Inteligência Artificial – ACH2016
Aula 16 – Regra de Bayes e Redes Bayesianas

Norton Trevisan Roman
(norton@usp.br)

12 de maio de 2019

Regra de Bayes

$$\left.\begin{array}{l} P(a \wedge b) = P(a|b)\,P(b) \\ P(a \wedge b) = P(b|a)\,P(a) \end{array}\right\} \rightarrow P(b|a) = \frac{P(a|b)\,P(b)}{P(a)}$$

$$\Downarrow$$

$$\left.\begin{array}{l} P(a \wedge b \wedge c) = P(a|b \wedge c)\,P(b|c)\,P(c) \\ P(a \wedge b \wedge c) = P(b|a \wedge c)\,P(a|c)\,P(c) \end{array}\right\} \rightarrow P(b|a \wedge c) = \frac{P(a|b \wedge c)\,P(b|c)}{P(a|c)}$$

Regra de Bayes

Exemplo

- Considere o seguinte problema:

Regra de Bayes

Exemplo

- Considere o seguinte problema:
  - 1% das mulheres na idade de 40 anos que fazem exames de rotina têm câncer de mama.

Regra de Bayes

Exemplo

- Considere o seguinte problema:
  - 1% das mulheres na idade de 40 anos que fazem exames de rotina têm câncer de mama.
  - 80% das mulheres com câncer de mama terão mamogramas positivos para câncer

Regra de Bayes

Exemplo

- Considere o seguinte problema:
  - 1% das mulheres na idade de 40 anos que fazem exames de rotina têm câncer de mama.
  - 80% das mulheres com câncer de mama terão mamogramas positivos para câncer
  - 9,6% das mulheres sem câncer de mama também terão mamogramas positivos

Regra de Bayes

Exemplo

- Considere o seguinte problema:
  - 1% das mulheres na idade de 40 anos que fazem exames de rotina têm câncer de mama.
  - 80% das mulheres com câncer de mama terão mamogramas positivos para câncer
  - 9,6% das mulheres sem câncer de mama também terão mamogramas positivos
  - Uma mulher de 40 anos teve um mamograma positivo em um exame de rotina. Qual a probabilidade dela efetivamente ter câncer de mama?

Regra de Bayes

Exemplo

- $P(\text{câncer}|\text{positivo}) = ?$

Regra de Bayes

Exemplo

- P(câncer|positivo) = ?

$$P(câncer|positivo) = \frac{P(positivo|câncer)P(câncer)}{P(positivo)}$$

Regra de Bayes

Exemplo

- $\mathrm{P(câncer|positivo)} = ?$

$$P(câncer|positivo) = \frac{P(positivo|câncer)P(câncer)}{P(positivo)}$$

$$P(positivo|câncer) = \frac{80}{100} = 0,8$$

Regra de Bayes

Exemplo

- P(câncer|positivo) = ?

$$P(câncer|positivo) = \frac{P(positivo|câncer)P(câncer)}{P(positivo)}$$

$$P(positivo|câncer) = \frac{80}{100} = 0,8$$

$$P(câncer) = \frac{1}{100} = 0,01$$

Regra de Bayes

Exemplo

- P(câncer|positivo) = ?

$$P(câncer|positivo) = \frac{P(positivo|câncer)P(câncer)}{P(positivo)}$$

$$P(positivo|câncer) = \frac{80}{100} = 0,8$$

$$P(câncer) = \frac{1}{100} = 0,01$$

$$P(positivo) = ?$$

Regra de Bayes

Exemplo

- Lembre que a probabilidade de um efeito é obtida somando-se as probabilidades de sua ocorrência em conjunto com todas suas causas

Regra de Bayes

Exemplo

- Lembre que a probabilidade de um efeito é obtida somando-se as probabilidades de sua ocorrência em conjunto com todas suas causas
  - Efeito: mamograma positivo

Regra de Bayes

Exemplo

- Lembre que a probabilidade de um efeito é obtida somando-se as probabilidades de sua ocorrência em conjunto com todas suas causas
  - Efeito: mamograma positivo
  - Causa: ter (ou não) câncer

Regra de Bayes

Exemplo

- Lembre que a probabilidade de um efeito é obtida somando-se as probabilidades de sua ocorrência em conjunto com todas suas causas
  - Efeito: mamograma positivo
  - Causa: ter (ou não) câncer

$$P(positivo) = P(positivo \wedge câncer) + P(positivo \wedge \neg câncer)]$$

Regra de Bayes

Exemplo

- Lembre que a probabilidade de um efeito é obtida somando-se as probabilidades de sua ocorrência em conjunto com todas suas causas
  - Efeito: mamograma positivo
  - Causa: ter (ou não) câncer

$$P(positivo) = P(positivo \land câncer) + P(positivo \land \neg câncer)]$$

$$P(positivo \land câncer) = P(positivo|câncer)P(câncer)$$

Regra de Bayes

Exemplo

- Lembre que a probabilidade de um efeito é obtida somando-se as probabilidades de sua ocorrência em conjunto com todas suas causas
  - Efeito: mamograma positivo
  - Causa: ter (ou não) câncer

$$P(positivo) = P(positivo \land câncer) + P(positivo \land \neg câncer)]$$

$$P(positivo \land câncer) = P(positivo|câncer)P(câncer)$$

$$P(positivo \land câncer) = 0,8 \times 0,01 = 0,008$$

Regra de Bayes

Exemplo

- Lembre que a probabilidade de um efeito é obtida somando-se as probabilidades de sua ocorrência em conjunto com todas suas causas
  - Efeito: mamograma positivo
  - Causa: ter (ou não) câncer

$$P(positivo) = P(positivo \land câncer) + P(positivo \land \neg câncer)]$$

$$P(positivo \land câncer) = P(positivo|câncer)P(câncer)$$

$$P(positivo \land câncer) = 0,8 \times 0,01 = 0,008$$

$$P(positivo \land \neg câncer) = P(positivo|\neg câncer)(1 - P(câncer))$$

Regra de Bayes

Exemplo

- Lembre que a probabilidade de um efeito é obtida somando-se as probabilidades de sua ocorrência em conjunto com todas suas causas
  - Efeito: mamograma positivo
  - Causa: ter (ou não) câncer

$$P(positivo) = P(positivo \wedge câncer) + P(positivo \wedge \neg câncer)]$$

$$P(positivo \wedge câncer) = P(positivo|câncer)P(câncer)$$

$$P(positivo \wedge câncer) = 0,8 \times 0,01 = 0,008$$

$$P(positivo \wedge \neg câncer) = P(positivo|\neg câncer)(1 - P(câncer))$$

$$P(positivo \wedge \neg câncer) = \frac{9,6}{100} \times (1 - 0,01) = 0,09504$$

Regra de Bayes

Exemplo

- $P(\text{câncer}|\text{positivo}) = ?$

Regra de Bayes

Exemplo

- P(câncer|positivo) = ?

$$P(câncer|positivo) = \frac{P(positivo|câncer)P(câncer)}{P(positivo)}$$

Regra de Bayes

Exemplo

- P(câncer|positivo) = ?

$$P(câncer|positivo) = \frac{P(positivo|câncer)P(câncer)}{P(positivo)}$$

$$P(positivo|câncer) = \frac{80}{100} = 0,8$$

Regra de Bayes

Exemplo

- P(câncer|positivo) = ?

$$P(câncer|positivo) = \frac{P(positivo|câncer)P(câncer)}{P(positivo)}$$

$$P(positivo|câncer) = \frac{80}{100} = 0,8$$

$$P(câncer) = \frac{1}{100} = 0,01$$

Regra de Bayes

Exemplo

- P(câncer|positivo) = ?

$$P(câncer|positivo) = \frac{P(positivo|câncer)P(câncer)}{P(positivo)}$$

$$P(positivo|câncer) = \frac{80}{100} = 0,8$$

$$P(câncer) = \frac{1}{100} = 0,01$$

$$P(positivo) = 0,008 + 0,09504 = 0,10304$$

Regra de Bayes

Exemplo

- P(câncer|positivo) = ?

$$P(câncer|positivo) = \frac{P(positivo|câncer)P(câncer)}{P(positivo)}$$

$$P(positivo|câncer) = \frac{80}{100} = 0,8$$

$$P(câncer) = \frac{1}{100} = 0,01$$

$$P(positivo) = 0,008 + 0,09504 = 0,10304$$

$$P(câncer|positivo) = \frac{0,8 \times 0,01}{0,10304} \approx 0,0776 = 7,76\%$$

Redes Bayesianas

Redes Bayesianas

- Trata-se de uma estrutura de dados

Redes Bayesianas

- Trata-se de uma estrutura de dados
  - Representa a distribuição conjunta de probabilidade para um conjunto de variáveis

Redes Bayesianas

- Trata-se de uma estrutura de dados
  - Representa a distribuição conjunta de probabilidade para um conjunto de variáveis
  - Representa dependências entre variáveis

Redes Bayesianas

- Trata-se de uma estrutura de dados
  - Representa a distribuição conjunta de probabilidade para um conjunto de variáveis
  - Representa dependências entre variáveis
- Grafo dirigido acíclico

Redes Bayesianas

- Trata-se de uma estrutura de dados
  - Representa a distribuição conjunta de probabilidade para um conjunto de variáveis
  - Representa dependências entre variáveis
- Grafo dirigido acíclico
  - Nós: variáveis aleatórias (discretas ou contínuas)

Redes Bayesianas

- Trata-se de uma estrutura de dados
  - Representa a distribuição conjunta de probabilidade para um conjunto de variáveis
  - Representa dependências entre variáveis
- Grafo dirigido acíclico
  - Nós: variáveis aleatórias (discretas ou contínuas)
  - Cada nó tem uma tabela de distribuição de probabilidade que quantifica a influência dos pais nesse nó

Redes Bayesianas

- Trata-se de uma estrutura de dados
  - Representa a distribuição conjunta de probabilidade para um conjunto de variáveis
  - Representa dependências entre variáveis
- Grafo dirigido acíclico
  - Nós: variáveis aleatórias (discretas ou contínuas)
  - Cada nó tem uma tabela de distribuição de probabilidade que quantifica a influência dos pais nesse nó
    - $P(X_i|pais(X_i))$

Redes Bayesianas

- Trata-se de uma estrutura de dados
  - Representa a distribuição conjunta de probabilidade para um conjunto de variáveis
  - Representa dependências entre variáveis
- Grafo dirigido acíclico
  - Nós: variáveis aleatórias (discretas ou contínuas)
  - Cada nó tem uma tabela de distribuição de probabilidade que quantifica a influência dos pais nesse nó
    - $P(X_i|pais(X_i))$
  - Arestas: unem variáveis interdependentes

Redes Bayesianas

Topologia

Redes Bayesianas

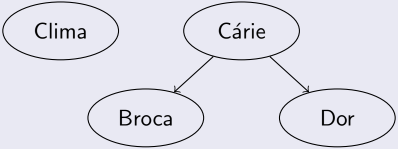
Topologia

- A topologia da rede especifica as relações de independência condicional existentes:

Redes Bayesianas

Topologia

- A topologia da rede especifica as relações de independência condicional existentes:

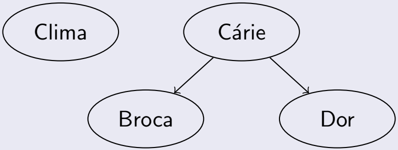


Fonte: Adaptado de AIMA. Russell & Norvig.

