Inteligência Artificial

Lidando com incerteza (parte 2)

Profa Debora Medeiros

Naïve-Bayes

- Classificador
  - Linear em n

Naive-Bayes

$$\mathbf{P}(\mathrm{Causa}, \mathrm{Efeito}_1, \ldots, \mathrm{Efeito}_n) = \mathbf{P}(\mathrm{causa}) \prod_i \mathbf{P}(\mathrm{Efeito}_i \mid \mathrm{Causa})$$

Modelagem Estatística (Bayesiana):

- Naive Bayes (NB) usa todos os atributos. Baseado em duas premissas:
  - Atributos igualmente importantes e condicionalmente independentes
    - o valor de um atributo não influencia no valor de outro atributo, dada a informação da classe;
- Na prática, tais premissas são frequentemente violadas, mas ainda assim o NB é muito competitivo:
  - Probabilidades estimadas não precisam necessariamente ser corretas, o que importa são as avaliações relativas.
- Parece ser consenso entre os mineradores de dados que, na prática, deve ser o primeiro algoritmo a usar.

Witten&Frank

Noção Intuitiva (base de dados “weather”):

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Desejamos estimar:

$$P(C_k|x) = \frac{P(x|C_k)P(C_k)}{P(x)}$$

- $P(C_k)$ pode ser estimada a partir da frequência relativa de classes;
- $P(\mathbf{x})$ é a *constante de normalização*

→ Como estimar $P(\mathbf{x}/C_k)$?

→ Assumindo independência condicional temos:

Frequências relativas:

Outlook			Temperature			Humidity			Windy			Play	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

Para um novo exemplo:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Verosssimilhança para as duas classes:

Para "yes" = 2/9 × 3/9 × 3/9 × 3/9 × 9/14 = 0.0053

Para "no" = 3/5 × 1/5 × 4/5 × 3/5 × 5/14 = 0.0206

Convertendo para probabilidades por meio de normalização:

P("yes") = 0.0053 / (0.0053 + 0.0206) = 0.205

P("no") = 0.0206 / (0.0053 + 0.0206) = 0.795

Witten&Frank

Regra de Bayes:

- Probabilidade de um evento *H* dada a evidência *E* :

$$\Pr[H|E] = \frac{\Pr[E|H]\Pr[H]}{\Pr[E]}$$

- Probabilidade *a priori* para H : $\Pr[H]$
  - Probabilidade de um evento antes de verificar a evidência
- Probabilidade *a posteriori* para H : $\Pr[H|E]$
  - Probabilidade de um evento após verificar a evidência

Thomas Bayes

(1702-1761)

Witten&Frank

Naïve Bayes para classificação:

- Qual é a probabilidade da classe dado um exemplo?
  - Evidência *E* = exemplo (valores dos atributos previsores);
  - Evento *H* = classe para um exemplo;
- Premissa *naïve*: evidência *dividida* em partes (i.e. atributos) independentes.

$$Pr[H|E] = \frac{Pr[E_1|H]Pr[E_2|H]\ldots Pr[E_n|H]Pr[H]}{Pr[E]}$$

Para o nosso exemplo:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Evidência E

$$\Pr[yes \mid E] = \Pr[Outlook = Sunny \mid yes] \times \Pr[Temperature = Cool \mid yes] \times \Pr[Humidity = High \mid yes] \times \Pr[Windy = True \mid yes] \times \frac{\Pr[yes]}{\Pr[E]}$$

$$= \frac{\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{\Pr[E]}$$

Probabilidade da classe "yes"

Witten&Frank

Problema da frequência *zero:*

- O que acontece se um determinado valor de atributo não aparece na base de treinamento, mas aparece no exemplo de teste?
(e.g. “outlook=overcast” para classe “no”)
    - Probabilidade correspondente será zero.
    - *Probabilidade a posteriori* será também zero.

- Possível solução: adicionar 1 ao contador para cada combinação de valor-classe (Estimador de *Laplace*)*.* Como resultado, as probabilidades nunca serão *zero.*

Estimativas das probabilidades modificadas:

- No caso geral, pode-se adicionar uma constante μ diferente de 1;
- Exemplo: atributo *outlook* para a classe *yes:*

$$\frac{2+\mu/3}{9+\mu} \qquad \frac{4+\mu/3}{9+\mu} \qquad \frac{3+\mu/3}{9+\mu}$$

Sunny ***Overcast*** ***Rainy***

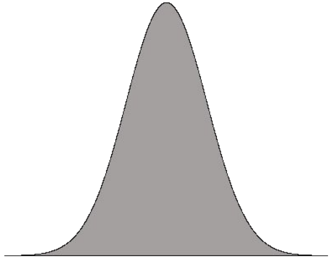
Atributos numéricos:

- Por exemplo, pode-se assumir uma distribuição Gaussiana para estimar as probabilidades:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\sigma = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \mu)^2$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Estatísticas para "weather":

