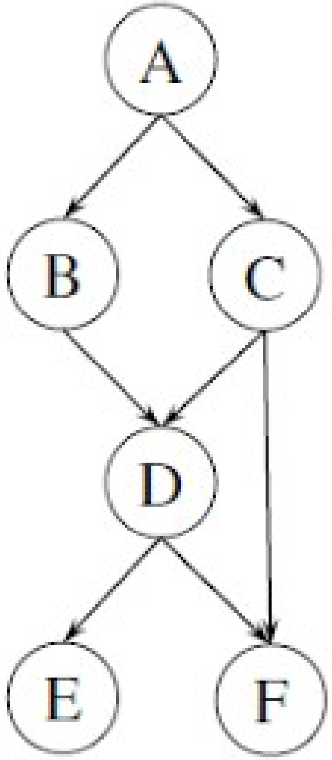
Redes Bayesianas

O que são redes Bayesianas?

Redes Bayesianas

O que são redes Bayesianas?

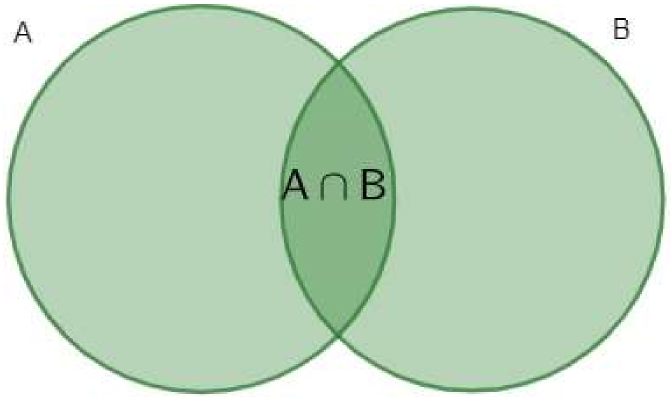
- Redes Bayesianas são modelos estatísticos que representam incertezas e dependências entre variáveis. Elas consistem em nós e arestas, onde cada nó é uma variável e cada aresta indica uma dependência probabilística.



Redes Bayesianas

Axiomas da probabilidade

- O axioma da probabilidade é um conjunto de regras fundamentais que regem a teoria das probabilidades. Esses axiomas estabelecem que:
  - A probabilidade de um evento é sempre não-negativa ($0 \leq P(A) \leq 1$),
  - A probabilidade da união de eventos mutuamente exclusivos é a soma das probabilidades individuais ($P(A \cup B) = P(A) + P(B) - (P(A \cap B))$),
  - A probabilidade do espaço amostral é igual a 1 ($P(S) = 1$).



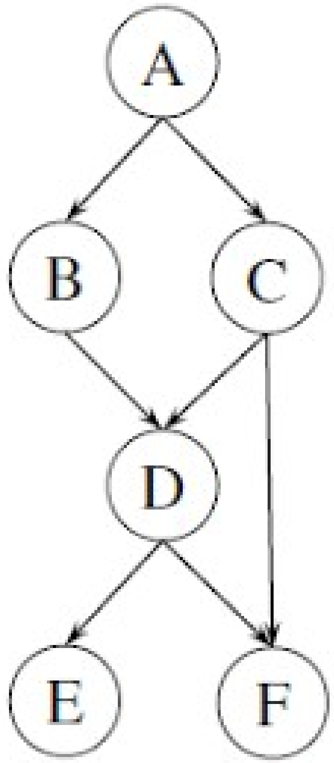
Redes Bayesianas

Probabilidade

- A probabilidade de uma proposição é igual à soma das probabilidades dos eventos atômicos em que ela é válida:

$$P(a) = \sum_{ei \; \epsilon \; e(a)} P(ei)$$

- Essa equação permite calcular a probabilidade de qualquer proposição dada uma distribuição conjunta total que especifique todos os eventos atômicos.



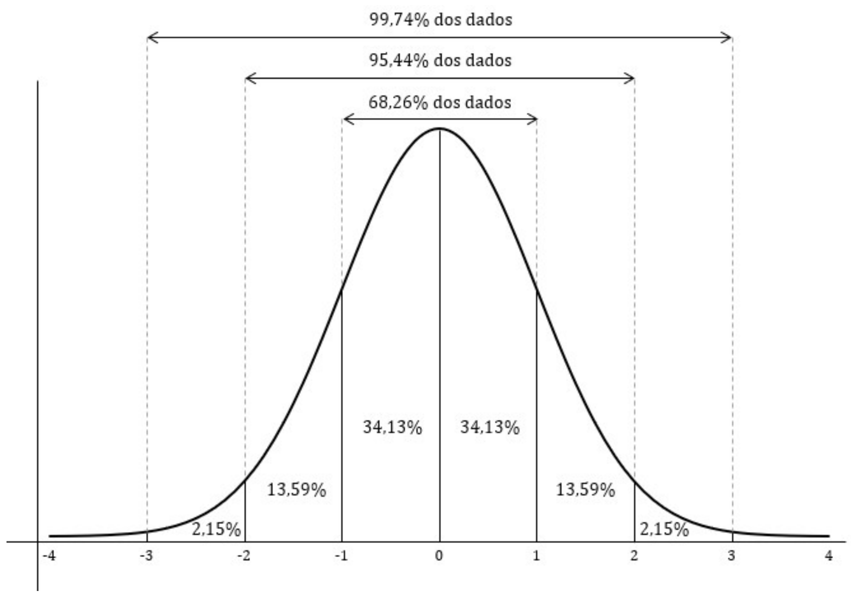
Redes Bayesianas

Probabilidade incondicional ou 'a priori'

- É a probabilidade de ocorrência de um evento, independentemente de qualquer informação adicional.

