Redes de Bayes

Capítulo 14.1–3

Resumo

◊ Sintaxe

◊ Semântica

◊ Distribuições parametrizadas

Redes de Bayes

Têm uma notação gráfica simples para representar independência condicional.
São uma especificação compacta de distribuições conjuntas completas.

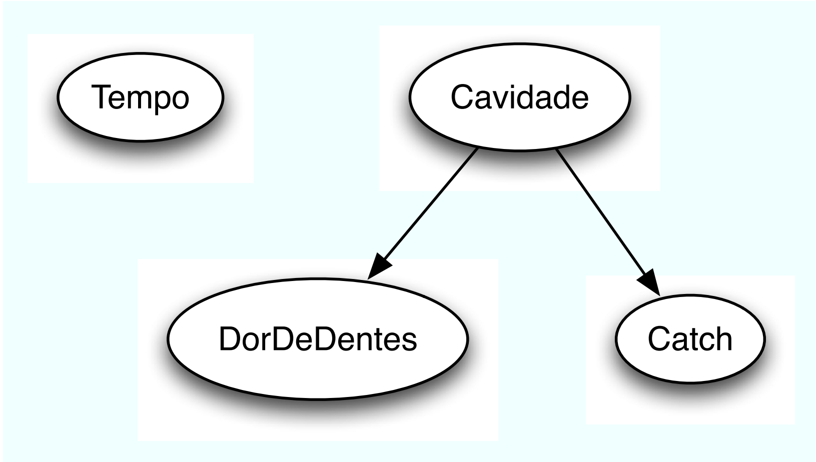
Sintaxe:

um conjunto de nós, um para cada variável
um grafo directo sem ciclos (link $\approx$ "influencia directamente")
uma distribuição condicional para cada nó dados os pais:
$\mathbf{P}(X_i|Pai(X_i))$

No caso mais simples, uma distribuição condicional é representada como uma Tabela de probabilidades condicional (CPT) dando a
a distribuição de X_i para cada combinação dos valores dos pais.

Exemplo

Topologia da rede que codifica afirmações condicionalmente independentes:



$Tempo$ é independente das outras variáveis

$DorDeDentes$ e $Catch$ são condicionalmente independentes dada $Cavidade$

Exemplo

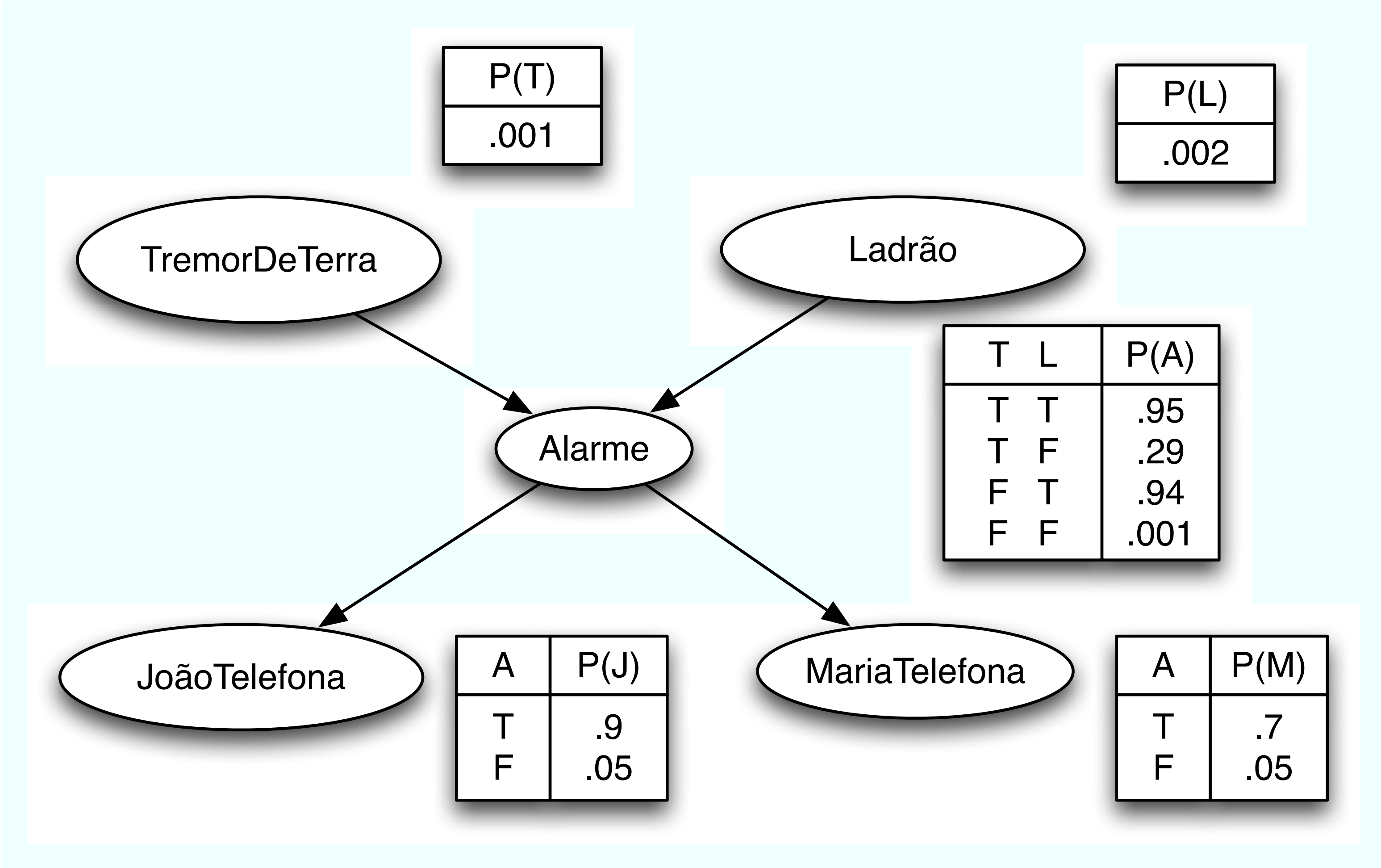
Estou no trabalho, o meu vizinho João telefona dizendo que o meu alarme está a tocar, mas a minha vizinha Maria não me telefona. Às vezes o alarme dispara devido a pequenos tremores de terra. Há um ladrão?