Outlook			Temperature		Humidity		Windy			Play	
	Yes	*No*	*Yes*	*No*	*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	64, 68,	65, 71,	65, 70,	70, 85,	False	6	2	9	5
Overcast	4	0	69, 70,	72, 80,	70, 75,	90, 91,	True	3	3		
Rainy	3	2	72, …	85, …	80, …	95, …					
Sunny	2/9	3/5	μ =73	μ =75	μ =79	μ =86	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	σ =6.2	σ =7.9	σ =10.2	σ =9.7	True	3/9	3/5		
Rainy	3/9	2/5									

$$f(temperature=66 | yes)=\frac{1}{\sqrt{2\pi}6.2}e^{-\frac{(66-73)^2}{2*6.2^2}}=0.0340$$

Discussão para Naïve Bayes:

- Naïve Bayes **funciona bem**, mesmo quando suas **premissas são violadas**. Classificação **não requer estimativas precisas** da probabilidade, desde que a **probabilidade máxima** seja atribuída à **classe correta** (Domingos & Pazzani, On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, Machine Learning 29, 103-130, 1997).
- Entretanto, a existência de muitos **atributos redundantes** pode causar problemas;
- Muitos atributos numéricos **não seguem uma distribuição *Guassiana*** (→ *kernel density estimators*).

Naive Bayes Wrapper (NBW):

- Atributos irrelevantes e redundantes podem comprometer acurácia de classificação;
- Selecionar atributos com base no desempenho do classificador NB. Informalmente pode-se sumarizar o NBW como segue:
  - Construir um classificador NB para cada atributo Xi (i = 1,..., n). Escolher Xi para o qual o NB apresenta a melhor acurácia e inseri-lo em AS = {atributos selecionados};
  - Para todo Xi $\notin$ AS construir um NB formado por {Xi } $\cup$ AS. Escolher o melhor classificador dentre os disponíveis e verificar se é melhor do que o obtido anteriormente:
    - SE sim, ENTÃO atualizar AS, inserindo o atributo adicional e repetindo o passo 2);
    - SE não, ENTÃO parar e usar o classificador obtido anteriormente.

Naive Bayes Wrapper (NBW):

- Atributos irrelevantes e redundantes podem comprometer acurácia de classificação;
- Selecionar atributos com base no desempenho do classificador NB. Informalmente pode-se sumarizar o NBW como segue:
  - Construir um classificador NB para cada atributo Xi (i = 1,..., n). Escolher Xi para o qual o NB apresenta a melhor acurácia e inseri-lo em AS = {atributos selecionados};
  - Para todo Xi $\notin$ AS construir um NB formado por {Xi } $\cup$ AS. Escolher o melhor classificador dentre os disponíveis e verificar se é melhor do que o obtido anteriormente:

Este arcabouço pode ser aplicado com todos os classificadores, mas pela simplicidade e rapidez do Naïve Bayes, o NBW é uma boa escolha.

Redes Bayesianas

Cap 14
Seções 1 – 3

Redes Bayesianas

Estrutura de dados para representar as dependências entre variáveis e fornecer uma especificação concisa de *qualquer* distribuição de probabilidade conjunta total.

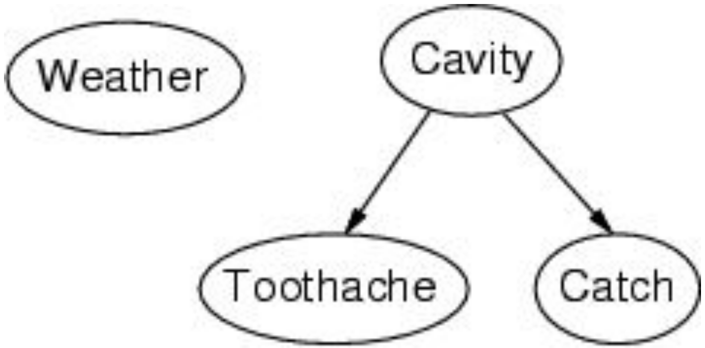
Sintaxe:

- um conjunto de nós, um para cada variável aleatória
- grafo direcionado e acíclico (seta = ”influência direta")
- cada nó tem uma distribuição condicional $\mathbf{P}(X_i \mid Pais\ (X_i))$ que quantifica o efeito dos pais sobre o nó

No caso mais simples, a distribuição condicional é representada como uma tabela de probabilidade condicional (TPC) dada uma distribuição sobre X_i para cada combinação de valores dos pais.

Exemplo

A topologia de uma rede representa relações de independência condicional :



Clima é independente de outras variáveis
Toothache e *Catch* são condicionalmente independentes dado *Cavity*

Exemplo

“ *I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar*? ”

Variáveis?

A topologia da rede?