Redes Bayesianas

Topologia

- A topologia da rede especifica as relações de independência condicional existentes:



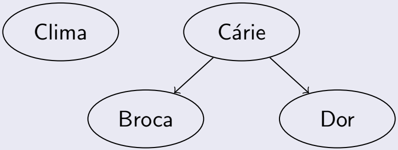
Fonte: Adaptado de AIMA. Russell & Norvig.

- *Clima* é independente das demais variáveis

Redes Bayesianas

Topologia

- A topologia da rede especifica as relações de independência condicional existentes:



Fonte: Adaptado de AIMA. Russell & Norvig.

- *Clima* é independente das demais variáveis
- *Broca* e *Dor* são condicionalmente independentes, dada *Cárie*

Redes Bayesianas

Exemplo

- Você está no trabalho...

Exemplo

- Você está no trabalho...
  - Seu vizinho, João, liga dizendo que o alarme da sua casa disparou

Redes Bayesianas

Exemplo

- Você está no trabalho...
  - Seu vizinho, João, liga dizendo que o alarme da sua casa disparou
  - Contudo, a vizinha, Maria, não ligou

Redes Bayesianas

Exemplo

- Você está no trabalho...
  - Seu vizinho, João, liga dizendo que o alarme da sua casa disparou
  - Contudo, a vizinha, Maria, não ligou
  - Algumas vezes o alarme dispara por conta de pequenos tremores

Redes Bayesianas

Exemplo

- Você está no trabalho...
  - Seu vizinho, João, liga dizendo que o alarme da sua casa disparou
  - Contudo, a vizinha, Maria, não ligou
  - Algumas vezes o alarme dispara por conta de pequenos tremores
  - Há um ladrão?

Redes Bayesianas

Exemplo

- Você está no trabalho...
  - Seu vizinho, João, liga dizendo que o alarme da sua casa disparou
  - Contudo, a vizinha, Maria, não ligou
  - Algumas vezes o alarme dispara por conta de pequenos tremores
  - Há um ladrão?
- Variáveis:

Redes Bayesianas

Exemplo

- Você está no trabalho...
  - Seu vizinho, João, liga dizendo que o alarme da sua casa disparou
  - Contudo, a vizinha, Maria, não ligou
  - Algumas vezes o alarme dispara por conta de pequenos tremores
  - Há um ladrão?
- Variáveis:
  - *Ladrão*, *Terremoto*, *Alarme*, *JoãoLiga*, *MariaLiga*

Exemplo – Relações causais (i.e., regras)

Redes Bayesianas

Exemplo – Relações causais (i.e., regras)

- Um ladrão pode disparar o alarme

Exemplo – Relações causais (i.e., regras)

- Um ladrão pode disparar o alarme
- Um terremoto pode disparar o alarme

Redes Bayesianas

Exemplo – Relações causais (i.e., regras)

- Um ladrão pode disparar o alarme
- Um terremoto pode disparar o alarme
- O alarme pode fazer com que Maria ligue

Redes Bayesianas

Exemplo – Relações causais (i.e., regras)

- Um ladrão pode disparar o alarme
- Um terremoto pode disparar o alarme
- O alarme pode fazer com que Maria ligue
  - Embora ela nem sempre o escute

Redes Bayesianas

Exemplo – Relações causais (i.e., regras)

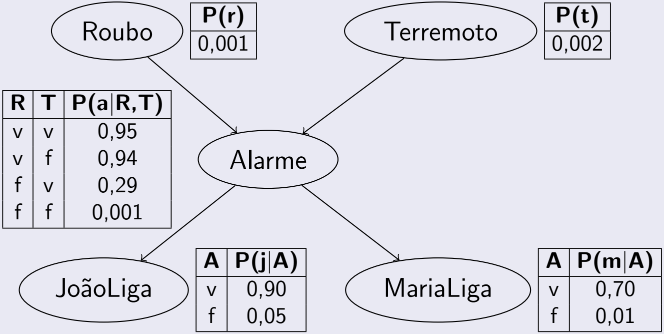
- Um ladrão pode disparar o alarme
- Um terremoto pode disparar o alarme
- O alarme pode fazer com que Maria ligue
  - Embora ela nem sempre o escute
- O alarme pode fazer com que João ligue

Redes Bayesianas

Exemplo – Relações causais (i.e., regras)

- Um ladrão pode disparar o alarme
- Um terremoto pode disparar o alarme
- O alarme pode fazer com que Maria ligue
  - Embora ela nem sempre o escute
- O alarme pode fazer com que João ligue
  - Embora algumas vezes ele confunda o telefone com o alarme

Exemplo



P(r)
0,001

P(t)
0,002

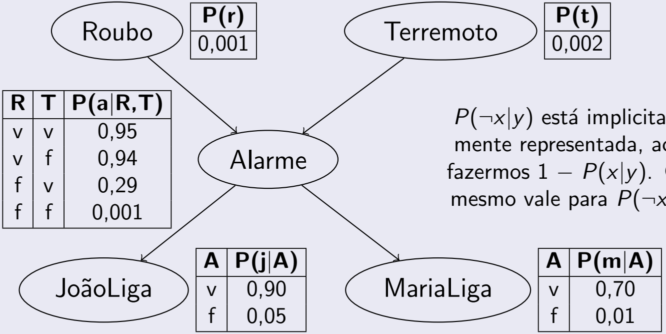
R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

A	P(j|A)
v	0,90
f	0,05

A	P(m|A)
v	0,70
f	0,01

Fonte: Adaptado de AIMA. Russell & Norvig.

Exemplo



P(r)
0,001

P(t)
0,002

R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

A	P(j|A)
v	0,90
f	0,05

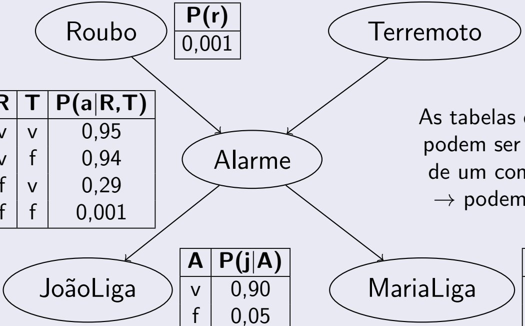
A	P(m|A)
v	0,70
f	0,01

$P(\neg x|y)$ está implicitamente representada, ao fazermos $1 - P(x|y)$. O mesmo vale para $P(\neg x)$

Fonte: Adaptado de AIMA. Russell & Norvig.

Redes Bayesianas

Exemplo



P(r)
0,001

P(t)
0,002

R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

A	P(j|A)
v	0,90
f	0,05

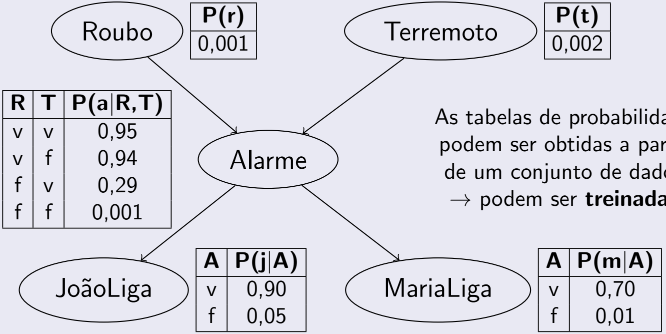
A	P(m|A)
v	0,70
f	0,01

As tabelas de probabilidade podem ser obtidas a partir de um conjunto de dados $\rightarrow$ podem ser **treinadas**

Fonte: Adaptado de AIMA. Russell & Norvig.

Redes Bayesianas

Exemplo



P(r)
0,001

P(t)
0,002

R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

A	P(j|A)
v	0,90
f	0,05

A	P(m|A)
v	0,70
f	0,01

As tabelas de probabilidade podem ser obtidas a partir de um conjunto de dados $\rightarrow$ podem ser **treinadas**

Fonte: Adaptado de AIMA. Russell & Norvig.

Probabilidades resumem um conjunto potencialmente infinito de causas, ao definirem valores para algo não acontecer (como o alarme não soar), sem se preocupar em definir suas causas.

Redes Bayesianas

Semântica Global

- Define a distribuição conjunta total como sendo o produto das distribuições condicionais locais

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pais(X_i))$$

  - $pais(X_i)$ – valores das variáveis pais de X_i na rede

Redes Bayesianas

Semântica Global

- Define a distribuição conjunta total como sendo o produto das distribuições condicionais locais

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pais(X_i))$$

  - $pais(X_i)$ – valores das variáveis pais de X_i na rede

- Exemplo:
  - João e Maria ligam porque o alarme soou, mas não há nem terremoto, nem ladrão

Semântica Global

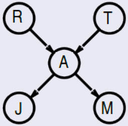
- Define a distribuição conjunta total como sendo o produto das distribuições condicionais locais

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pais(X_i))$$

  - $pais(X_i)$ – valores das variáveis pais de X_i na rede

- Exemplo:
  - João e Maria ligam porque o alarme soou, mas não há nem terremoto, nem ladrão

$$P(j \wedge m \wedge a \wedge \neg r \wedge \neg t)$$
$$= P(j|a)P(m|a)P(a|\neg r \wedge \neg t)P(\neg r)P(\neg t)$$
$$= 0{,}90 \times 0{,}70 \times 0{,}001 \times 0{,}999 \times 0{,}998 \approx 0{,}00063$$



Fonte: Slides de AIMA. Russell & Norvig.