Variáveis: $Ladrao$, $TremorDeTerra$, $Alarme$, $JoaoTelefona$, $MariaTelefona$

A tipologia da rede reflecte o conhecimento "causal"

- Um ladrão faz o alarme disparar
- Um tremor de terra faz o alarme disparar
- O som do alarme faz a Maria telefonar
- O som do alarme faz o João telefonar

Exemplo



P(T)
.001

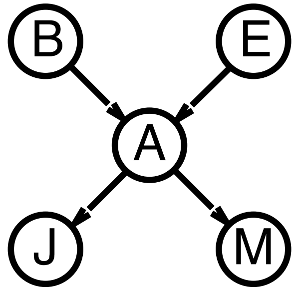
P(L)
.002

T	L	P(A)
T	T	.95
T	F	.29
F	T	.94
F	F	.001

A	P(J)
T	.9
F	.05

A	P(M)
T	.7
F	.05

Representação Compacta



Uma tabela CPT para Booleanos X_i com k pais Booleanos tem 2^k linhas para as combinações dos valores dos pais.

Cada linha requer um valor p para $X_i = true$ (o valor $X_i = false$ é $1 - p$)

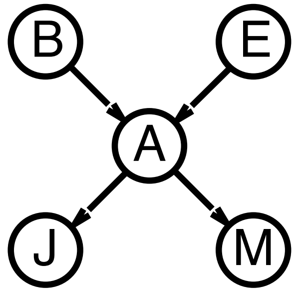
Cada variável não tem mais do que k pais, a rede completa necessita de $O(n \cdot 2^k)$ valores

I.e., cresce linearmente com n, vs. $O(2^n)$ para a distribuição conjunta completa.

Para a rede assalto, $1 + 1 + 4 + 2 + 2 = 10$ valores (vs. $2^5 - 1 = 31$)

Semântica (numérica)

A semântica define a distribuição conjunta completa como o produto das distribuições condicionais locais:



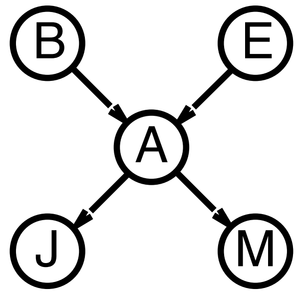
$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pais(X_i))$$

e.g., $P(j \wedge m \wedge a \wedge \neg l \wedge \neg t)$

$$=$$

Semântica numérica

A semântica define a distribuição conjunta completa como o produto das distribuições condicionais locais:



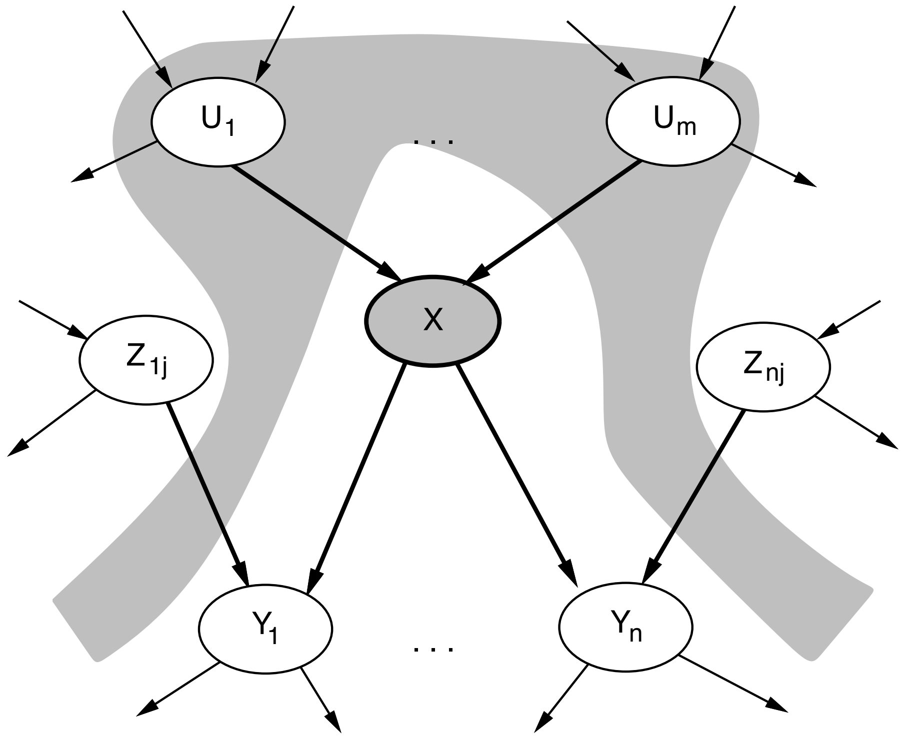
$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pais(X_i))$$

e.g., $P(j \wedge m \wedge a \wedge \neg l \wedge \neg t)$

$$\begin{aligned} &= P(j|a)P(m|a)P(a|\neg l, \neg t)P(\neg l)P(\neg t) \\ &= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\ &\approx 0.00063 \end{aligned}$$