Exemplo

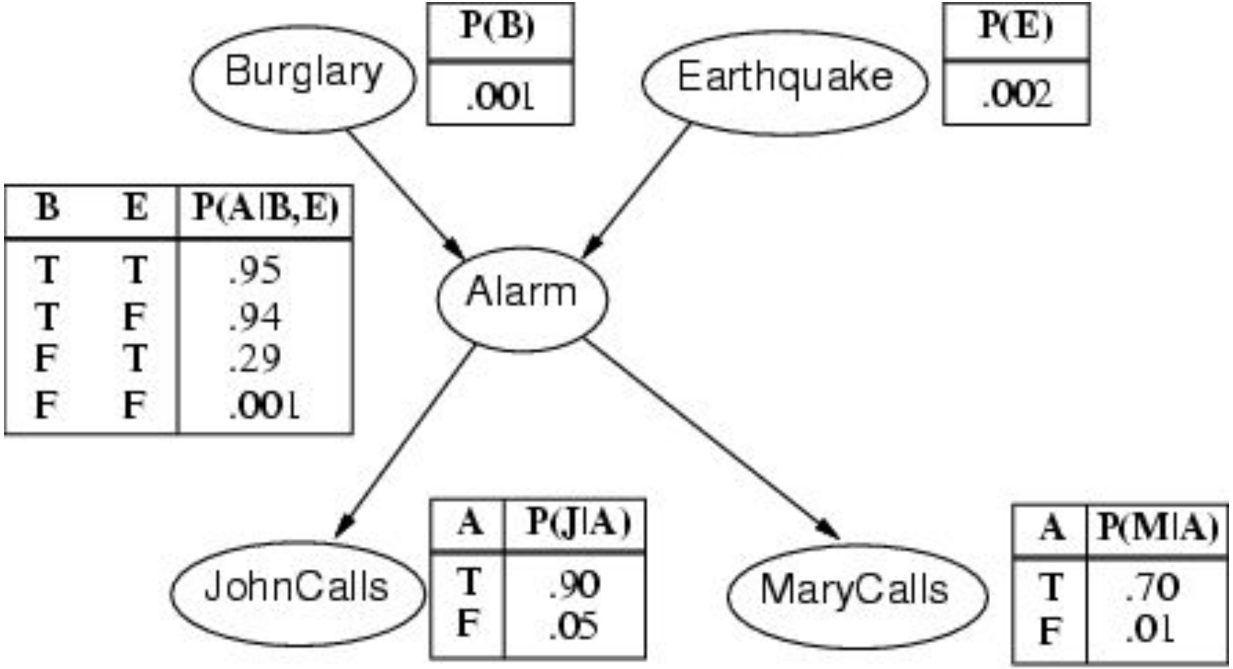
“ *I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar*? ”

Variáveis: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

A topologia da rede reflete conhecimento "causal":

- Um roubo (*burglar) pode ligar o alarme*
- Um terremoto (*earthquake*) pode ligar o alarme
- O alarme faz Mary telefonar
- O alarme faz John telefonar

Exemplo



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Da topologia da rede

Roubos e terremotos afetam diretamente a probabilidade do alarme tocar;

Mas o fato de Joao e Maria telefonarem só depende do alarme;

Desse modo, **a rede representa nossas suposições de que eles não percebem quaisquer roubos diretamente, não notam os terremotos e não verificam antes de ligar!**

As probabilidades...

... resumem um conjunto potencialmente infinito de circunstâncias (Maria ouve música alta, João liga qdo toca o telefone; umidade, falta de energia, etc podem interferir no alarme; Joao e maria não estão em casa, etc.

Preguiça e ignorância

Construindo uma rede Bayesiana

1. Escolher uma ordem para as variáveis aleatórias $X_1, \ldots, X_n$
2. Para i = 1 à n
   adicione X_i à rede
   selecione pais para $X_1, \ldots, X_{i-1}$ tais que
   $$\mathbf{P}(X_i \mid Pais(X_i)) = \mathbf{P}(X_i \mid X_1, \ldots X_{i-1})$$

Esta escolha de pais garante a semântica global:

$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})$ (regra da cadeia)

$= \prod_{i=1} \mathbf{P}(X_i \mid Pais(X_i))$ (por construção)

Ordem para as variáveis

A ordem correta em que os nós devem ser adicionados consiste em adicionar primeiro as “causas de raiz”, depois as variáveis que elas influenciam e assim por diante, até chegarmos às folhas, que não tem nenhuma influência causal direta sobre as outras variáveis.

E se escolhermos a ordem “errada”??

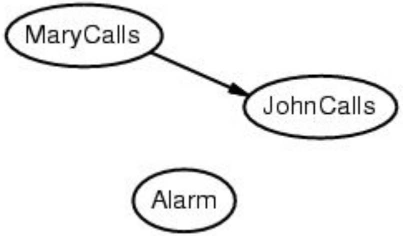
Exemplo

Assumindo a ordem: *M, J, A, B, E*

$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

Exemplo

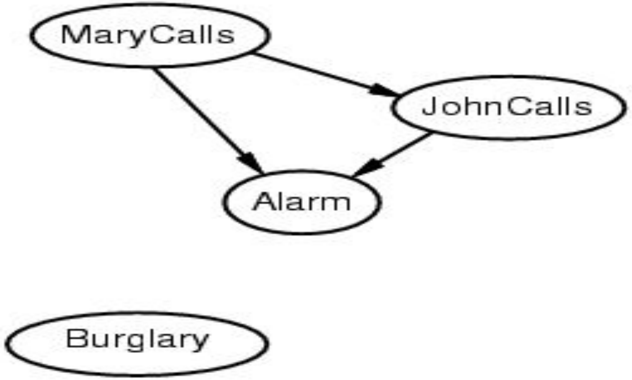
Assumindo a ordem: *M, J, A, B, E*



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**
$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$?