- É o grau de crença em uma proposição na ausência de outras informações.
  - P(Cárie = verdadeiro) = 0.1
  - P(Clima = ensolarado) = 0.72

- Distribuição de probabilidades
  - Dá probabilidades a todos os valores possíveis de uma variável aleatória.
  - Exemplo: a distribuição normal.



Redes Bayesianas

Probabilidade conjunta

- É a probabilidade de dois ou mais eventos ocorrerem simultaneamente.

- Probabilidades de todas as combinações de valores de um conjunto de variáveis aleatórias.

- Uma distribuição conjunta total especifica a probabilidade de qualquer evento atômico.
    - Qualquer probabilidade nesse domínio pode ser calculada a partir da distribuição conjunta total.

P(Clima,*Carie*) = tabela 4 × 2 de valores:

Clima =	ensolarado	chuvoso	nublado	neve
Cárie = verdadeiro	0.144	0.02	0.016	0.02
Cárie = falso	0.576	0.08	0.064	0.08

Redes Bayesianas

Probabilidade condicional ou ‘a posteriori’

- É a probabilidade de um evento ocorrer, dado que outro evento já ocorreu.

- É o grau de crença em uma proposição dada a presença de evidências (valores de variáveis aleatórias conhecidos). Exemplos:
  - P(Cárie = verdadeiro | DorDeDente = verdadeiro) = 0.8
  - P(Cárie = verdadeiro | DorDeDente = verdadeiro, Cárie = verdadeiro) = 1
  - P(Cárie = verdadeiro | DorDeDente = verdadeiro, Ensolarado = verdadeiro) = P(Cárie = verdadeiro | DorDeDente) = 0.8

- Distribuição condicional
  - P(Y|X) fornece o valor de P(Y=yj| X=xi) para cada valor de i e j possíveis.

Redes Bayesianas

Probabilidade condicional ou 'a posteriori'

- Pode ser definida em termos de probabilidades a priori:
  $\mathrm{P}(a|b) = \frac{P(a \cap b)}{P(b)}$ se P(b)>0.

- Regra do produto dá uma definição alternativa:
  $$\mathrm{P}(\mathrm{a} \cap b) = P(a|b).P(b) = P(b|a).P(a)$$

- Isso pode ser generalizado para distribuições totais:
  $P(Clima, Cárie) = P(Clima|Cárie).P(Cárie)$ (que é um conjunto de 4x2 equações, não uma multiplicação matricial).

- Regra da cadeia é obtida a partir de aplicações sucessivas da regra do produto:
  $P(x1, \ldots, xn) = \prod_{i=1}^{n} P(xi|x1, \ldots, xi-1)$

Redes Bayesianas

Probabilidade condicional ou ‘a posteriori’

- Um exemplo de probabilidade condicional é o lançamento de um dado honesto de seis faces. Suponha que queremos calcular a probabilidade de obter um número par, dado que o número obtido é maior do que 3. Podemos usar a fórmula da probabilidade condicional para determinar essa probabilidade específica.

$\mathrm{P}(a|b) = \frac{P(a \cap b)}{P(b)}$ se P(b)>0.

$$\mathrm{P}(número\ par|número > 3) = \frac{P(número\ par\ e\ número > 3)}{P(númer\qquad)}$$

Redes Bayesianas

Probabilidade condicional ou ‘a posteriori’

$$\mathrm{P}(número\ par|número > 3) = \frac{P(número\ par\ e\ número{>}3)}{P(número{>}3)}$$

- Números pares maiores que três: 4 e 6 (2 eventos)
- Números maiores que três: 4, 5 e 6 (3 eventos)

$$\mathrm{P}(número\ par|número > 3) = \frac{2}{3}$$

- Isso significa que, se sabemos que o número obtido é maior que 3, a chance de obter um número par é de aproximadamente 66.67%.

Redes Bayesianas

Inferência probabilística

- Inferência probabilística: a computação a partir de evidências observadas de probabilidades posteriores para proposições de consulta.

- Inferência com o uso de distribuições conjuntas totais: base de conhecimento a partir da qual são derivadas respostas para todas as consultas.

Redes Bayesianas

Exemplo: Inferência probabilística

- Suponha um domínio com a distribuição conjunta total da figura ao lado.

- Para qualquer proposição 'a', P(a) é a soma dos eventos atômicos 'w' onde 'a' ocorre:

$$P(a) = \sum_{w:w \models a} P(w)$$

	dordedente		¬dordedente	
	boticão	¬boticão	boticão	¬boticão
cárie	.108	.012	.072	.008
¬cárie	.016	.064	.144	.576

- P(dordedente) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2
- P(dordedente V cárie) = 0.108 + 0.012 + 0.016 + 0.064 + 0.072 + 0.008 = 0.28

Redes Bayesianas

Exercício – Calcule P(~Cárie|dordedente)

	dordedente		¬dordedente	
	botição	¬botição	botição	¬botição
cárie	.108	.012	.072	.008
¬cárie	.016	.064	.144	.576

$$P(\sim Cárie|dordedente) = \frac{P(\sim Cárie) \cap P(dordedente)}{P(dordedente)}$$

$$P(\sim Cárie|dordedente) = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064}$$

$$P(\sim Cárie|dordedente) = 0.4$$

Redes Bayesianas

Teorema de Bayes

- O reverendo 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.