Semântica da tipologia da rede

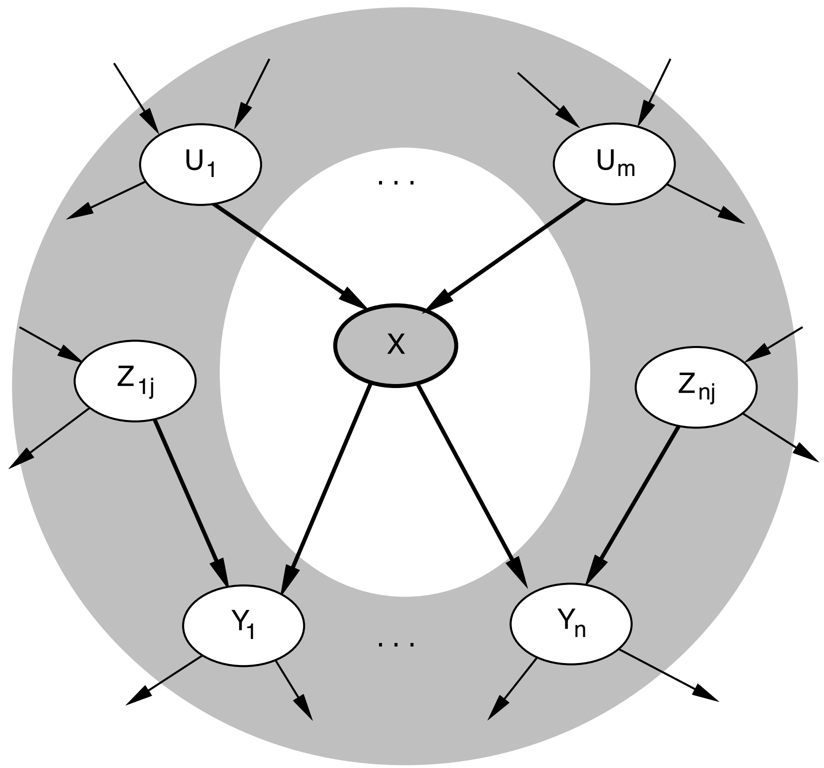
Semântica: cada nó é condicionalmente independente dos nós que não são seus descendentes dados os pais.



Teorema: Semântica da tipologia de rede $\Leftrightarrow$ Semântica numérica

Cobertor de Markov (Markov blanket)

Cada nó é condicionalmente independente de outros dado o seu cobertor de Markov, Markov blanket: pais + filhos + pais dos filhos



Construção de redes de Bayes

Um método que dado um conjunto independências condicionais garante a semântica da tipologia da rede.

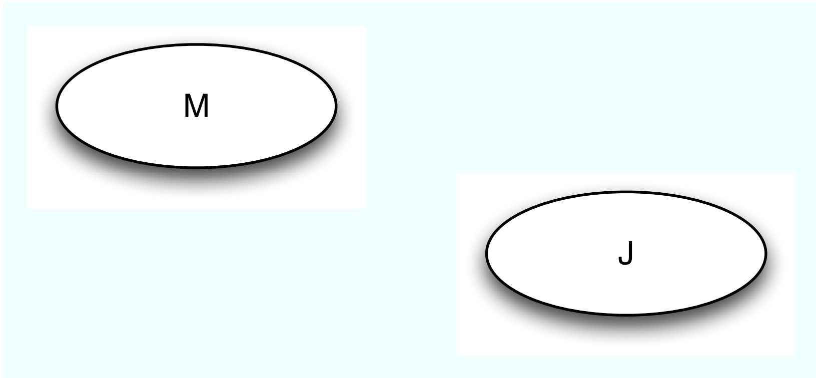
1. Escolher uma ordem nas variáveis $X_1, \ldots, X_n$
2. Para $i = 1$ até n
   juntar X_i à rede
   escolher pais em $X_1, \ldots, X_{i-1}$ tal que
   $$\mathbf{P}(X_i | pais(X_i)) = \mathbf{P}(X_i | X_1,\ \ldots,\ X_{i-1})$$

Esta escolha de pais garante a semântica numérica:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \textstyle\prod_{i=1}^{n} \mathbf{P}(X_i | X_1,\ \ldots,\ X_{i-1}) \quad \text{(regra da cadeia)} \\ &= \textstyle\prod_{i=1}^{n} \mathbf{P}(X_i | Pais(X_i)) \quad \text{(na construção)}\end{aligned}$$