Exemplo

Assumindo a ordem: *M, J, A, B, E*



$\boldsymbol{P}(J \mid M) = \boldsymbol{P}(J)$? **No**

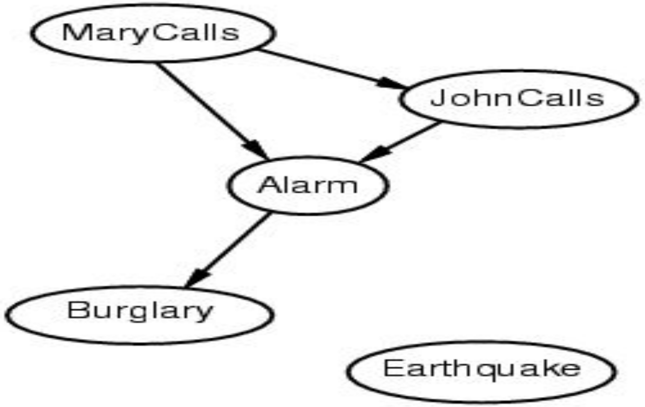
$\boldsymbol{P}(A \mid J, M) = \boldsymbol{P}(A \mid J)$? $\boldsymbol{P}(A \mid J, M) = \boldsymbol{P}(A)$? **No**

$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B \mid A)$?

$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B)$?

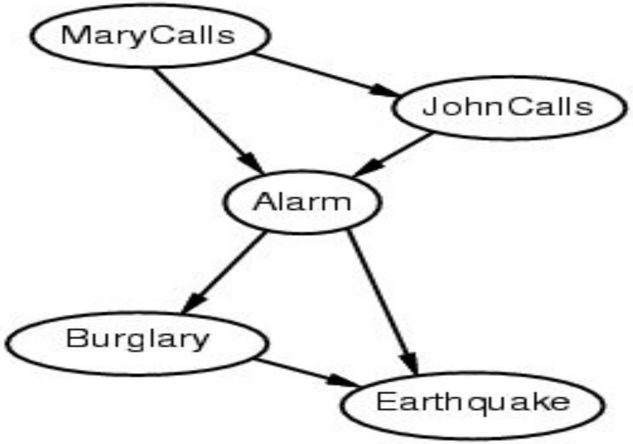
Exemplo

Assumindo a ordem: *M, J, A, B, E*



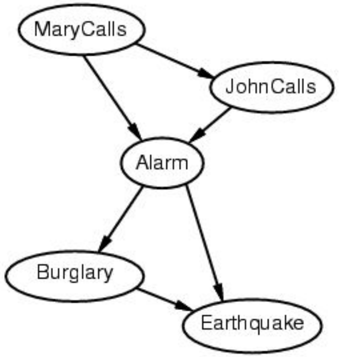
$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**
$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**
$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$? **Yes**
$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$? **No**
$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A)$?
$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A, B)$?

Exemplo



P(*J* | *M*) = ***P***(*J*)? **No**
P(*A* | *J*, *M*) = ***P***(*A* | *J*)? ***P***(*A* | *J*, *M*) = ***P***(*A*)? **No**
P(*B* | *A*, *J*, *M*) = ***P***(*B* | *A*)? **Yes**
P(*B* | *A*, *J*, *M*) = ***P***(*B*)? **No**
P(*E* | *B*, *A* ,*J*, *M*) = ***P***(*E* | *A*)? **No**
P(*E* | *B*, *A*, *J*, *M*) = ***P***(*E* | *A*, *B*)? **Yes**

Exemplo

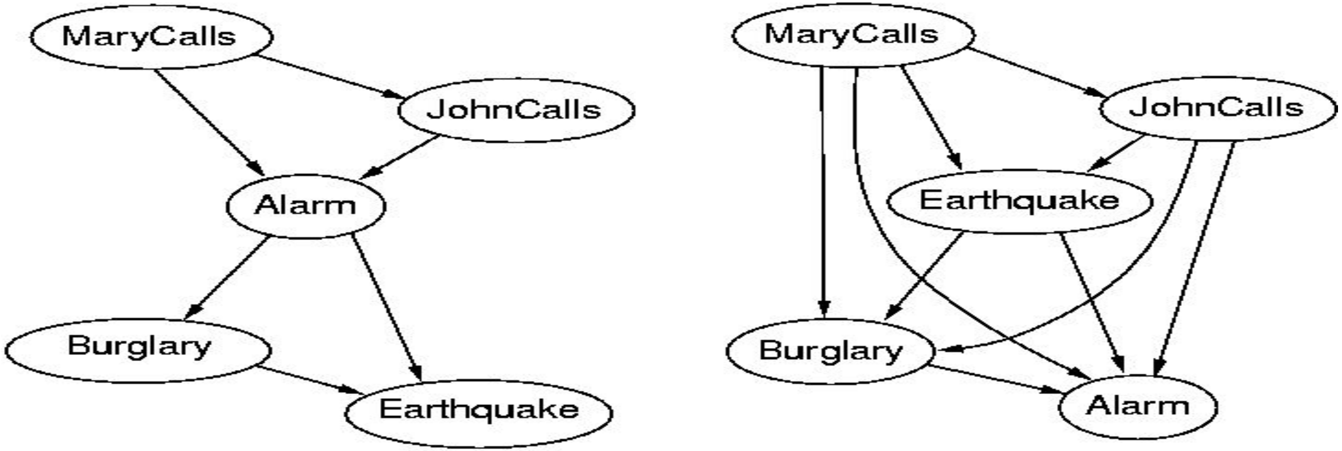


A rede resultante terá **dois vínculos a mais** que a rede original e exigirá outras probabilidades para serem especificadas