Redes Bayesianas

Construção de Redes Bayesianas

Redes Bayesianas

Construção de Redes Bayesianas

- Considere a regra da cadeia:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)$$

Verdadeira para todo conjunto de variáveis

Redes Bayesianas

Construção de Redes Bayesianas

- Considere a regra da cadeia:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)$$

Verdadeira para todo conjunto de variáveis

- Compare-a com a interpretação semântica de uma rede bayesiana:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pais(x_i))$$

Redes Bayesianas

Construção de Redes Bayesianas

- Considere a regra da cadeia:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)$$

Verdadeira para todo conjunto de variáveis

- Compare-a com a interpretação semântica de uma rede bayesiana:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pais(x_i))$$

- $P(x_i | x_{i-1}, \ldots, x_1) = P(x_i | pais(x_i))$, se $pais(x_i) \subseteq \{x_{i-1}, \ldots, x_1\}$

Construção de Redes Bayesianas

- Considere a regra da cadeia:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)$$ Verdadeira para todo conjunto de variáveis

- Compare-a com a interpretação semântica de uma rede bayesiana:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pais(x_i))$$

- $P(x_i | x_{i-1}, \ldots, x_1) = P(x_i | pais(x_i))$, se $pais(x_i) \subseteq \{x_{i-1}, \ldots, x_1\}$
  - Condição satisfeita se rotularmos os nós em uma ordem consistente com a ordem parcial implícita no grafo

Redes Bayesianas

Construção de Redes Bayesianas

- $P(X_i|X_{i-1},\ldots,X_1) = P(X_i|pais(X_i))$

Redes Bayesianas

Construção de Redes Bayesianas

- $P(X_i|X_{i-1},\ldots,X_1) = P(X_i|pais(X_i))$
  - Redes bayesianas são representações corretas do domínio somente se cada nó for condicionalmente independente de seus predecessores na ordenação, dados seus pais

Redes Bayesianas

Construção de Redes Bayesianas

- $P(X_i|X_{i-1},\ldots,X_1) = P(X_i|pais(X_i))$
  - Redes bayesianas são representações corretas do domínio somente se cada nó for condicionalmente independente de seus predecessores na ordenação, dados seus pais
- Método:

Redes Bayesianas

Construção de Redes Bayesianas

- $P(X_i|X_{i-1}, \ldots, X_1) = P(X_i|pais(X_i))$
  - Redes bayesianas são representações corretas do domínio somente se cada nó for condicionalmente independente de seus predecessores na ordenação, dados seus pais
- Método:
  - Escolha uma ordem para as variáveis $X_n, \ldots, X_1$

Construção de Redes Bayesianas

- $P(X_i|X_{i-1},\ldots,X_1) = P(X_i|pais(X_i))$
  - Redes bayesianas são representações corretas do domínio somente se cada nó for condicionalmente independente de seus predecessores na ordenação, dados seus pais
- Método:
  - Escolha uma ordem para as variáveis $X_n,\ldots,X_1$
  - Para $i = 1$ até n:

Redes Bayesianas

Construção de Redes Bayesianas

- $P(X_i|X_{i-1}, \ldots, X_1) = P(X_i|pais(X_i))$
  - Redes bayesianas são representações corretas do domínio somente se cada nó for condicionalmente independente de seus predecessores na ordenação, dados seus pais
- Método:
  - Escolha uma ordem para as variáveis $X_n, \ldots, X_1$
  - Para $i = 1$ até n:
    - Adicione X_i à rede

Construção de Redes Bayesianas

- $P(X_i|X_{i-1},\ldots,X_1) = P(X_i|pais(X_i))$
  - Redes bayesianas são representações corretas do domínio somente se cada nó for condicionalmente independente de seus predecessores na ordenação, dados seus pais
- Método:
  - Escolha uma ordem para as variáveis $X_n,\ldots,X_1$
  - Para $i = 1$ até n:
    - Adicione X_i à rede
    - Selecione pais de $X_1,\ldots,X_{i-1}$ tais que $P(X_i|pais(X_i)) = P(X_i|X_{i-1},\ldots,X_1)$

Construção de Redes Bayesianas

- Exemplo de teste para os pais:

Construção de Redes Bayesianas

- Exemplo de teste para os pais:
  - Ordem: R, T, A, J, M

Redes Bayesianas

Construção de Redes Bayesianas

- Exemplo de teste para os pais:
  - Ordem: R, T, A, J, M
  - Maria ligar (M) é influenciada pelo fato de ter um ladrão (R) ou terremoto (T)

Redes Bayesianas

Construção de Redes Bayesianas

- Exemplo de teste para os pais:
  - Ordem: R, T, A, J, M
  - Maria ligar (M) é influenciada pelo fato de ter um ladrão (R) ou terremoto (T)
    - Não diretamente, apenas por meio do alarme (A)

Construção de Redes Bayesianas

- Exemplo de teste para os pais:
  - Ordem: R, T, A, J, M
  - Maria ligar (M) é influenciada pelo fato de ter um ladrão (R) ou terremoto (T)
    - Não diretamente, apenas por meio do alarme (A)
  - Dado o estado do alarme, João ligar (J) não influencia o fato de Maria ligar ou não

Construção de Redes Bayesianas

- Exemplo de teste para os pais:
  - Ordem: R, T, A, J, M
  - Maria ligar (M) é influenciada pelo fato de ter um ladrão (R) ou terremoto (T)
    - Não diretamente, apenas por meio do alarme (A)
  - Dado o estado do alarme, João ligar (J) não influencia o fato de Maria ligar ou não
  - Com base nessas crenças, podemos dizer que

$$P(M|J, A, T, R) = P(M|A)$$

Construção de Redes Bayesianas

- Exemplo de teste para os pais:
  - Ordem: R, T, A, J, M
  - Maria ligar (M) é influenciada pelo fato de ter um ladrão (R) ou terremoto (T)
    - Não diretamente, apenas por meio do alarme (A)
  - Dado o estado do alarme, João ligar (J) não influencia o fato de Maria ligar ou não
  - Com base nessas crenças, podemos dizer que

$$P(M|J, A, T, R) = P(M|A)$$

    - E nossa escolha foi correta, para mariaLigar

Redes Bayesianas

Construção de Redes Bayesianas

- Qual então é a ordem correta de nós?

Construção de Redes Bayesianas

- Qual então é a ordem correta de nós?
  - Uma dica é começar com quem influencia outros nós, mas não sofre influência de ninguém

Redes Bayesianas

Construção de Redes Bayesianas

- Qual então é a ordem correta de nós?
  - Uma dica é começar com quem influencia outros nós, mas não sofre influência de ninguém
  - A ordem correta de inclusão de nós é, então:

Redes Bayesianas

Construção de Redes Bayesianas

- Qual então é a ordem correta de nós?
  - Uma dica é começar com quem influencia outros nós, mas não sofre influência de ninguém
  - A ordem correta de inclusão de nós é, então:
    - Primeiro adicione as "causas principais"

Construção de Redes Bayesianas

- Qual então é a ordem correta de nós?
  - Uma dica é começar com quem influencia outros nós, mas não sofre influência de ninguém
  - A ordem correta de inclusão de nós é, então:
    - Primeiro adicione as “causas principais”
    - Em seguida, as variáveis que elas influenciam

Redes Bayesianas

Construção de Redes Bayesianas

- Qual então é a ordem correta de nós?
  - Uma dica é começar com quem influencia outros nós, mas não sofre influência de ninguém
  - A ordem correta de inclusão de nós é, então:
    - Primeiro adicione as “causas principais”
    - Em seguida, as variáveis que elas influenciam
    - Repita, até atingir as folhas – variáveis que não influenciam ninguém

Redes Bayesianas

Construção de Redes Bayesianas

- Qual então é a ordem correta de nós?
  - Uma dica é começar com quem influencia outros nós, mas não sofre influência de ninguém
  - A ordem correta de inclusão de nós é, então:
    - Primeiro adicione as "causas principais"
    - Em seguida, as variáveis que elas influenciam
    - Repita, até atingir as folhas – variáveis que não influenciam ninguém
- E se escolhermos a ordem errada?

Redes Bayesianas

Construção de Redes Bayesianas

- Qual então é a ordem correta de nós?
  - Uma dica é começar com quem influencia outros nós, mas não sofre influência de ninguém
  - A ordem correta de inclusão de nós é, então:
    - Primeiro adicione as "causas principais"
    - Em seguida, as variáveis que elas influenciam
    - Repita, até atingir as folhas – variáveis que não influenciam ninguém
- E se escolhermos a ordem errada?
  - Suponha que escolhemos a ordem: M,J,A,R,T

Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T

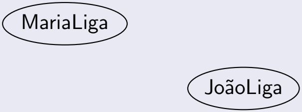
Consequências da Ordem Errada

Devemos incluir na ordem definida: M,J,A,R,T

Redes Bayesianas

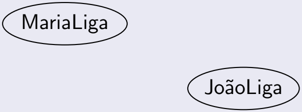
Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T



Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T

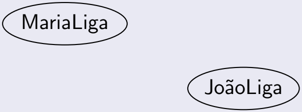


$P(J|M) = P(J)$?

Redes Bayesianas

Consequências da Ordem Errada

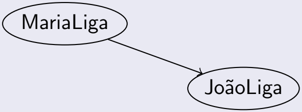
Devemos incluir
na ordem definida:
M,J,A,R,T



$P(J|M) = P(J)$? Não. Se Maria ligar, provavelmente o alarme disparou, o que aumenta a probabilidade de João ligar

Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T

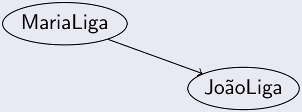


$P(J|M) = P(J)$? Não. Se Maria ligar, provavelmente o alarme disparou, o que aumenta a probabilidade de João ligar

Redes Bayesianas

Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T

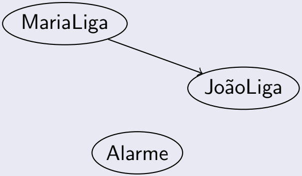


Note que os candidatos a pai de X_i têm que estar entre os X_{i-1} anteriores

$P(J|M) = P(J)$? Não. Se Maria ligar, provavelmente o alarme disparou, o que aumenta a probabilidade de João ligar

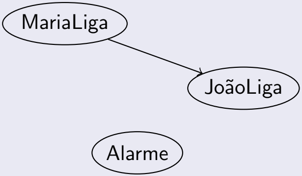
Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T