Exemplo

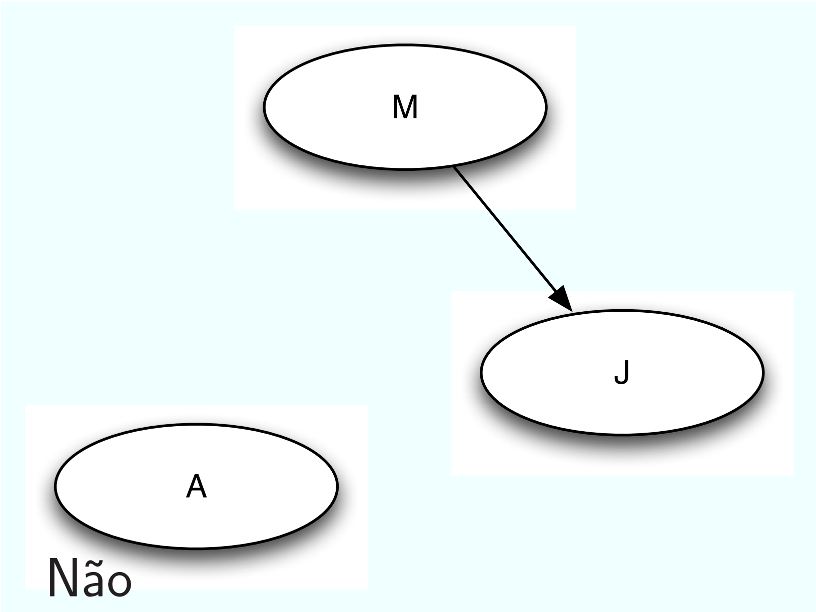
Suponha que se escolhe a ordem M, J, A, L, T



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

Exemplo

Suponha que se escolhe a ordem M, J, A, L, T

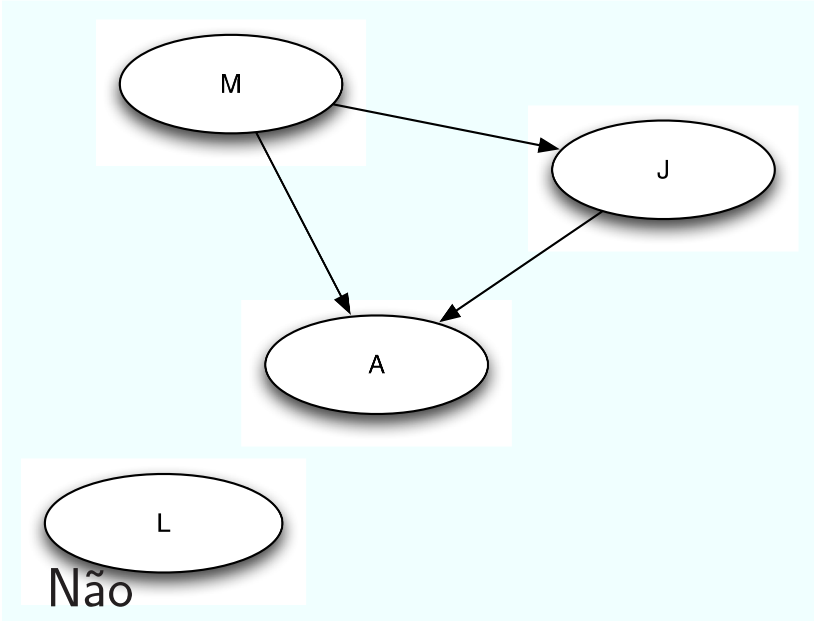


$P(J|M) = P(J)$? Não

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

Exemplo

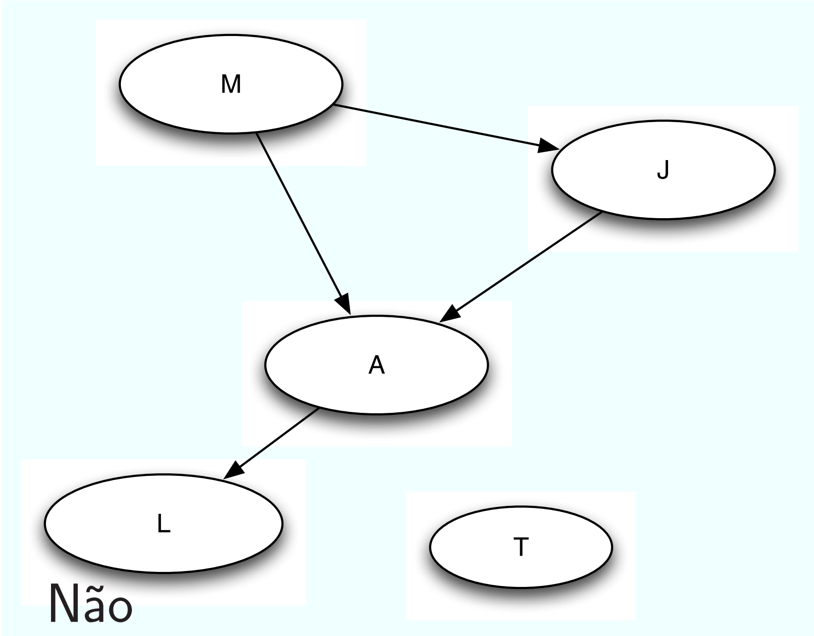
Suponha que se escolhe a ordem M, J, A, L, T



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

Exemplo

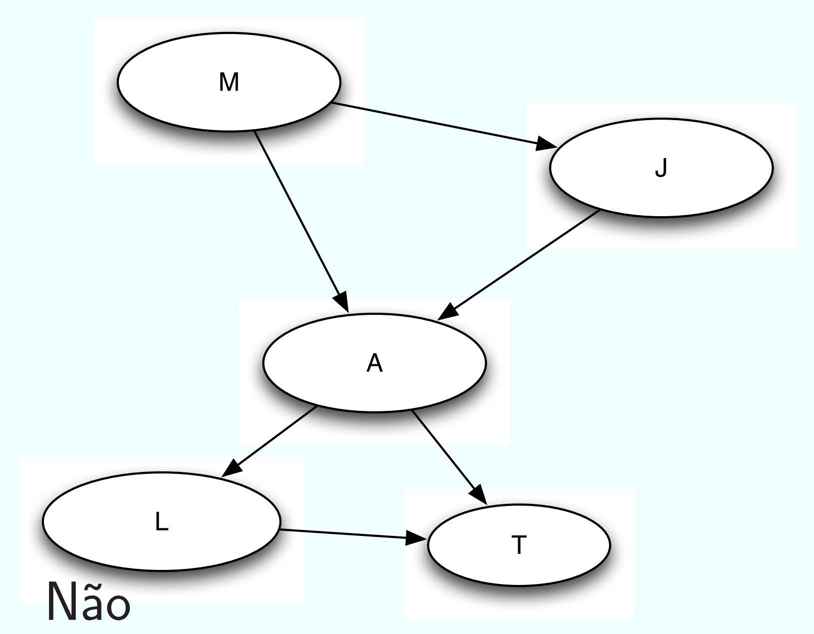
Suponha que se escolhe a ordem M, J, A, L, T