Alguns dos vínculos apresentam **relacionamentos tênues** que exigem julgamentos de probabilidade difíceis e antinaturais (prob de *Terremoto*, dados *Roubo* e *Alarme*)

(Em geral) é melhor pensar de *causas* para *efeitos* (modelo causal) e não do contrário (modelo de diagnóstico)

Exemplo



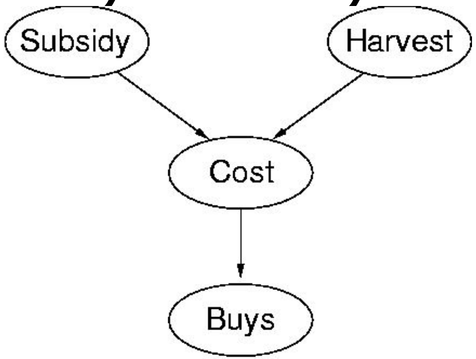
(a) (b)

Uma ordenação de nós ruim: *MarryCalls*, *JohnCalls*, *Earthquake*, *Burglary* e *Alarm*

Entretanto, todas as três redes devem representar a mesma distribuição conjunta. As duas últimas só não expressam todas as independências condicionais

Redes de Bayes Híbridas

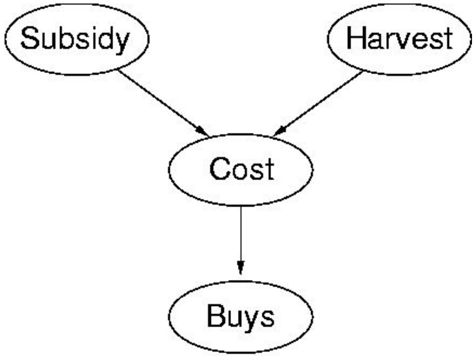
Discretas: *Subsidy?* e *Buys?*



Dois novos tipos de distr. condicionais:

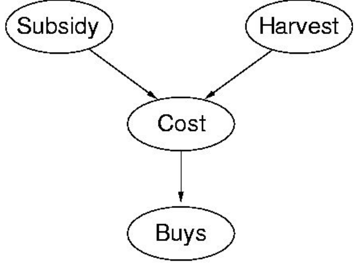
- **variável contínua, com pais contínuos e discretos *(Cost)***
- **Variável discreta com pais contínuos *(Buys?)***

Redes de Bayes Hibridas



Manipular variáveis contínuas:
- **Discretização**: repartir os valores possíveis em um conjunto fixo de intervalos
- **Definir funções de probabilidade padrão** especificadas por um número finito de parâmetros

variável contínua, com pais contínuos e discretos *(Cost)*



Para custo: **P**(Custo|Colheita, Subsídio)

O pai discreto (Subsídio) é manipulado por enumeração explícita:

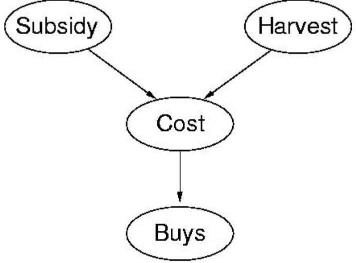
P(Custo|Colheita, subsídio) e *P(Custo|Colheita, ¬subsídio)*

Para *Colheita* especificamos como a distribuição sobre o custo **c** depende do valor contínuo **h** de colheita.

I.e., os parâmetros da distribuição de custo como função de **h**

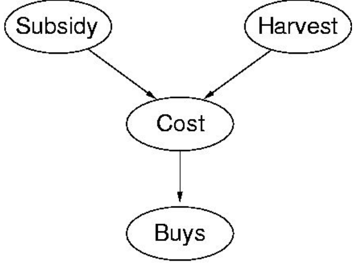
em geral: **distribuição Gaussiana linear**

variável contínua, com pais contínuos e discretos *(Cost)*



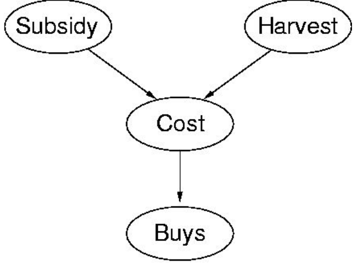
distribuição gaussiana linear: o filho (*Cost*) tem uma distribuição gaussiana cuja média varia linearmente com o valor do pai (*Harvest*), e cujo desvio padrão é fixo:

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



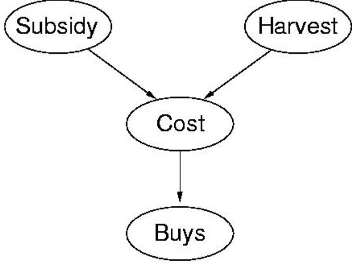
distribuição gaussiana linear: o filho (*Cost*) tem uma distribuição gaussiana cuja <u>média varia linearmente com o valor do pai (*Harvest*),</u> e cujo desvio padrão é fixo:

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

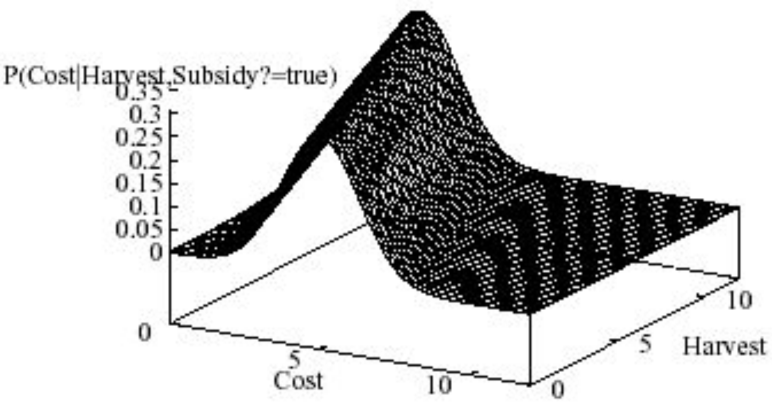


distribuição gaussiana linear: o filho (*Cost*) tem uma distribuição gaussiana cuja média varia linearmente com o valor do pai (*Harvest*), e <u>cujo desvio padrão é fixo:</u>

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



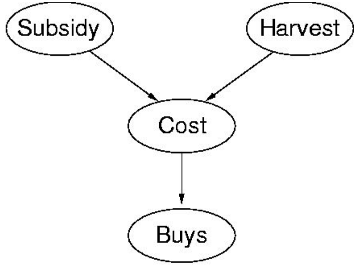
$$P(Cost = c | Harvest = h, Subsidy? = true)$$
$$= N(a_t h + b_t, \sigma_t)(c)$$
$$= \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{c - (a_t h + b_t)}{\sigma_t}\right)^2\right)$$



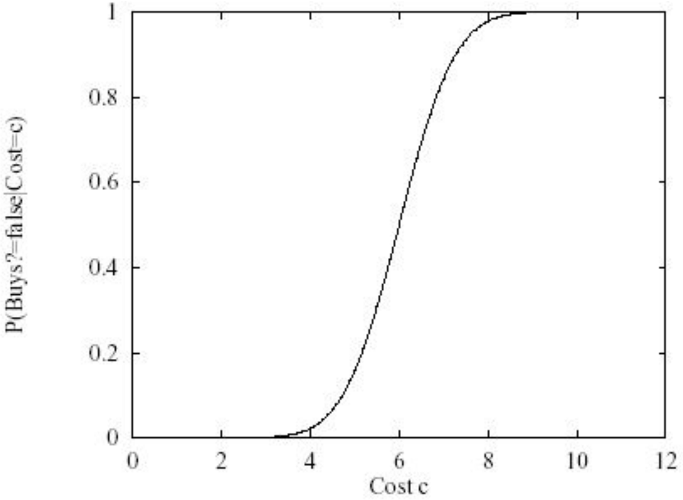
A inclinação é negativa, porque o preço diminui à medida que a quantidade oferecida aumenta

variáveis discretas com pais contínuos

- Ex. *Compras*:
  - Podemos supor que o cliente comprará se o preço for baixo e não comprará se for alto e que:
  - A probabilidade de compra varia suavemente em alguma região intermediária
    - A distribuição condicional é semelhante a uma função de limiar “suave” (*soft threshold*)
    - Distribuição **probit** é uma possibilidade...



v.discretas, pais contínuos



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

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

- Probabilidade de *Compra (buys)* dado *Custo (Cost)*: ***limiar suave***
- Distribuição **Probit (probability unit)**:
  - integral da Gaussiana
  - Limites do Custo ocorrem em volta de μ, a amplitude da região dos limites é proporcional a σ, e a probabilidade de comprar diminui a medida que o custo aumenta

Inferência

Computar a probabilidade de uma consulta dado um evento observado

$P(A_1,A_2,\ldots,A_n) = \Pi_{i=1}^{n} P(A_i|parents(A_i))$

Variáveis de evidência observadas

$\mathbf{E} = E_1,\ldots,E_m$

Variável de consulta X

Sem evidência (escondidas)

$\mathbf{Y} = Y_1\ldots Y_l$

Inferência

$P(X|\mathbf{E})$

Teorema de Bayes:
$P(A|B) = P(A,B) / P(B)$

Revisando a regra da cadeia:
$P(A_1,A_2,\dots,A_n) = \Pi_{i=1}^{n} P(A_i|parents(A_i))$

Inferência

Marginalização:

$P(A_i) = \sum_i P(A_i,B_i)$

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6 \,.$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4 \,.$$

$$\begin{aligned}
&\mathbf{P}(Cavity \mid toothache) = \alpha\, \mathbf{P}(Cavity, toothache) \\
&= \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\
&= \alpha\, [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] = \alpha\, \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle \,.
\end{aligned}$$

Inferência

$P(X|\mathbf{E}) = \alpha\, P(X,\mathbf{E})$

Marginalização:

$P(A_i) = \sum_j P(A_i,B_j)$

$P(X|\mathbf{E}) = \alpha \sum_{\mathbf{Y}} P(X,\mathbf{E},\mathbf{Y})$

Inferência

$P(X|\mathbf{E}) = \alpha\, P(X,\mathbf{E})$

Marginalização:

$P(A_i) = \sum_j P(A_i,B_j)$

$P(X|\mathbf{E}) = \alpha \sum_{\mathbf{Y}} P(X,\mathbf{E},\mathbf{Y})$

$P(X|\mathbf{E}) = \alpha \sum_{\mathbf{Y}} \Pi_{A=X}\, P(A|parents(A))$

Inferência - Exemplo

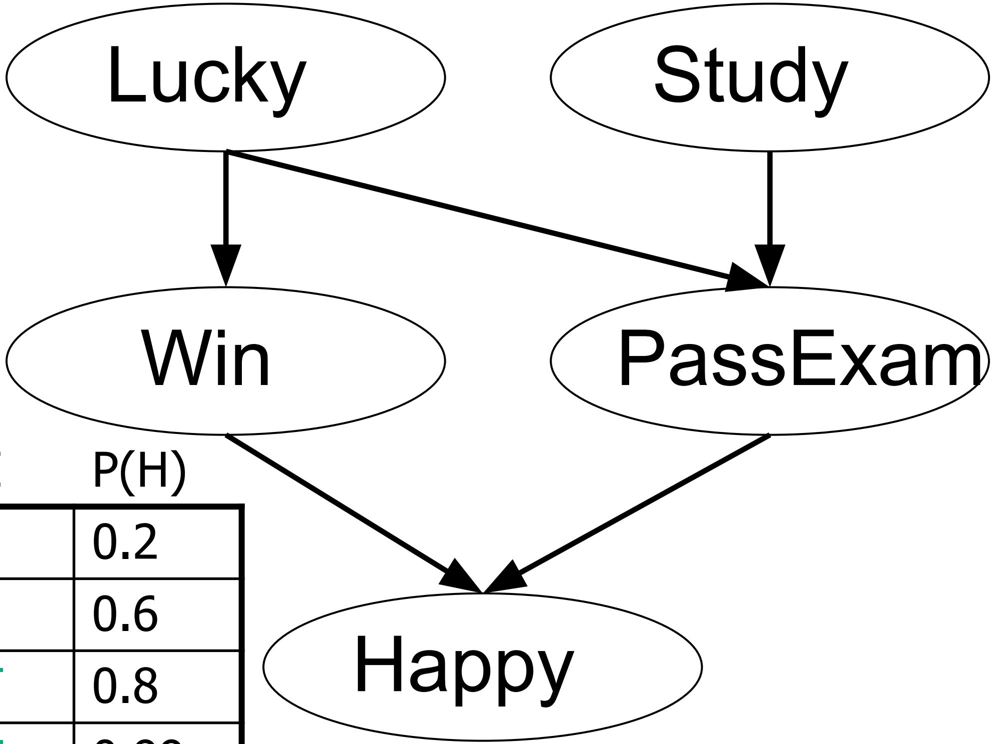
Problema: Bob pode:

- estar ou não feliz
- ter ou não ter passado no exame de cálculo
- ter ou não ter ganho na loteria
- ser ou não sortudo
- ter ou não ter estudado

Inference Example

P(L) = 0.75

P(S) = 0.2



L	P(W)
F	0.01
T	0.4

L	S	P(E)
F	F	0.01
T	F	0.5
F	T	0.9
T	T	0.99

W	E	P(H)
F	F	0.2
T	F	0.6
F	T	0.8
T	T	0.99

Exemplo 1

Com a informação da tabela apenas, qual a probabilidade de Bob ter ganho a loteria?

$P(W) = \sum_l P(W,l)$

$P(W) = \sum_l P(W|l)P(l)$

$P(W) = P(W|L)P(L) + P(W|\neg L)P(\neg L)$

$P(W) = 0.4*0.75 + 0.01*0.25$

$P(W) = 0.3025$

Exemplo 2

Dado que Bob está feliz, qual é a probabilidade de que ele tenha ganho a loteria?

Da rede sabemos que:

$P(h,e,w,s,l) = P(l)P(s)P(e|l,s)P(w|l)P(h|w,e)$

Queremos encontrar

$P(W|H) = \alpha \sum_l \sum_s \sum_e$

$P(l)P(s)P(e|l,s)P(W|l)P(H|W,e)$

$P(\neg W|H)$

Exemplo 2

l	s	e	P(s)	P(l)	P(e|l,s)	P(W|l)	P(H|W,e)	
F	F	F	0.8	0.25	0.99	0.01	0.6	0.001188
T	F	F	0.8	0.75	0.5	0.4	0.6	0.072
F	T	F	0.2	0.25	0.1	0.01	0.6	0.00003
T	T	F	0.2	0.75	0.01	0.4	0.6	0.00036
F	F	T	0.8	0.25	0.01	0.01	0.99	0.0000198
T	F	T	0.8	0.75	0.5	0.4	0.99	0.1188
F	T	T	0.2	0.25	0.9	0.01	0.99	0.0004455
T	T	T	0.2	0.75	0.99	0.4	0.99	0.058806