Consequências da Ordem Errada

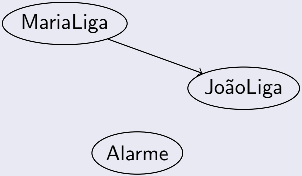
Devemos incluir
na ordem definida:
M,J,A,R,T



$$P(A|J,M) = P(A|J)?$$
$$P(A|J,M) = P(A|M)?$$
$$P(A|J,M) = P(A)?$$

Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T

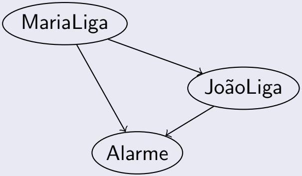


$$\left.\begin{array}{l} P(A|J,M) = P(A|J)? \\ P(A|J,M) = P(A|M)? \\ P(A|J,M) = P(A)? \end{array}\right\}$$

Não. Se ambos ligarem, as chances do alarme ter disparado são maiores do que se apenas um ligar, ou se nenhum ligar

Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T

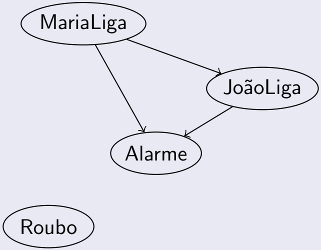


$$\left.\begin{array}{l} P(A|J,M) = P(A|J)? \\ P(A|J,M) = P(A|M)? \\ P(A|J,M) = P(A)? \end{array}\right\}$$

Não. Se ambos ligarem, as chances do alarme ter disparado são maiores do que se apenas um ligar, ou se nenhum ligar

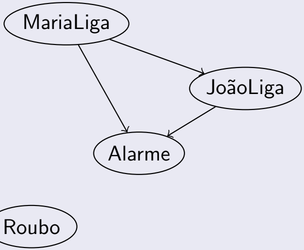
Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T



Consequências da Ordem Errada

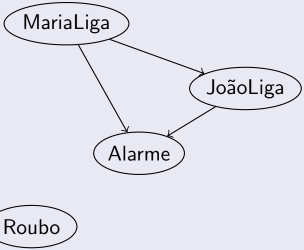
Devemos incluir
na ordem definida:
M,J,A,R,T



$$P(R|M, J, A) = P(R)?$$

Consequências da Ordem Errada

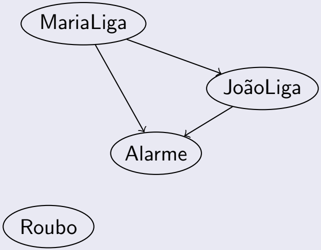
Devemos incluir
na ordem definida:
M,J,A,R,T



$P(R|M, J, A) = P(R)$? Não. Não é independente das demais

Consequências da Ordem Errada

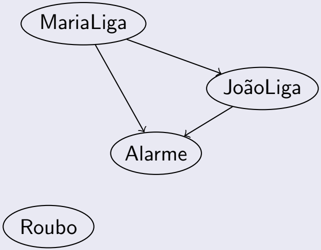
Devemos incluir
na ordem definida:
M,J,A,R,T



$P(R|M,J,A) = P(R)$? Não. Não é independente das demais
$P(R|M,J,A) = P(R|A)$?

Consequências da Ordem Errada

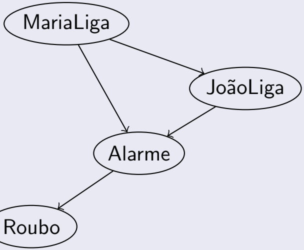
Devemos incluir
na ordem definida:
M,J,A,R,T



$P(R|M, J, A) = P(R)$? Não. Não é independente das demais
$P(R|M, J, A) = P(R|A)$? Sim. Se sabemos o estado do alarme, não importa se João ou Maria ligam

Consequências da Ordem Errada

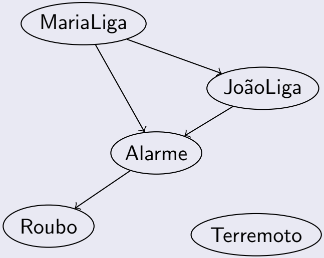
Devemos incluir
na ordem definida:
M,J,A,R,T



$P(R|M, J, A) = P(R)$? Não. Não é independente das demais
$P(R|M, J, A) = P(R|A)$? Sim. Se sabemos o estado do alarme, não importa se João ou Maria ligam

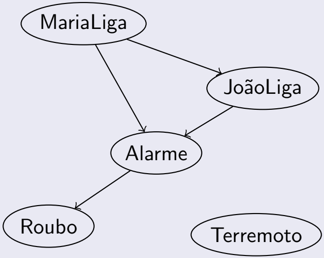
Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T



Consequências da Ordem Errada

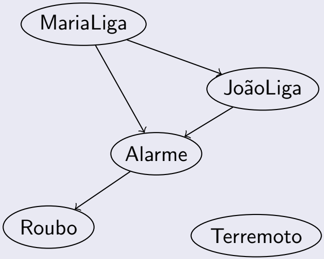
Devemos incluir
na ordem definida:
M,J,A,R,T



$$P(T|R, A, J, M) = P(T)?$$

Consequências da Ordem Errada

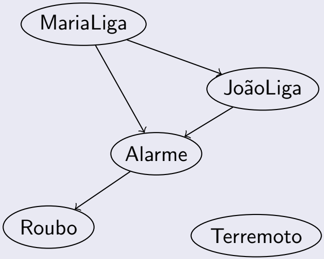
Devemos incluir
na ordem definida:
M,J,A,R,T



$P(T|R,A,J,M) = P(T)$? Não. Não é independente das demais

Consequências da Ordem Errada

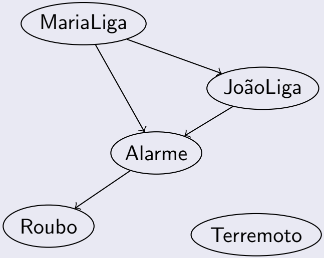
Devemos incluir
na ordem definida:
M,J,A,R,T



$P(T|R, A, J, M) = P(T)$? Não. Não é independente das demais
$P(T|R, A, J, M) = P(T|A)$?

Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T

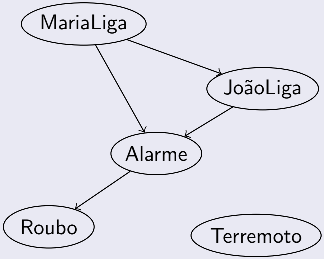


$P(T|R, A, J, M) = P(T)$? Não. Não é independente das demais
$P(T|R, A, J, M) = P(T|A)$? Não

Redes Bayesianas

Consequências da Ordem Errada

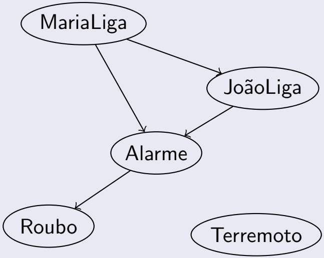
Devemos incluir
na ordem definida:
M,J,A,R,T



$P(T|R, A, J, M) = P(T)$? Não. Não é independente das demais
$P(T|R, A, J, M) = P(T|A)$? Não
$P(T|R, A, J, M) = P(T|R, A)$?

Consequências da Ordem Errada

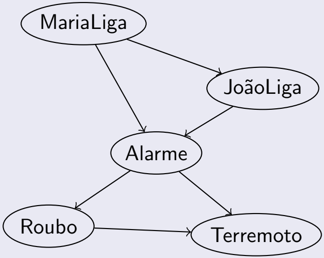
Devemos incluir
na ordem definida:
M,J,A,R,T



$P(T|R,A,J,M) = P(T)$? Não. Não é independente das demais
$P(T|R,A,J,M) = P(T|A)$? Não
$P(T|R,A,J,M) = P(T|R,A)$? Sim. Se o alarme estiver soando, aumenta a probabilidade de terremoto. Contudo, se soubermos que houve um roubo, isso explica o alarme, e a probabilidade do terremoto cai

Consequências da Ordem Errada

Devemos incluir
na ordem definida:
M,J,A,R,T

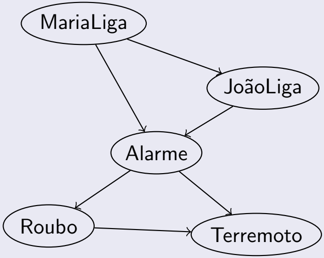


Fonte: Adaptado de AIMA. Russell & Norvig.

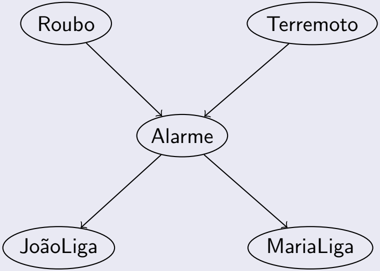
$P(T|R,A,J,M) = P(T)$? Não. Não é independente das demais
$P(T|R,A,J,M) = P(T|A)$? Não
$P(T|R,A,J,M) = P(T|R,A)$? Sim. Se o alarme estiver soando, aumenta a probabilidade de terremoto. Contudo, se soubermos que houve um roubo, isso explica o alarme, e a probabilidade do terremoto cai

Consequências da Ordem Errada

Ordem: M,J,A,R,T



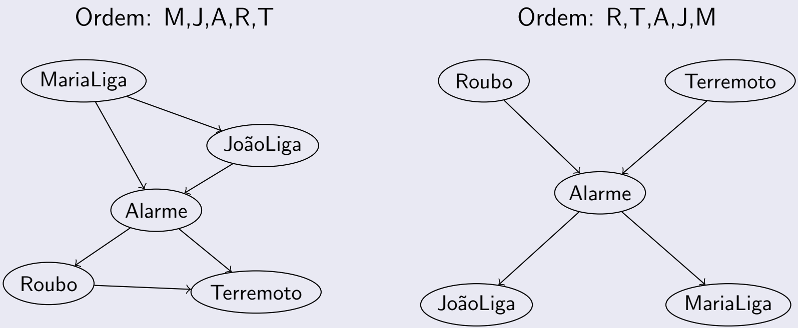
Ordem: R,T,A,J,M



Fonte: Adaptado de AIMA. Russell & Norvig.

Redes Bayesianas

Consequências da Ordem Errada



Fonte: Adaptado de AIMA. Russell & Norvig.