$P(J|M) = P(J)$? Não
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? Não
$P(L|A, J, M) = P(L|A)$? Sim
$P(L|A, J, M) = P(L)$? Não
$P(T|L, A, J, M) = P(T|A)$?
$P(T|L, A, J, M) = P(T|A, L)$?

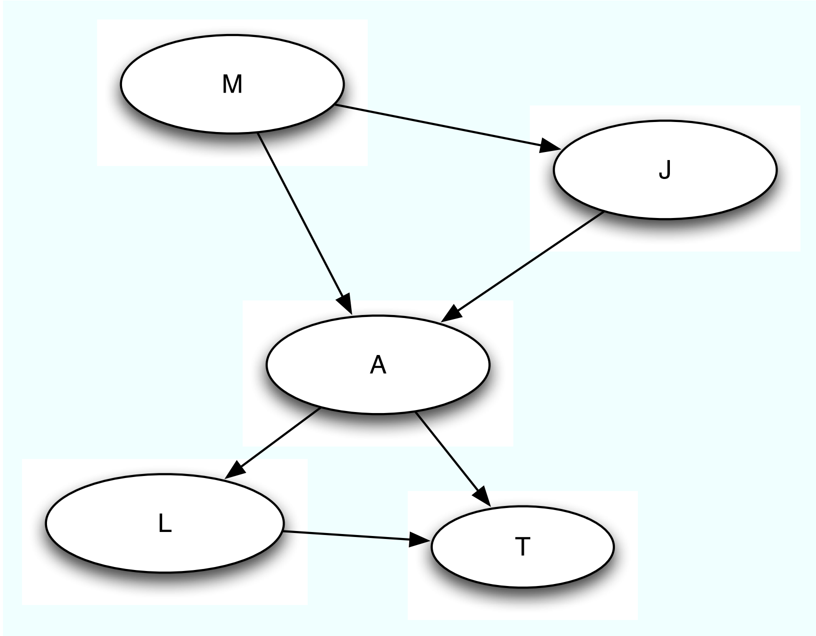
Exemplo

Suponha que se escolhe a ordem M, J, A, L, T



$P(J|M) = P(J)$? Não
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? Não
$P(L|A,J,M) = P(L|A)$? Sim
$P(L|A,J,M) = P(L)$? Não
$P(T|L,A,J,M) = P(T|A)$? Não
$P(T|L,A,J,M) = P(T|A,B)$? Sim

Exemplo



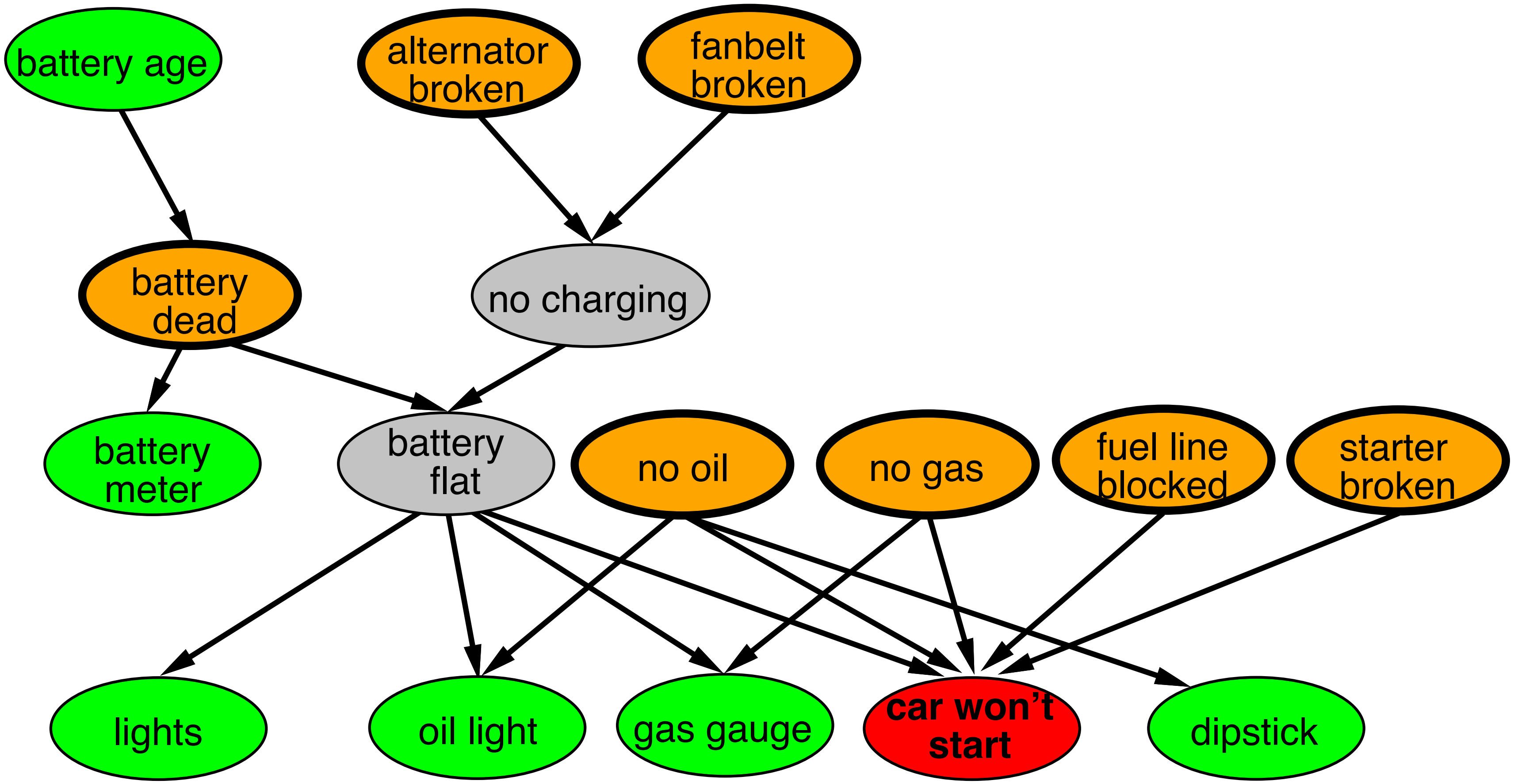
Decidir se há independência condicional é muito difícil quando não se vai no sentido da causa para o efeito.