Redes Bayesianas

Teorema de Bayes

- Equação do Teorema de Bayes:

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

- Esta simples equação é uma ferramenta útil para inferir a probabilidade a posteriori de um evento baseado na evidência e num conhecimento a priori de outros eventos.

Redes Bayesianas

Teorema de Bayes

- Exemplo:

Um médico sabe que a 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 - $P(M|S)$? Relembrando: $P(A|B) = P(B|A)P(A) / P(B)$.

$P(M|S) = P(S|M)P(M) / P(S) = 1/2 * 1/50000 / 1/20 = 0.0002$

Redes Bayesianas

Representação do conhecimento

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

- Uma alternativa é usar Redes Bayesianas!!

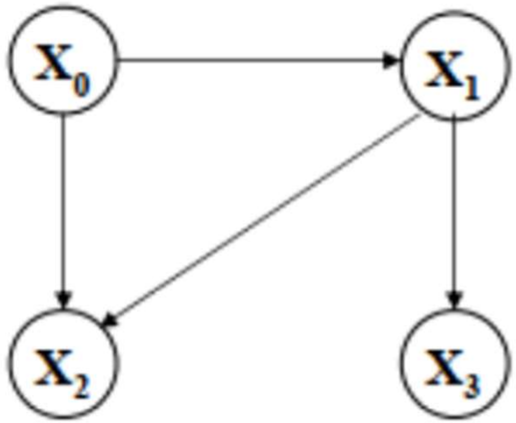
- Rede Bayesiana é uma ferramenta gráfica para raciocínio e representação de conhecimento frente a incertezas.

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

Redes Bayesianas

Representação do conhecimento

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



Redes Bayesianas

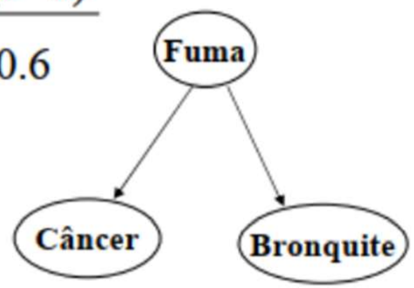
Representação do conhecimento

- Qual a probabilidade de alguém ter câncer, sabendo que fuma?
  P(C|F) = ? 0.7

- Qual a probabilidade de alguém ter bronquite, dado que fuma?
  P(B|F) = ? 0.8

- Qual a porcentagem de pessoas que fumam?
  P(F) = ? 0.4

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

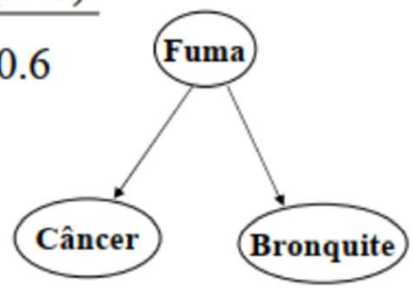
Representação do conhecimento

- Anteriormente, perguntamos: Qual a probabilidade de alguém 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!!!

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

Representação do conhecimento

P(F=‘s’|C=‘s’) = P(F=‘s’ e C=‘s’) / P(C=‘s’)

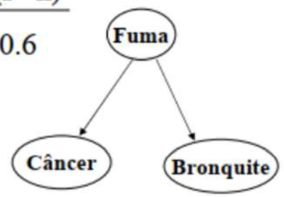
Primeira parte:

P(F=‘s’ e C=‘s’) = P(C=‘s’|F=‘s’) * P(F=‘s’)

P(F=‘s’ e C=‘s’) = 0.7 * 0.4

P(F=‘s’ e C=‘s’) = 0.28

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

Representação do conhecimento

P(F=‘s’|C=‘s’) = P(F=‘s’ e C=‘s’) / P(C=‘s’)

Segunda parte:

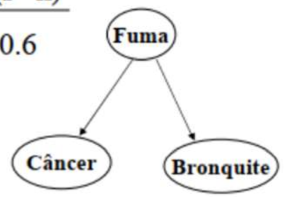
P(C=‘s’) = Σi P(C=‘s’|F=‘i’)*P(F=‘i’)

P(C=‘s’) = P(C=‘s’|F=‘s’)* P(F=‘s’) + P(C=‘s’|F=‘n’)* P(F=‘n’)

P(C=‘s’) = (0.7 * 0.4) + (0.1 * 0.6)

P(C=‘s’) = 0.34

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

Representação do conhecimento

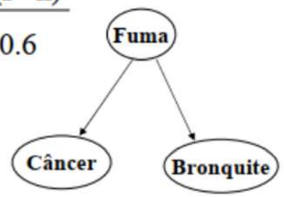
P(F='s'|C='s') = P(F='s' e C='s') / P(C='s')

Portanto:

P(F='s'|C='s') = 0.28 / 0.34 = 0.82

Para uma pessoa que tem câncer, podemos inferir com 82% de certeza que ela fuma.