A nova rede tem duas arestas a mais que a antiga

Construção Automática

- Vimos o caso em que a estrutura da rede é dada e temos os valores para todas as variáveis nos exemplos de treino

Construção Automática

- Vimos o caso em que a estrutura da rede é dada e temos os valores para todas as variáveis nos exemplos de treino
  - Estimamos as tabelas de probabilidade condicional a partir dos dados

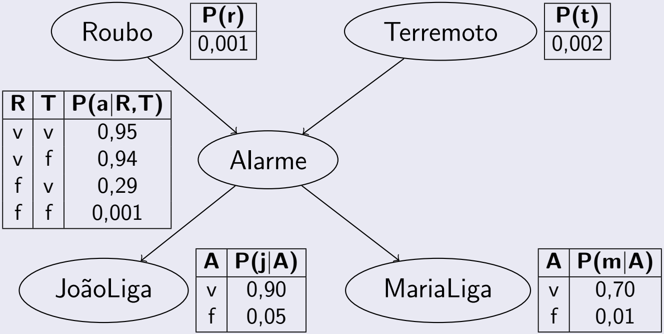
Redes Bayesianas

Construção Automática

- Vimos o caso em que a estrutura da rede é dada e temos os valores para todas as variáveis nos exemplos de treino
  - Estimamos as tabelas de probabilidade condicional a partir dos dados
- Podemos aprender também a estrutura a partir dos dados. Ex:
  - K2: Cooper and Herskovits (1992)
  - TAN: Friedman et al. (1997)
  - Simulated Annealing: Bouckaert (1995)

Inferência em Redes Bayesianas

- Considere a rede bayesiana:



P(r)
0,001

P(t)
0,002

R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

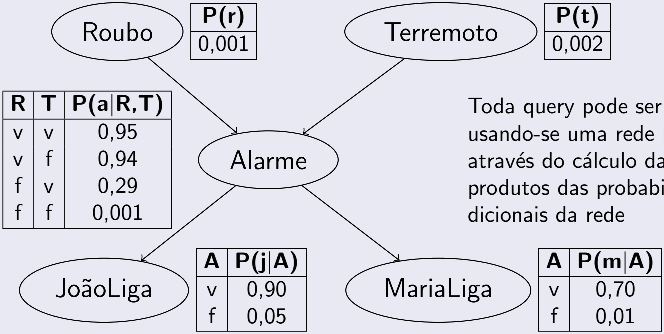
A	P(j|A)
v	0,90
f	0,05

A	P(m|A)
v	0,70
f	0,01

Fonte: Adaptado de AIMA. Russell & Norvig.

Inferência em Redes Bayesianas

- Considere a rede bayesiana:



P(r)
0,001

P(t)
0,002

R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

A	P(j|A)
v	0,90
f	0,05

A	P(m|A)
v	0,70
f	0,01

Toda query pode ser respondida usando-se uma rede bayesiana através do cálculo das somas dos produtos das probabilidades condicionais da rede

Fonte: Adaptado de AIMA. Russell & Norvig.

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

- Query: $P(r|j, m) = ?$

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

- Query: $P(r|j, m) =?$

$$P(r|j, m) = \frac{P(r, j, m)}{P(j, m)}$$

Inferência em Redes Bayesianas – Exemplo

- Query: $P(r|j,m) = ?$

$$P(r|j,m) = \frac{P(r,j,m)}{P(j,m)}$$

$$P(r,j,m) = \sum_{T=\{v,f\}} \sum_{A=\{v,f\}} P(r,j,m,T,A)$$

Inferência em Redes Bayesianas – Exemplo

- Query: $P(r|j,m) = ?$

Como não definimos valores para T e A, temos que fazer todas as combinações de valores possíveis

$$P(r|j,m) = \frac{P(r,j,m)}{P(j,m)}$$

$$P(r,j,m) = \sum_{T=\{v,f\}} \sum_{A=\{v,f\}} P(r,j,m,T,A)$$

Inferência em Redes Bayesianas – Exemplo

- Query: $P(r|j,m) = ?$

Como não definimos valores para T e A, temos que fazer todas as combinações de valores possíveis

$$P(r|j,m) = \frac{P(r,j,m)}{P(j,m)}$$

$$P(r,j,m) = \sum_{T=\{v,f\}} \sum_{A=\{v,f\}} P(r,j,m,T,A)$$

$$= \sum_{T=\{v,f\}} P(r,j,m,T,a) + P(r,j,m,T,\neg a)$$

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

- Query: $P(r|j,m) = ?$

$$
\begin{aligned}
P(r|j,m) &= \frac{P(r,j,m)}{P(j,m)} \\
P(r,j,m) &= \sum_{T=\{v,f\}} \sum_{A=\{v,f\}} P(r,j,m,T,A) \\
&= \sum_{T=\{v,f\}} P(r,j,m,T,a) + P(r,j,m,T,\neg a) \\
&= P(r,j,m,t,a) + P(r,j,m,t,\neg a) + \\
&\quad P(r,j,m,\neg t,a) + P(r,j,m,\neg t,\neg a)
\end{aligned}
$$

Inferência em Redes Bayesianas – Exemplo

- Query: $P(r|j,m) = ?$

A query foi reescrita como uma soma de probabilidades conjuntas

$$
\begin{aligned}
P(r|j,m) &= \frac{P(r,j,m)}{P(j,m)} \\
P(r,j,m) &= \sum_{T=\{v,f\}} \sum_{A=\{v,f\}} P(r,j,m,T,A) \\
&= \sum_{T=\{v,f\}} P(r,j,m,T,a) + P(r,j,m,T,\neg a) \\
&= P(r,j,m,t,a) + P(r,j,m,t,\neg a) + \\
&\quad P(r,j,m,\neg t,a) + P(r,j,m,\neg t,\neg a)
\end{aligned}
$$

Inferência em Redes Bayesianas – Exemplo

$$P(r, j, m, t, a) = P(r)P(j, m, t, a|r)$$

Inferência em Redes Bayesianas – Exemplo

$$\begin{aligned} P(r, j, m, t, a) &= P(r)P(j, m, t, a|r) \\ &= P(r)P(t)P(j, m, a|r, t) \end{aligned}$$

Inferência em Redes Bayesianas – Exemplo

$$
\begin{aligned}
P(r, j, m, t, a) &= P(r)P(j, m, t, a|r) \\
&= P(r)P(t)P(j, m, a|r, t)
\end{aligned}
$$

A escolha de R e T deu-se por serem raízes na árvore

Inferência em Redes Bayesianas – Exemplo

$$
\begin{aligned}
P(r,j,m,t,a) &= P(r)P(j,m,t,a|r) \\
&= P(r)P(t)P(j,m,a|r,t) \\
&= P(r)P(t)P(a|r,t)P(j,m|r,t,a)
\end{aligned}
$$

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

$$
\begin{aligned}
P(r, j, m, t, a) &= P(r)P(j, m, t, a|r) \\
&= P(r)P(t)P(j, m, a|r, t) \\
&= P(r)P(t)P(a|r, t)P(j, m|r, t, a)
\end{aligned}
$$

A foi escolhido por depender unicamente das variáveis já isoladas (T e R)

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

$$
\begin{aligned}
P(r,j,m,t,a) &= P(r)P(j,m,t,a|r) \\
&= P(r)P(t)P(j,m,a|r,t) \\
&= P(r)P(t)P(a|r,t)P(j,m|r,t,a) \\
&= P(r)P(t)P(a|r,t)P(j,m|a)
\end{aligned}
$$

Inferência em Redes Bayesianas – Exemplo

$$\begin{aligned}
P(r, j, m, t, a) &= P(r)P(j, m, t, a|r) \\
&= P(r)P(t)P(j, m, a|r, t) \\
&= P(r)P(t)P(a|r, t)P(j, m|r, t, a) \\
&= P(r)P(t)P(a|r, t)P(j, m|a)
\end{aligned}$$

j não depende de r e t

Inferência em Redes Bayesianas – Exemplo

$$\begin{aligned}
P(r,j,m,t,a) &= P(r)P(j,m,t,a|r) \\
&= P(r)P(t)P(j,m,a|r,t) \\
&= P(r)P(t)P(a|r,t)P(j,m|r,t,a) \\
&= P(r)P(t)P(a|r,t)P(j,m|a) \\
&= P(r)P(t)P(a|r,t)P(j|a)P(m|a)
\end{aligned}$$

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

$$
\begin{aligned}
P(r, j, m, t, a) &= P(r)P(j, m, t, a|r) \\
&= P(r)P(t)P(j, m, a|r, t) \\
&= P(r)P(t)P(a|r, t)P(j, m|r, t, a) \\
&= P(r)P(t)P(a|r, t)P(j, m|a) \\
&= P(r)P(t)P(a|r, t)P(j|a)P(m|a)
\end{aligned}
$$

Note que se trata de cada nó da rede

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

$$
\begin{aligned}
P(r,j,m,t,a) &= P(r)P(j,m,t,a|r) \\
&= P(r)P(t)P(j,m,a|r,t) \\
&= P(r)P(t)P(a|r,t)P(j,m|r,t,a) \\
&= P(r)P(t)P(a|r,t)P(j,m|a) \\
&= P(r)P(t)P(a|r,t)P(j|a)P(m|a)
\end{aligned}
$$

Assim, derivando-se as demais de modo similar, teremos:

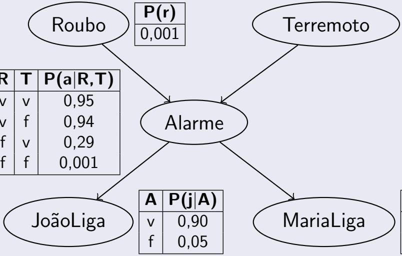
Inferência em Redes Bayesianas – Exemplo

$$\begin{aligned}
P(r, j, m, t, a) &= P(r)P(j, m, t, a|r) \\
&= P(r)P(t)P(j, m, a|r, t) \\
&= P(r)P(t)P(a|r, t)P(j, m|r, t, a) \\
&= P(r)P(t)P(a|r, t)P(j, m|a) \\
&= P(r)P(t)P(a|r, t)P(j|a)P(m|a)
\end{aligned}$$

Assim, derivando-se as demais de modo similar, teremos:

$$\begin{aligned}
P(r, j, m, t, \neg a) &= P(r)P(t)P(\neg a|r, t)P(j|\neg a)P(m|\neg a) \\
P(r, j, m, \neg t, a) &= P(r)P(\neg t)P(a|r, \neg t)P(j|a)P(m|a) \\
P(r, j, m, \neg t, \neg a) &= P(r)P(\neg t)P(\neg a|r, \neg t)P(j|\neg a)P(m|\neg a)
\end{aligned}$$

Inferência em Redes Bayesianas – Exemplo



P(r)
0,001

P(t)
0,002

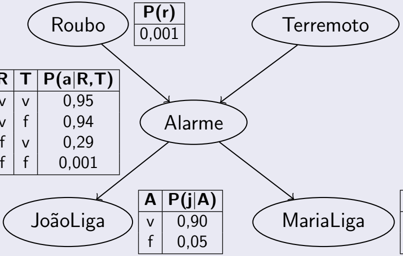
R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

A	P(j|A)
v	0,90
f	0,05

A	P(m|A)
v	0,70
f	0,01

Fonte: Adaptado de AIMA. Russell & Norvig.