(Definir modelos causais e independência condicional é uma tarefa intuitiva)

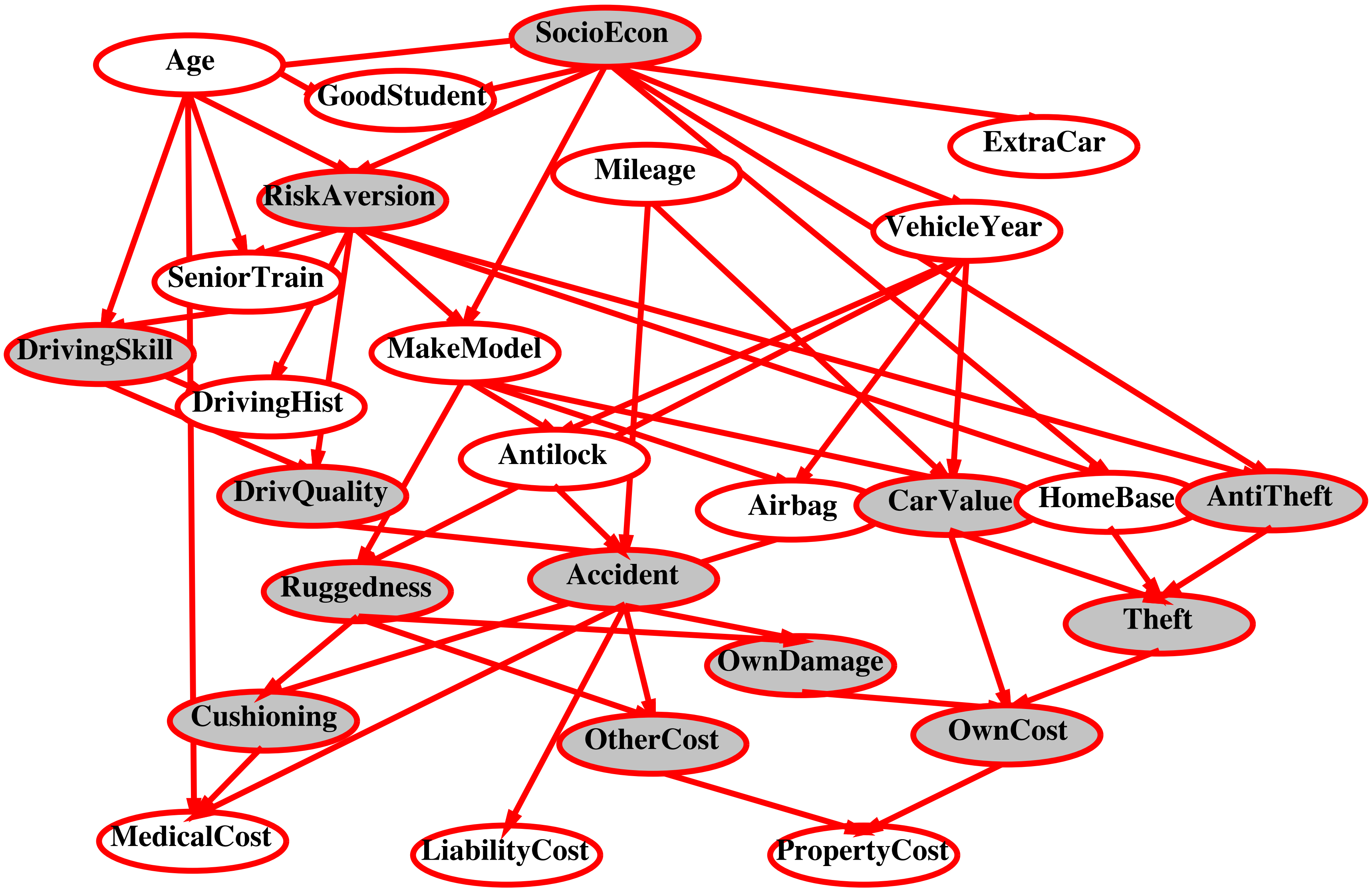
A rede final é menos compacta que a inicial: $1+2+4+2+4=13$ valores.