$P(W|H) = \alpha\ 0.2516493$

Exemplo 2

l	s	e	P(s)	P(l)	P(e|l,s)	P(¬W|l)	P(H|¬W,e)	
F	F	F	0.8	0.25	0.99	0.99	0.2	0.039204
T	F	F	0.8	0.75	0.5	0.6	0.2	0.036
F	T	F	0.2	0.25	0.1	0.99	0.2	0.00099
T	T	F	0.2	0.75	0.01	0.6	0.2	0.00018
F	F	T	0.8	0.25	0.01	0.99	0.8	0.001584
T	F	T	0.8	0.75	0.5	0.6	0.8	0.144
F	T	T	0.2	0.25	0.9	0.99	0.8	0.03564
T	T	T	0.2	0.75	0.99	0.6	0.8	0.07128

$$P(\neg W|H) = \alpha\, 0.328878$$

Exemplo 2

$P(W|H) = \alpha <0.2516493, 0.328878>$

$P(W|H) = <0.4335, 0.5665>$

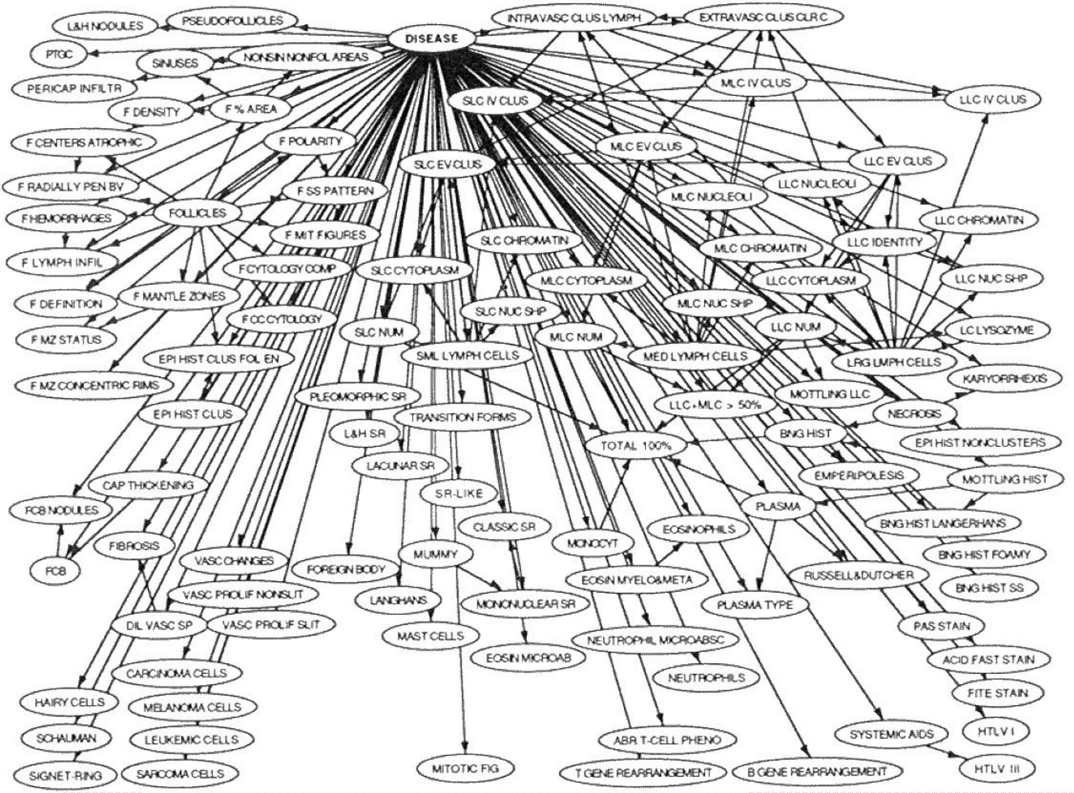
Sistemas especialistas

- Redes bayesianas para implementar sistemas especialistas
  - Sistemas de diagnóstico
  - Conhecimento (nós, relaçoes, probabilidades) tipicamente fornecidos por humanos
  - Sistema observa evidências peguntando ao usuário

Pathfinder

- Doenças do linfonodo
  - Mais de 60 doenças
  - Mais de 100 características de linfonodos
  - Mais de 30 características de informação clínica
- Muito trabalho realizado por especialistas
  - 8 horas para definir características e doenças
  - 35 horas para construir a topologia
  - 40 horas para avaliar as probabilidades

Pathfinder



- Testing the network
  - 53 casos de teste (diagnosticos reais)
  - Precisão tão boa quanto a dos especialistas

Resumo

Redes Bayesianas são representações explícitas de independência condicional

Topologia + TPCs = representações compactas de distribuições conjuntas totais

Ferramentas poderosas para construir uma representação de um domínio que envolva incerteza.

Bibliografia

- G. Bittencourt
  - Capítulo 2
- S. O. Rezende
  - Capítulos 1 e 2
- Slides
  - Ana Carolina Lorena, Unifesp
  - Ronaldo Prati, UFABC
  - Richard Khoury, University of Waterloo
  - Profa Anne Magaly Canuto, UFRN
  - Profa Josiane
  - Profa Solange O. Rezende, ICMC-USP
  - Prof Guilherme Bittencourt, UFSC
  - Prof Ricardo Campello, ICMC-USP