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

Aplicações

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

Redes Bayesianas

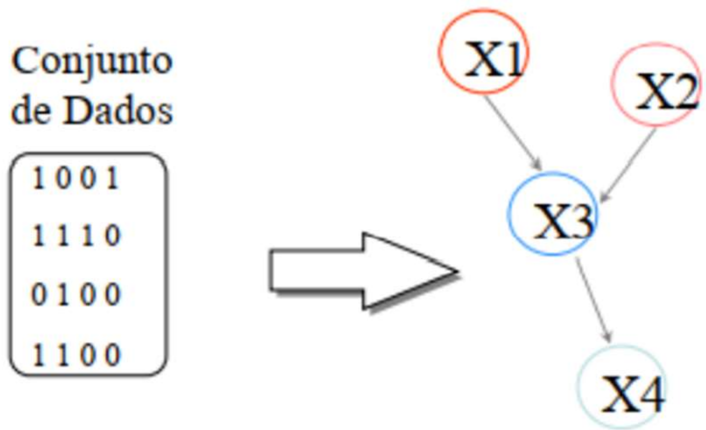
Construção de Redes Bayesianas

- 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.

Redes Bayesianas

Construção de Redes Bayesianas

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



Redes Bayesianas

Aprendizado de 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:
  - Um mecanismo de busca.
  - Uma forma de avaliar cada solução candidata.

Redes Bayesianas

Aprendizado de Redes Bayesianas - Abordagem baseada em busca e pontuação

- Mecanismo de busca:
  - Consiste em encontrar diferentes tipos de redes para o problema em questão. A busca utiliza heurísticas, uma vez que varrer todo o espaço de busca é um problema computacional difícil.
  - Estão sendo utilizados como mecanismos de busca: *hill climbing*, algoritmos genéticos, sistemas imunológicos artificiais, colônia de formigas.

Redes Bayesianas

Aprendizado de Redes Bayesianas - Abordagem baseada em busca e pontuação

- Forma de avaliar a rede:
  - São métodos que calculam a verossimilhança da rede com os dados, podendo incorporar algum termo penalizando modelos complexos. São métodos para seleção de modelos.
    - ✓ *Bayesian Information Criteron* (BIC).
    - ✓ *Akaike Information Criterion* (AIC).
    - ✓ *Minimum Description Length* (MDL).

Redes Bayesianas

Aprendizado de Redes Bayesianas - Abordagem baseada em busca e pontuação

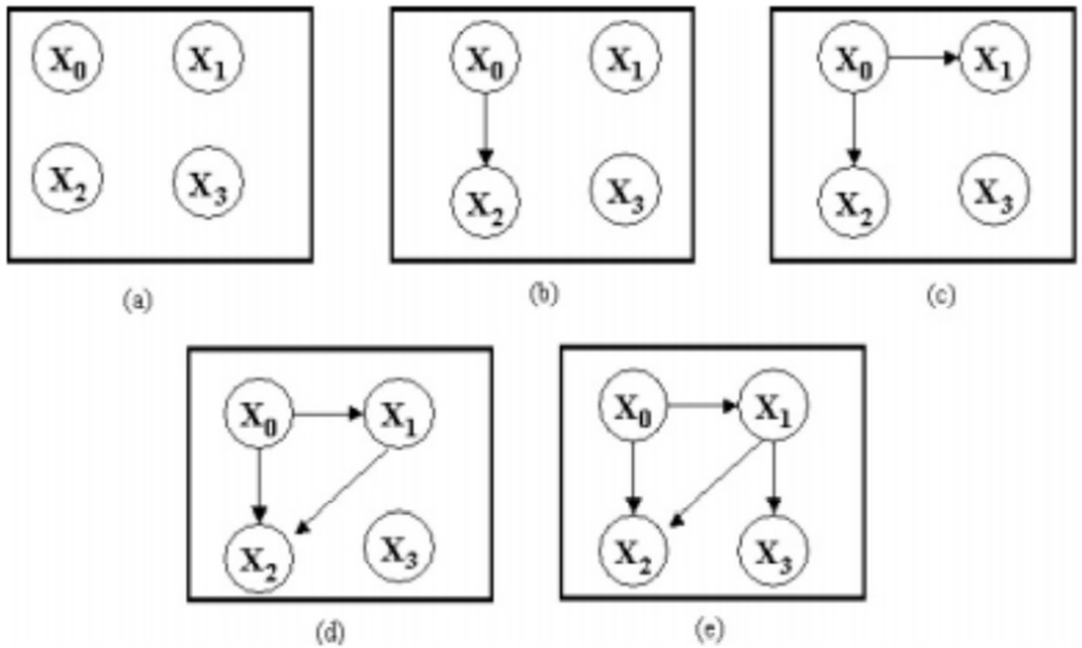
- O aprendizado começa com um grafo sem arcos. Em seguida, ele é avaliado. O próximo passo é perturbar a estrutura previamente encontrada e avaliá-la novamente.

- Este processo continua até que nenhuma outra estrutura melhor que a anterior seja encontrada.