Exemplo: Diagnostico de um Carro

Evidência inicial: o carro não trabalha, variáveis de teste (verde), variáveis "avarias" (laranja) Variáveis escondidas (cinzento) assegura uma estrutura local ou esparsa, parâmetros reduzidos.



Exemplo: Segurança Rodoviária



Distribuições Condicionais Compactas

CPT cresce exponencialmente com o número de pais
CPT é infinita quando os pais ou os filhos têm valores contínuos.

Solução: distribuições canónicas definidas de forma compacta

nós Deterministicos são o caso mais simples:
$X = f(Pais(X))$ para alguma função f

E.g., Funções Booleanas
$AmericanosDoNorte \Leftrightarrow Canadianos \vee EstadosUnidos \vee Mexicanos$

E.g., relações entre variáveis continuas

$$\frac{\partial Nivel}{\partial t} = \text{fluxoEntrada} + \text{precipitação - fluxoSaida - evaporação}$$

Distribuições Condicionais Compactas

Noisy-OR Modelo de distribuição para causa múltiplas não interactivas
    1) Pais $U_1 \ldots U_k$ incluem todas as causas (pode-se adicionar leak node)
    2) probabilidade independente de falha q_i para cada causa sozinha

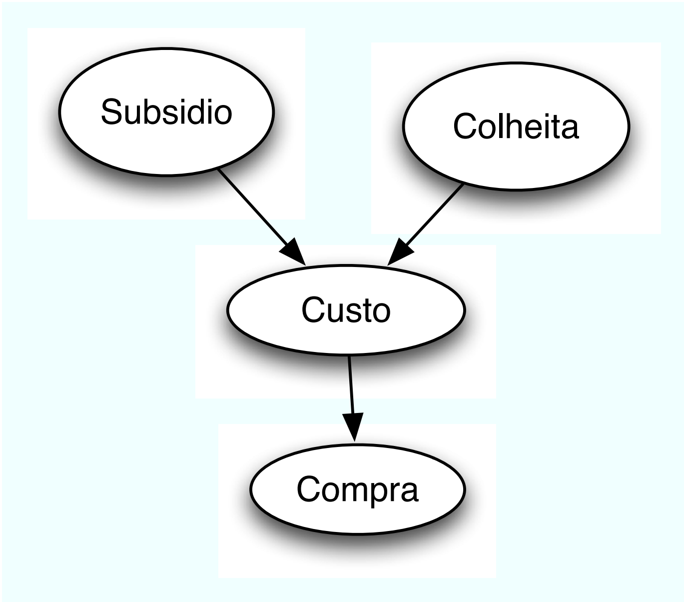
$$\Rightarrow P(X|U_1 \ldots U_j, \neg U_{j+1} \ldots \neg U_k) = 1 - \prod_{i=1}^{j} q_i$$

$Constipacao$	$Gripe$	$Malaria$	$P(Febre)$	$P(\neg Febre)$
F	F	F	$\mathbf{0.0}$	1.0
F	F	T	0.9	$\mathbf{0.1}$
F	T	F	0.8	$\mathbf{0.2}$
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	$\mathbf{0.6}$
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Número de parâmetros é **linear** no número de pais.

Rede Híbrida (discreta+continua)

Variáveis discretas ($Subsidio?$ e $Compra?$); continuas ($Colheita$ e $Custo$)



Opção 1: discretização— possibilidade de cometer erros, tabelas CPT grandes
Opção 2: parametrização de funções canónicas

1) Variáveis continuas, pais discretos+contínuos (e.g., $Custo$)
2) Variáveis discretas, pais contínuos (e.g., $Compra?$)

Filhos que são variáveis continuas

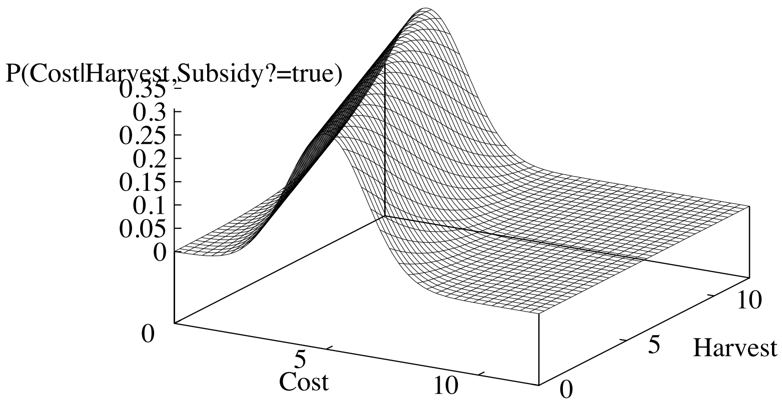
É necessária uma função de densidade condicional para a variável filha dadas as variáveis continuas dos pais, para cada afectação possível dos pais que são variáveis discretas.

A função mais comum é o modelo linear de Gauss, e.g.,:

$$
\begin{aligned}
&P(Custo = c | Colheita = h, Subsidio? = true) \\
&= N(a_t h + b_t, \sigma_t)(c) \\
&= \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{c - (a_t h + b_t)}{\sigma_t}\right)^2\right)
\end{aligned}
$$

A média $Custo$ varia linearmente com $Colheita$, o desvio padrão é fixo.