Inferência em Redes Bayesianas – Exemplo



P(r)
0,001

P(t)
0,002

R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

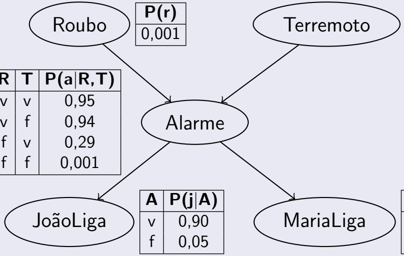
A	P(j|A)
v	0,90
f	0,05

A	P(m|A)
v	0,70
f	0,01

Fonte: Adaptado de AIMA. Russell & Norvig.

$$P(r, j, m, t, a) = P(r)P(t)P(a|r, t)P(j|a)P(m|a) \approx 1,197 \times 10^{-6}$$

Inferência em Redes Bayesianas – Exemplo



P(r)
0,001

P(t)
0,002

R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

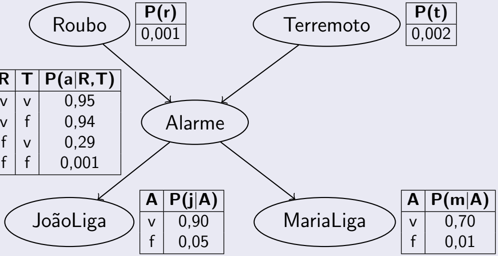
A	P(j|A)
v	0,90
f	0,05

A	P(m|A)
v	0,70
f	0,01

Fonte: Adaptado de AIMA. Russell & Norvig.

$$P(r, j, m, t, a) = P(r)P(t)P(a|r, t)P(j|a)P(m|a) \approx 1,197 \times 10^{-6}$$
$$P(r, j, m, t, \neg a) = P(r)P(t)P(\neg a|r, t)P(j|\neg a)P(m|\neg a) \approx 5 \times 10^{-11}$$

Inferência em Redes Bayesianas – Exemplo



P(r)
0,001

P(t)
0,002

R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

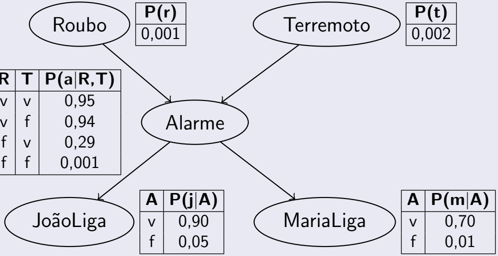
A	P(j|A)
v	0,90
f	0,05

A	P(m|A)
v	0,70
f	0,01

Fonte: Adaptado de AIMA. Russell & Norvig.

$$\begin{aligned}
P(r,j,m,t,a) &= P(r)P(t)P(a|r,t)P(j|a)P(m|a) &\approx 1,197\times 10^{-6}\\
P(r,j,m,t,\neg a) &= P(r)P(t)P(\neg a|r,t)P(j|\neg a)P(m|\neg a) &\approx 5\times 10^{-11}\\
P(r,j,m,\neg t,a) &= P(r)P(\neg t)P(a|r,\neg t)P(j|a)P(m|a) &\approx 5,91\times 10^{-4}
\end{aligned}$$

Inferência em Redes Bayesianas – Exemplo



P(r)
0,001

P(t)
0,002

R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

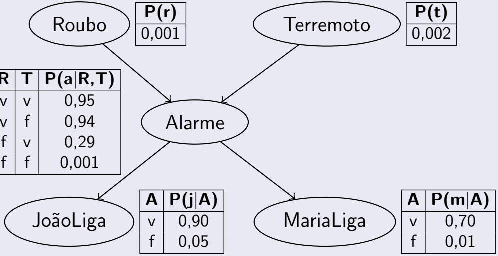
A	P(j|A)
v	0,90
f	0,05

A	P(m|A)
v	0,70
f	0,01

Fonte: Adaptado de AIMA. Russell & Norvig.

$$\begin{aligned}
P(r,j,m,t,a) &= P(r)P(t)P(a|r,t)P(j|a)P(m|a) &\approx 1,197 \times 10^{-6}\\
P(r,j,m,t,\neg a) &= P(r)P(t)P(\neg a|r,t)P(j|\neg a)P(m|\neg a) &\approx 5 \times 10^{-11}\\
P(r,j,m,\neg t,a) &= P(r)P(\neg t)P(a|r,\neg t)P(j|a)P(m|a) &\approx 5,91 \times 10^{-4}\\
P(r,j,m,\neg t,\neg a) &= P(r)P(\neg t)P(\neg a|r,\neg t)P(j|\neg a)P(m|\neg a) &\approx 2,994 \times 10^{-6}
\end{aligned}$$

Inferência em Redes Bayesianas – Exemplo



P(r)
0,001

P(t)
0,002

R	T	P(a|R,T)
v	v	0,95
v	f	0,94
f	v	0,29
f	f	0,001

A	P(j|A)
v	0,90
f	0,05

A	P(m|A)
v	0,70
f	0,01

Fonte: Adaptado de AIMA. Russell & Norvig.

$$\begin{aligned}
P(r,j,m,t,a) &= P(r)P(t)P(a|r,t)P(j|a)P(m|a) &&\approx 1,197 \times 10^{-6}\\
P(r,j,m,t,\neg a) &= P(r)P(t)P(\neg a|r,t)P(j|\neg a)P(m|\neg a) &&\approx 5 \times 10^{-11}\\
P(r,j,m,\neg t,a) &= P(r)P(\neg t)P(a|r,\neg t)P(j|a)P(m|a) &&\approx 5,91 \times 10^{-4}\\
P(r,j,m,\neg t,\neg a) &= P(r)P(\neg t)P(\neg a|r,\neg t)P(j|\neg a)P(m|\neg a) &&\approx 2,994 \times 10^{-6}
\end{aligned}$$

$$P(r,j,m) = \sum_{T=\{v,f\}} \sum_{A=\{v,f\}} P(r,j,m,T,A) \approx 0,00059224$$

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

- Lembre que queríamos $P(r|j,m) = \dfrac{P(r,j,m)}{P(j,m)}$

Inferência em Redes Bayesianas – Exemplo

- Lembre que queríamos $P(r|j,m) = \dfrac{P(r,j,m)}{P(j,m)}$
  - Já calculamos $P(r,j,m)$

Inferência em Redes Bayesianas – Exemplo

- Lembre que queríamos $P(r|j,m) = \dfrac{P(r,j,m)}{P(j,m)}$
  - Já calculamos $P(r,j,m)$
  - Falta $P(j,m)$

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

- Lembre que queríamos $P(r|j,m) = \dfrac{P(r,j,m)}{P(j,m)}$
  - Já calculamos $P(r,j,m)$
  - Falta $P(j,m)$
- Contudo, $P(j,m)$ não precisa ser calculado diretamente:

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

- Lembre que queríamos $P(r|j,m) = \dfrac{P(r,j,m)}{P(j,m)}$
  - Já calculamos $P(r,j,m)$
  - Falta $P(j,m)$
- Contudo, $P(j,m)$ não precisa ser calculado diretamente:
  - Podemos calcular $P(\neg r,j,m)$ pelo mesmo método que calculamos $P(r,j,m)$

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

- Lembre que queríamos $P(r|j,m) = \dfrac{P(r,j,m)}{P(j,m)}$
  - Já calculamos $P(r,j,m)$
  - Falta $P(j,m)$
- Contudo, $P(j,m)$ não precisa ser calculado diretamente:
  - Podemos calcular $P(\neg r,j,m)$ pelo mesmo método que calculamos $P(r,j,m)$
    - Nesse caso teríamos $P(\neg r,j,m) \approx 0,0014919$

Inferência em Redes Bayesianas – Exemplo

- Lembre que $P(\neg r|j, m) + P(r|j, m) = 1$

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

- Lembre que $P(\neg r|j,m) + P(r|j,m) = 1$
- Assim:

$$\begin{aligned} P(r|j,m) + P(\neg r|j,m) &= \frac{P(r,j,m)}{P(j,m)} + \frac{P(\neg r,j,m)}{P(j,m)} \\ &= \alpha \times [P(r,j,m) + P(\neg r,j,m)] \end{aligned}$$

com $\alpha = \dfrac{1}{P(j,m)}$ sendo a **constante de normalização**

Inferência em Redes Bayesianas – Exemplo

- Como $P(\neg r|j, m) + P(r|j, m) = 1$, temos que

$$\alpha \times [P(r, j, m) + P(\neg r, j, m)] = 1$$

$$\Rightarrow \alpha \times [0,00059224 + 0,0014919] = 1$$
$$\Rightarrow \alpha = \frac{1}{P(j, m)} \approx 479,81$$

Redes Bayesianas

Inferência em Redes Bayesianas – Exemplo

- Como $P(\neg r|j,m) + P(r|j,m) = 1$, temos que

$$\alpha \times [P(r,j,m) + P(\neg r,j,m)] = 1$$

$$\Rightarrow \alpha \times [0,00059224 + 0,0014919] = 1$$
$$\Rightarrow \alpha = \frac{1}{P(j,m)} \approx 479,81$$

Levando a

$$P(r|j,m) = \alpha \times P(r,j,m) \approx 0,284 = 28,4\%$$
$$P(\neg r|j,m) = \alpha \times P(\neg r,j,m) \approx 0,716 = 71,6\%$$

Inferência em Redes Bayesianas – Em Suma

- Para calcular qualquer combinação de valores:

Redes Bayesianas

Inferência em Redes Bayesianas – Em Suma

- Para calcular qualquer combinação de valores:
  - Identifique que variáveis estão na rede e não receberam valor

Inferência em Redes Bayesianas – Em Suma

- Para calcular qualquer combinação de valores:
  - Identifique que variáveis estão na rede e não receberam valor
  - Faça todas as combinações de valores com essas variáveis + os valores fixos definidos no problema, somando-as

Redes Bayesianas

Inferência em Redes Bayesianas – Em Suma

- Para calcular qualquer combinação de valores:
  - Identifique que variáveis estão na rede e não receberam valor
  - Faça todas as combinações de valores com essas variáveis + os valores fixos definidos no problema, somando-as
  - Cada combinação é calculada multiplicando-se as probabilidades de cada nó na rede, para os valores definidos pela combinação