Redes Bayesianas

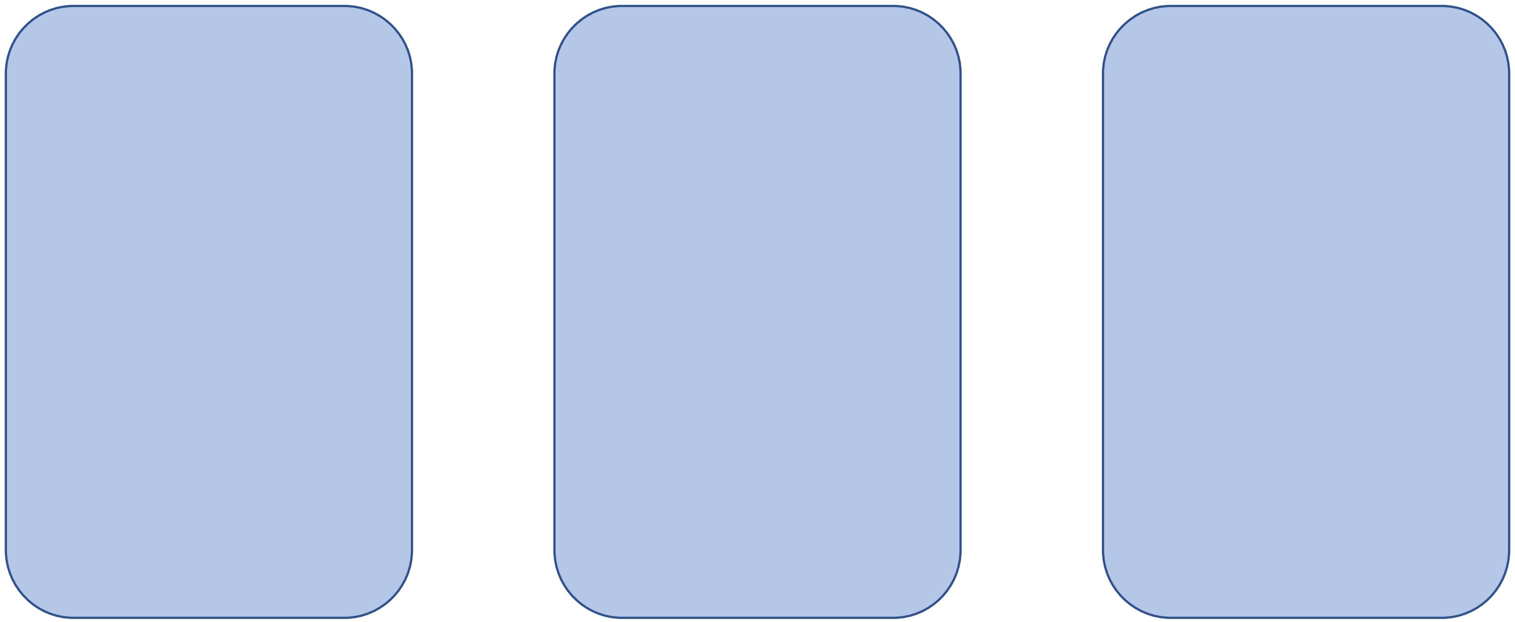
Aprendizado de Redes Bayesianas

- Exemplo de busca:



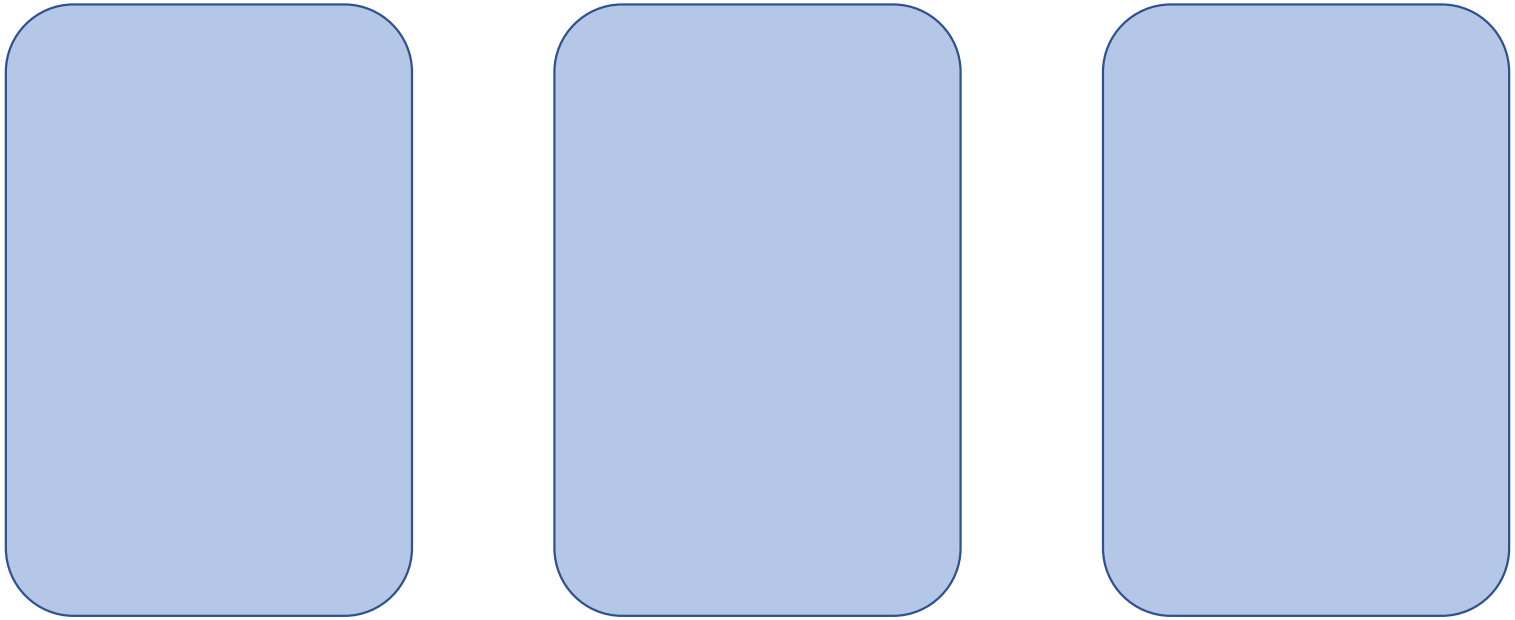
Redes Bayesianas

O problema de MONTY HALL



Redes Bayesianas

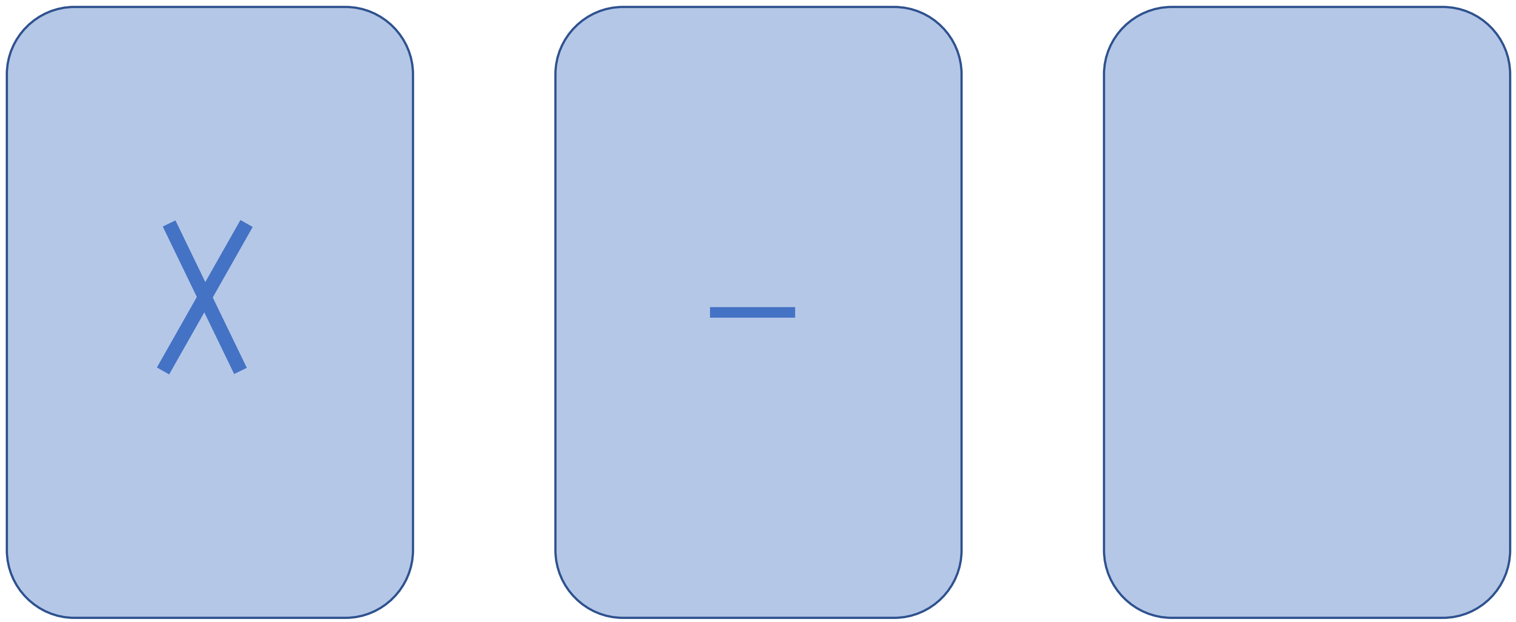
O problema de MONTY HALL



1/3 1/3 1/3

Redes Bayesianas

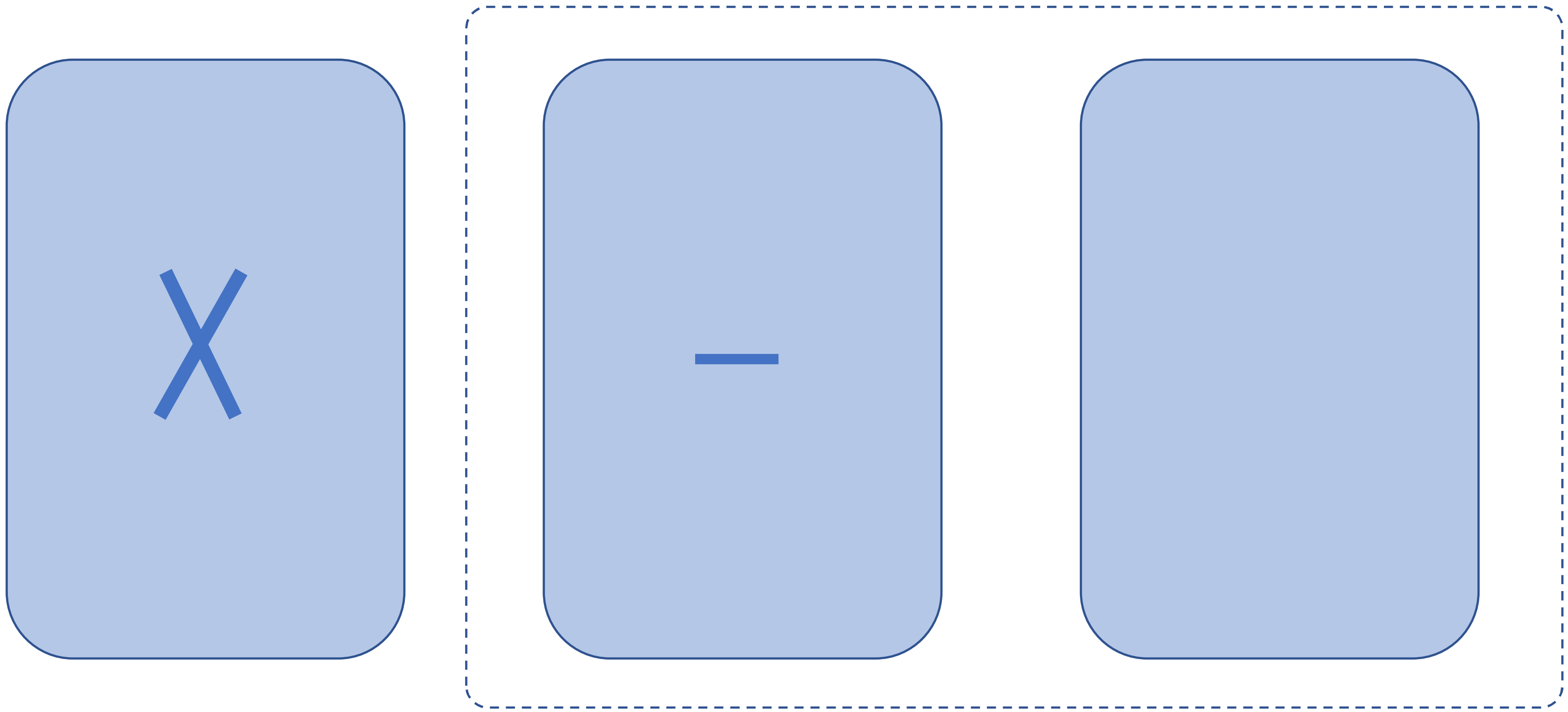
O problema de MONTY HALL



1/3 1/3 1/3

Redes Bayesianas

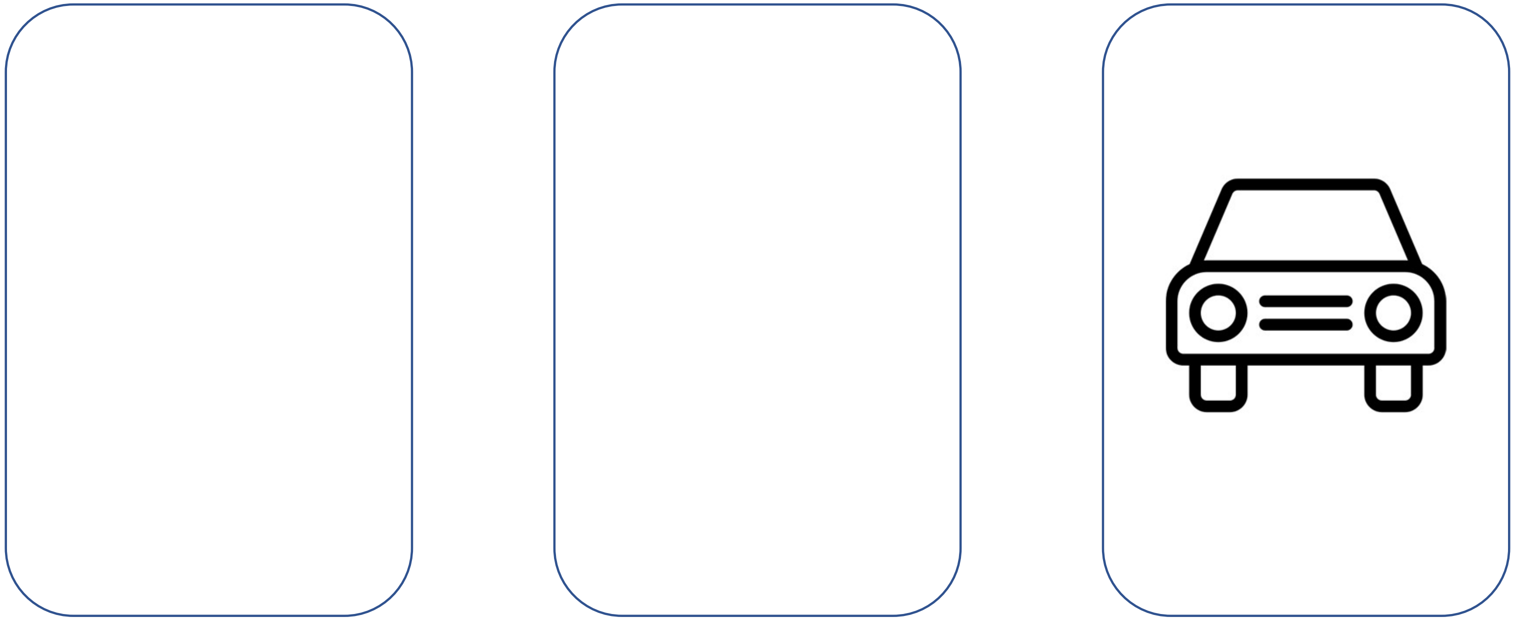
O problema de MONTY HALL



1/3 2/3

Redes Bayesianas