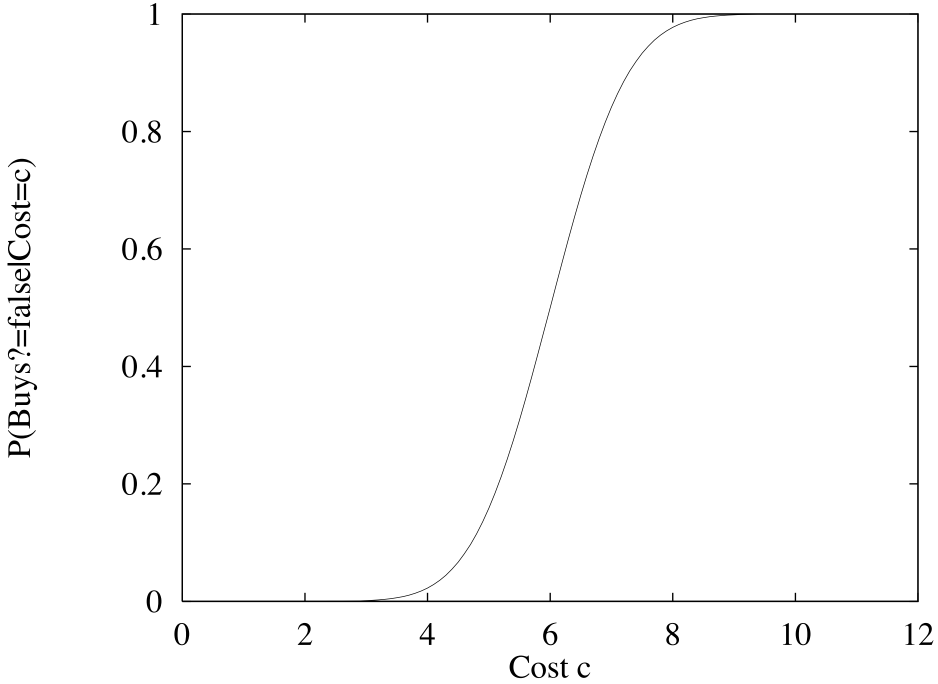
Filhos que são variáveis continuas



Uma rede linear de Gauss para variáveis Discretas+continuas é uma rede condicional de Gauss i.e., uma função de Gauss multivariada sobre todas as variáveis continuas para cada combinação dos valores das variáveis discretas.

Variável discreta com pais contínuos

Probabilidade de $Compra?$ dado $Custo$ deve ser um "soft threshold":



Probit esta distribuição usa o integral da função de Gauss:

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

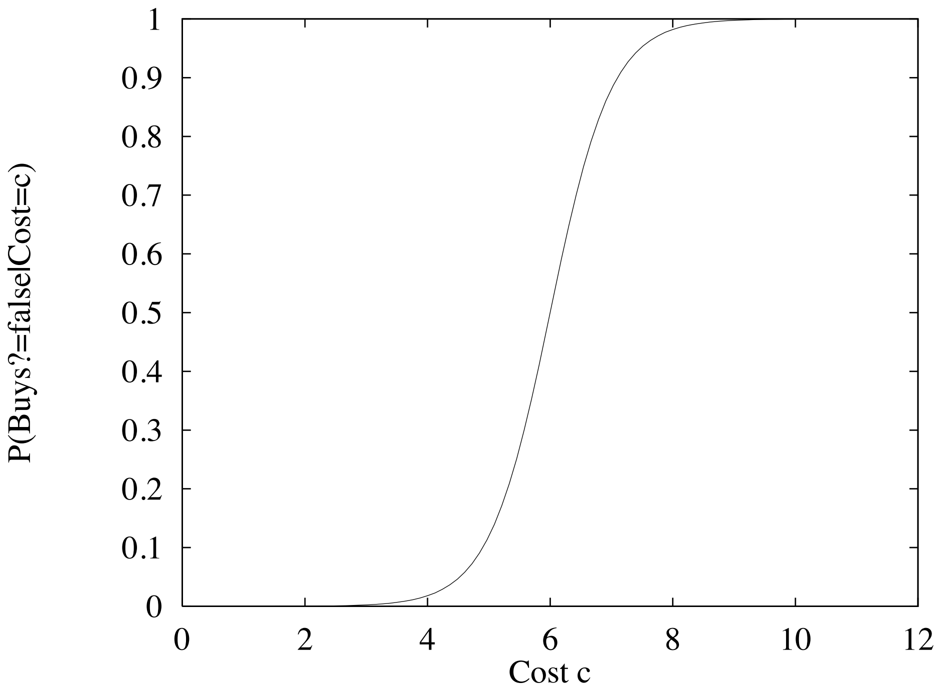
$P(Buys? = true \mid Cost = c) = \Phi((-c + \mu)/\sigma)$

Variável discreta com pais contínuos

A distribuição Sigmoid (ou logit) também é usada em redes neuronais:

$$P(Compra? = true \mid Custo = c) = \frac{1}{1 + exp(-2\frac{-c+\mu}{\sigma})}$$

A função Sigmoid tem uma forma semelhante à proibit:



Resumo

As redes de Bays providenciam uma representação natural para a independência condicional (induzida por causas)

Topologia + CPT = representação compacta para a distribuição conjunta.

Fácil de construir por não peritos.

Distribuições canónicas (e.g., noisy-OR) = representação compacta da CPTs

Variáveis continuas $\Rightarrow$ distribuições parametrizadas (e.g., funções lineares de Gauss)