Redes Bayesianas

Inferência em Redes Bayesianas

- Problemas:

Redes Bayesianas

Inferência em Redes Bayesianas

- Problemas:
  - Método bastante trabalhoso

Redes Bayesianas

Inferência em Redes Bayesianas

- Problemas:
  - Método bastante trabalhoso
  - No pior caso, intratável (NP-difícil)

Redes Bayesianas

Inferência em Redes Bayesianas

- Problemas:
  - Método bastante trabalhoso
  - No pior caso, intratável (NP-difícil)
- Solução:

Inferência em Redes Bayesianas

- Problemas:
  - Método bastante trabalhoso
  - No pior caso, intratável (NP-difícil)
- Solução:
  - Usar métodos para inferência aproximada

Inferência em Redes Bayesianas

- Problemas:
  - Método bastante trabalhoso
  - No pior caso, intratável (NP-difícil)
- Solução:
  - Usar métodos para inferência aproximada
    - Amostragem direta

Redes Bayesianas

Inferência em Redes Bayesianas

- Problemas:
  - Método bastante trabalhoso
  - No pior caso, intratável (NP-difícil)
- Solução:
  - Usar métodos para inferência aproximada
    - Amostragem direta
    - Algoritmos do tipo Markov Chain Monte Carlo (MCMC), como *simulated annealing* e *Gibbs sampling*

Redes Bayesianas

Inferência em Redes Bayesianas

- Problemas:
  - Método bastante trabalhoso
  - No pior caso, intratável (NP-difícil)
- Solução:
  - Usar métodos para inferência aproximada
    - Amostragem direta
    - Algoritmos do tipo Markov Chain Monte Carlo (MCMC), como *simulated annealing* e *Gibbs sampling*
    - Mesmo esses podem ser NP-difíceis, embora em muitos casos úteis

Variáveis Contínuas

Redes Bayesianas

Variáveis Contínuas

- Duas abordagens:

Redes Bayesianas

Variáveis Contínuas

- Duas abordagens:
  - Discretização

Variáveis Contínuas

- Duas abordagens:
  - Discretização
    - Separe os valores contínuos em intervalos discretos

Redes Bayesianas

Variáveis Contínuas

- Duas abordagens:
  - Discretização
    - Separe os valores contínuos em intervalos discretos
    - Há perda de precisão

Redes Bayesianas

Variáveis Contínuas

- Duas abordagens:
  - Discretização
    - Separe os valores contínuos em intervalos discretos
    - Há perda de precisão
    - Pode resultar em grande número de valores

Redes Bayesianas

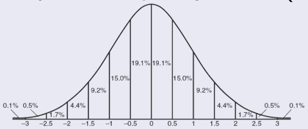
Variáveis Contínuas

- Duas abordagens:
  - Discretização
    - Separe os valores contínuos em intervalos discretos
    - Há perda de precisão
    - Pode resultar em grande número de valores
  - Use uma função de densidade de probabilidade como aproximação para $P(A|B)$. Ex:

Redes Bayesianas

Variáveis Contínuas

- Duas abordagens:
  - Discretização
    - Separe os valores contínuos em intervalos discretos
    - Há perda de precisão
    - Pode resultar em grande número de valores
  - Use uma função de densidade de probabilidade como aproximação para $P(A|B)$. Ex:



Fonte: https://coloneltedcampbell.blog/2016/02/13/the-800-pound-gorillas/normal67/

$$\begin{aligned} P(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= N(\mu, \sigma^2) \end{aligned}$$

Redes Bayesianas – Variáveis Contínuas

Usando uma função de densidade:

Redes Bayesianas – Variáveis Contínuas

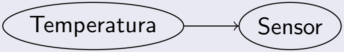
Usando uma função de densidade:

- Suponha que queremos modelar a medição de temperatura em uma determinada sala, usando um termômetro

Redes Bayesianas – Variáveis Contínuas

Usando uma função de densidade:

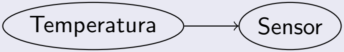
- Suponha que queremos modelar a medição de temperatura em uma determinada sala, usando um termômetro



Redes Bayesianas – Variáveis Contínuas

Usando uma função de densidade:

- Suponha que queremos modelar a medição de temperatura em uma determinada sala, usando um termômetro

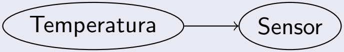


- O termômetro (sensor), contudo, não é perfeito

Redes Bayesianas – Variáveis Contínuas

Usando uma função de densidade:

- Suponha que queremos modelar a medição de temperatura em uma determinada sala, usando um termômetro



- O termômetro (sensor), contudo, não é perfeito
  - Mede em torno da temperatura real, mas não exatamente

Redes Bayesianas – Variáveis Contínuas

Usando uma função de densidade:

- Suponha que queremos modelar a medição de temperatura em uma determinada sala, usando um termômetro

$$\text{Temperatura} \longrightarrow \text{Sensor}$$

- O termômetro (sensor), contudo, não é perfeito
  - Mede em torno da temperatura real, mas não exatamente
  - $S \approx N(t; \sigma_S^2) \rightarrow$ normal centrada em t, com desvio padrão σ_S

Redes Bayesianas – Variáveis Contínuas

Usando uma função de densidade:

- Agora imagine que queremos poder inferir como estará a temperatura na sala daqui a alguns instantes

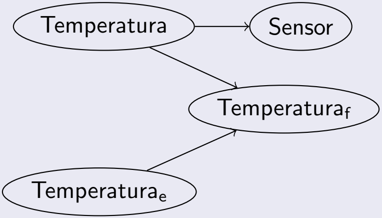
Usando uma função de densidade:

- Agora imagine que queremos poder inferir como estará a temperatura na sala daqui a alguns instantes
  - Essa depende, no entanto, da temperatura atual e da temperatura externa, dado que sempre há uma troca de calor com o exterior

Redes Bayesianas – Variáveis Contínuas

Usando uma função de densidade:

- Agora imagine que queremos poder inferir como estará a temperatura na sala daqui a alguns instantes
  - Essa depende, no entanto, da temperatura atual e da temperatura externa, dado que sempre há uma troca de calor com o exterior

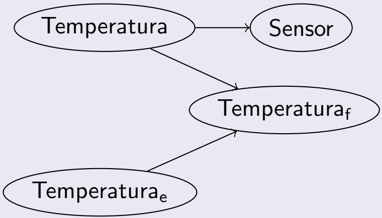


  - Temos então:

Redes Bayesianas – Variáveis Contínuas

Usando uma função de densidade:

- Um possível modelo para T_f seria



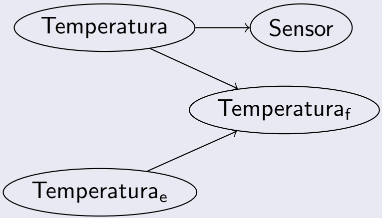
Fonte: Coursera (Cont. Variables). Koller.

$$T_f \approx N(\alpha T + (1-\alpha) T_e, \sigma^2_{T_f})$$

Redes Bayesianas – Variáveis Contínuas

Usando uma função de densidade:

- Um possível modelo para T_f seria



Fonte: Coursera (Cont. Variables). Koller.

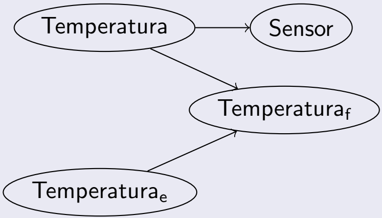
$$T_f \approx N(\alpha T + (1-\alpha) T_e, \sigma^2_{T_f})$$

centrada na média ponderada das temperaturas no momento atual (T) e externa (T_e)

Redes Bayesianas – Variáveis Contínuas

Usando uma função de densidade:

- Um possível modelo para T_f seria



Fonte: Coursera (Cont. Variables). Koller.

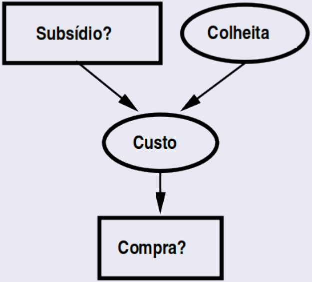
$$T_f \approx N(\alpha T + (1-\alpha) T_e, \sigma^2_{T_f})$$

centrada na média ponderada das temperaturas no momento atual (T) e externa (T_e)

<u>Modelo Linear Gaussiano</u>

Mistura Discretas + Contínuas – Exemplo

- Alguém vai comprar frutas, dependendo do preço

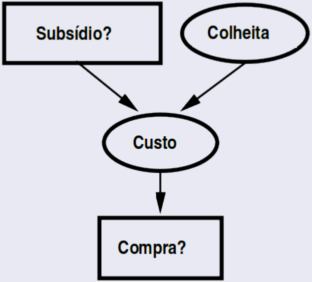


Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas

Mistura Discretas + Contínuas – Exemplo

- Alguém vai comprar frutas, dependendo do preço
  - O preço depende do tamanho da colheita e do fato do governo ter dado ou não subsídios

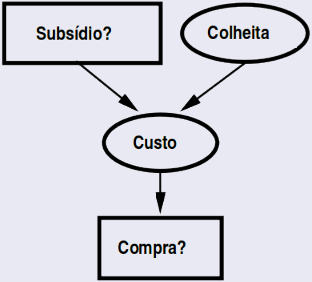


Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas

Mistura Discretas + Contínuas – Exemplo

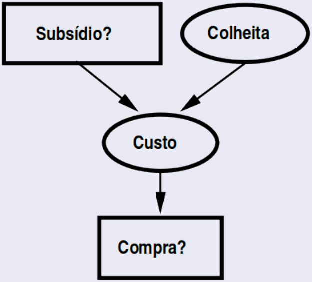
- Alguém vai comprar frutas, dependendo do preço
  - O preço depende do tamanho da colheita e do fato do governo ter dado ou não subsídios
- Variáveis contínuas



Fonte: Slides de AIMA. Russell & Norvig.

Mistura Discretas + Contínuas – Exemplo

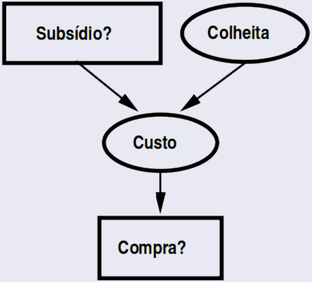
- Alguém vai comprar frutas, dependendo do preço
  - O preço depende do tamanho da colheita e do fato do governo ter dado ou não subsídios
- Variáveis contínuas
  - Custo e Colheita



Fonte: Slides de AIMA. Russell & Norvig.