O problema de MONTY HALL



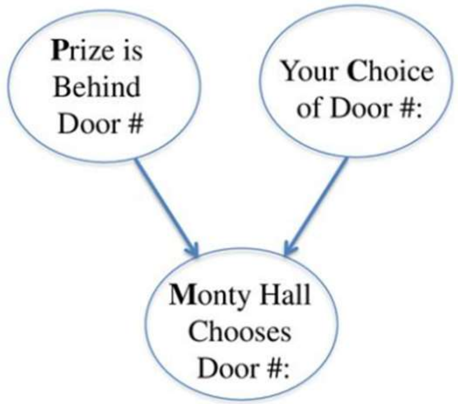
Redes Bayesianas

O problema de MONTY HALL

Como usar as redes Bayesianas para resolver esta questão?

Prize is Behind Door #:		Pr(P)
	door 1	0.333
	door 2	0.333
	door 3	0.333

Your Choice of Door #:		Pr(C)
	door 1	0.333
	door 2	0.333
	door 3	0.333



Pr(M|P,C)	Prize is Behind Door #:	door 1	door 1	door 1	door 2	door 2	door 2	door 3	door 3	door 3
	Your Choice of Door #:	door 1	door 2	door 3	door 1	door 2	door 3	door 1	door 2	door 3
Monty Hall Chooses Door #:	door 1	0	0	0	0	0.5	1	0	1	0.5
	door 2	0.5	0	1	0	0	0	1	0	0.5
	door 3	0.5	1	0	1	0.5	0	0	0	0

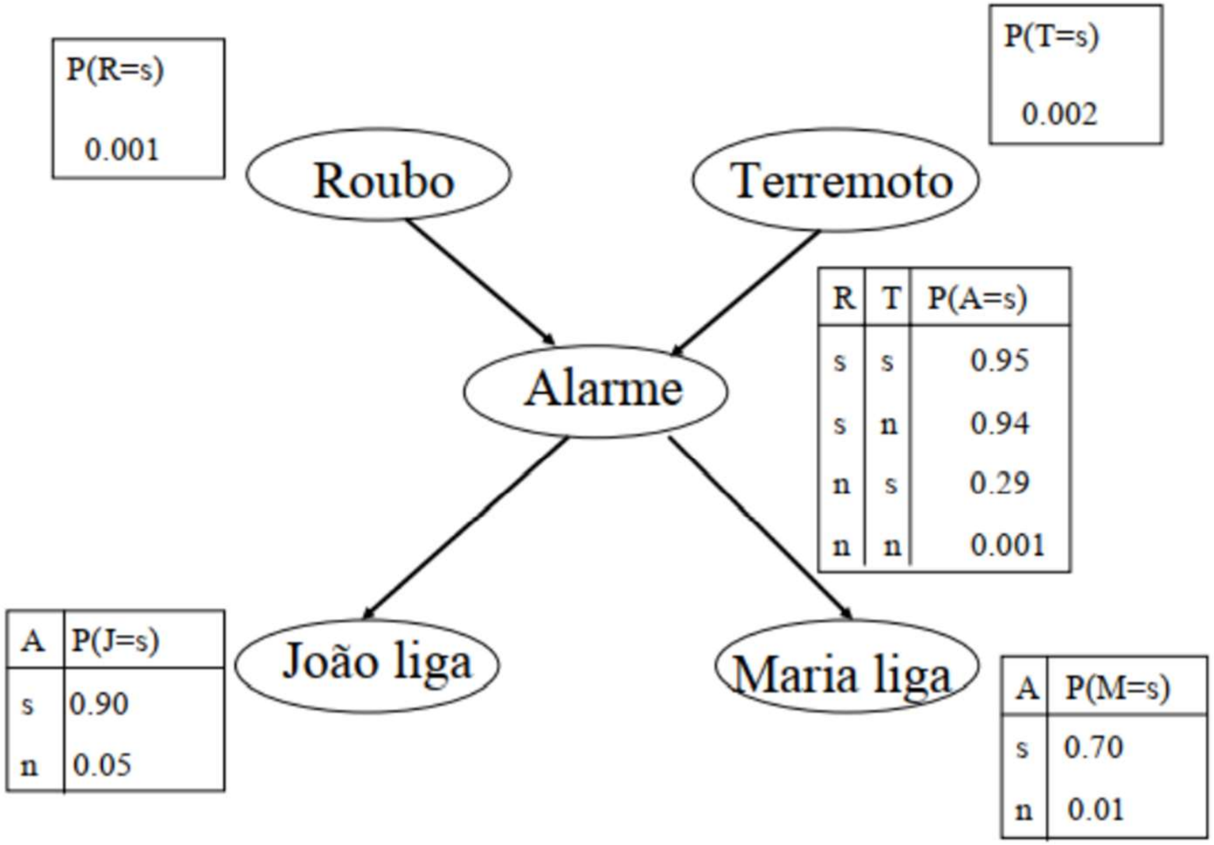
Redes Bayesianas

Exercício

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

Redes Bayesianas

Exercício



E agora?

Comentários