Mistura Discretas + Contínuas – Exemplo

- Alguém vai comprar frutas, dependendo do preço
  - O preço depende do tamanho da colheita e do fato do governo ter dado ou não subsídios
- Variáveis contínuas
  - Custo e Colheita
- Variáveis discretas

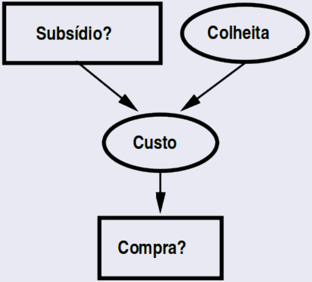


Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas

Mistura Discretas + Contínuas – Exemplo

- Alguém vai comprar frutas, dependendo do preço
  - O preço depende do tamanho da colheita e do fato do governo ter dado ou não subsídios
- Variáveis contínuas
  - Custo e Colheita
- Variáveis discretas
  - Subsídio (Sim ou não)

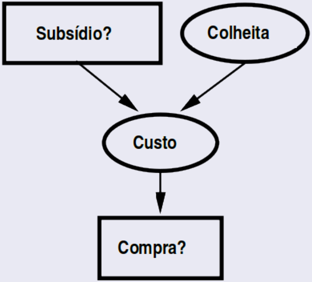


Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas

Mistura Discretas + Contínuas – Exemplo

- Alguém vai comprar frutas, dependendo do preço
  - O preço depende do tamanho da colheita e do fato do governo ter dado ou não subsídios
- Variáveis contínuas
  - Custo e Colheita
- Variáveis discretas
  - Subsídio (Sim ou não)
  - Compra (Sim ou não)

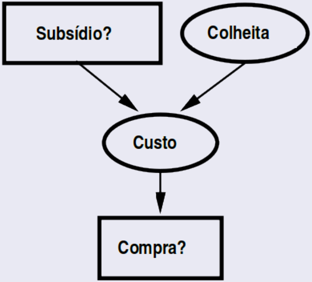


Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas – Exemplo

Variáveis contínuas com pais mistos

- Modelamos uma dependência contínua-contínua com um dos pais, para cada valor discreto do outro pai

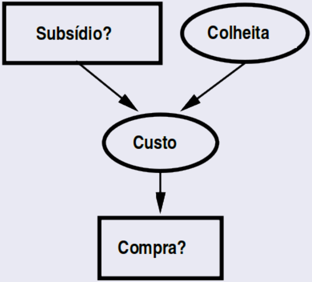


Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas – Exemplo

Variáveis contínuas com pais mistos

- Modelamos uma dependência contínua-contínua com um dos pais, para cada valor discreto do outro pai
  - $P(Custo|Colheita, subsídio) = \frac{1}{\sigma_v\sqrt{2\pi}} e^{\frac{-(Custo-(a_v \times Colheita + b_v))^2}{2\sigma_v^2}}$

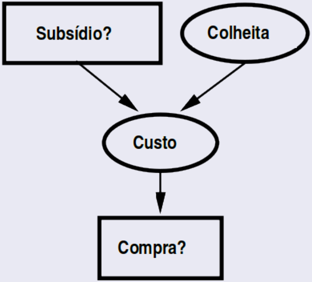


Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas – Exemplo

Variáveis contínuas com pais mistos

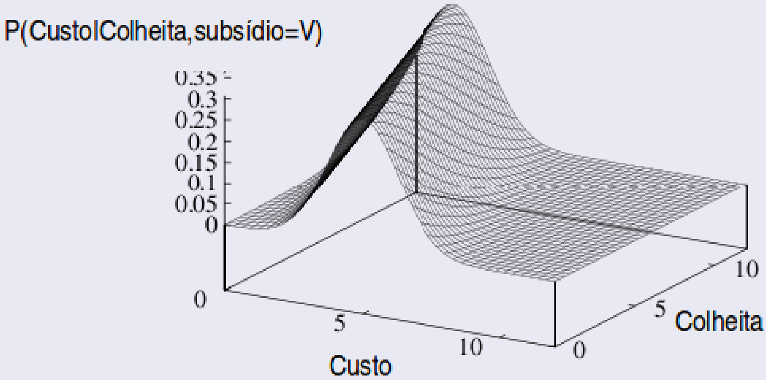
- Modelamos uma dependência contínua-contínua com um dos pais, para cada valor discreto do outro pai
  - $P(Custo|Colheita, subsídio) = \frac{1}{\sigma_v\sqrt{2\pi}} e^{\frac{-(Custo-(a_v \times Colheita + b_v))^2}{2\sigma_v^2}}$
  - $P(Custo|Colheita, \neg subsídio) = \frac{1}{\sigma_f\sqrt{2\pi}} e^{\frac{-(Custo-(a_f \times Colheita + b_f))^2}{2\sigma_f^2}}$



Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas – Exemplo

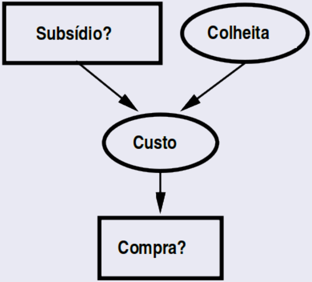
Variáveis contínuas com pais discretos



Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas – Exemplo

Variáveis discretas com pais contínuos:

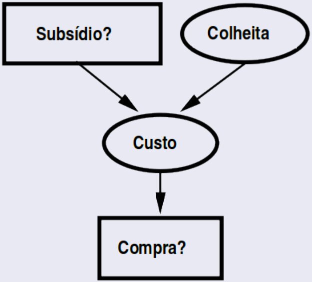


Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas – Exemplo

Variáveis discretas com pais contínuos:

- A distribuição deve seguir um degrau suave

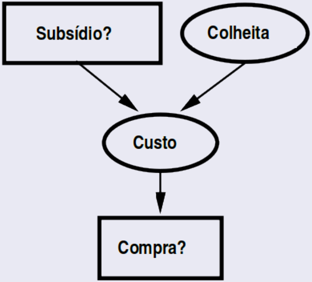


Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas – Exemplo

Variáveis discretas com pais contínuos:

- A distribuição deve seguir um degrau suave
- Espera-se que o cliente compre se o custo for baixo e não compre se for alto, sendo que a probabilidade entre esses valores varia suavemente

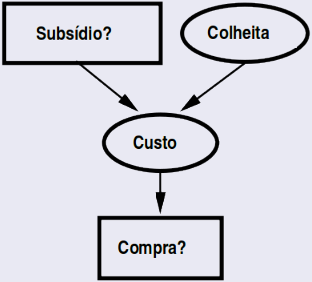


Fonte: Slides de AIMA. Russell & Norvig.

Redes Híbridas – Exemplo

Variáveis discretas com pais contínuos:

- A distribuição deve seguir um degrau suave
- Espera-se que o cliente compre se o custo for baixo e não compre se for alto, sendo que a probabilidade entre esses valores varia suavemente



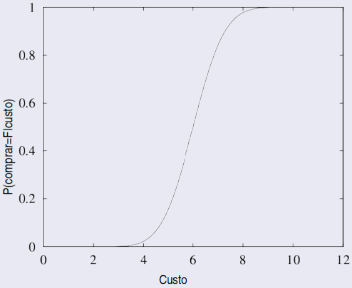
Fonte: Slides de AIMA. Russell & Norvig.

- Um meio de fazer tais degraus é usar a integral da distribuição normal

Redes Híbridas – Exemplo

Variáveis discretas com pais contínuos:

$$\Phi(x) = \int_{-\infty}^{x} N(0,1)(x)dx$$



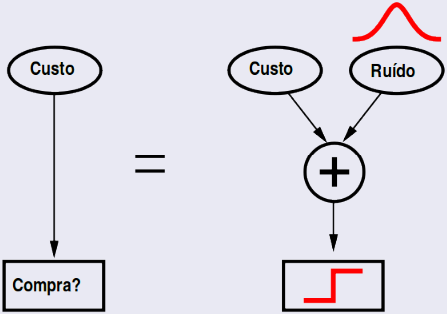
Fonte: Slides de AIMA. Russell & Norvig.

$$P(comprar|Custo = c) = \Phi(\frac{-c + \mu}{\sigma})$$

O degrau vai ocorrer ao redor de μ

Justificativa para a Gaussiana

- Pode ser vista como um degrau normal, cuja localização precisa está sujeita a ruído aleatório normalmente distribuído:

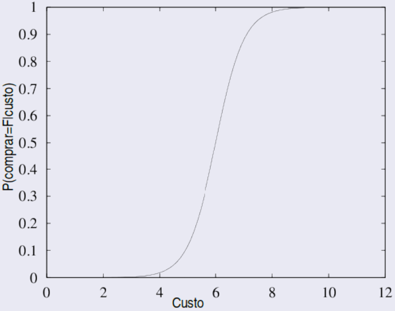


Fonte: Slides de AIMA. Russell & Norvig.

Alternativa à Integral da Gaussiana

- Usar uma função sigmoide:

$$P(comprar|Custo = c) = \frac{1}{1 + e^{-2\frac{-c+\mu}{\sigma}}}$$



Fonte: Slides de AIMA. Russell & Norvig.

Similar à integral da Gaussiana, porém mais alongada

Referências

- Russell, S.; Norvig P. (2010): Artificial Intelligence: A Modern Approach. Prentice Hall. 3a ed.
  - Slides do livro: `aima.eecs.berkeley.edu/slides-pdf/`
- Mitchell, T.M.: Machine Learning. McGraw-Hill. 1997.
- Murphy, K. P.: Machine Learning: A Probabilistic Perspective. MIT Press. 2012.
- Cover, T.M.; Thomas, J.A.: Elements of Information Theory. 2 ed. Wiley. 2006.
- Alpaydın, E.: Introduction to Machine Learning. 2 ed. MIT Press. 2010.

Referências

- Bouckaert, R.R.: Bayesian Belief Networks: from Construction to Inference. Ph.D. thesis. University of Utrecht. 1995.
- Cheng, J.; Greiner, R.: Comparing bayesian network classifiers. Proceedings UAI, 101–107. 1999.
- Cooper, G.; Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. Machine Learning, 9: 309–347. 1992.
- Duda, R.O.: Hart, P.E.: Stork, D.G.: Pattern Classification. 2 ed. Wiley. 2001.
- Friedman, N.; Geiger, D.; Goldszmidt, M.: Bayesian Network Classifiers. Machine Learning, 29: 131–163. 1997.
- Koller, D.: `https://pt.coursera.org/lecture/probabilistic-graphical-models/continuous-variables-wkNvM`