Conhecimento e Raciocínio

Aula 6

Redes Bayesianas

Viriato M. Marques

Licenciatura em Engenharia Informática

DEIS – Departamento de Engenharia Informática e de Sistemas

ISEC – Instituto Superior de Engenharia de Coimbra

6. Redes Bayesianas

1. Introdução

- *O mundo não é preto e branco…é cinzento* ☺
- A incerteza está presente em muitos aspectos da realidade
- Em muitas tarefas o desfecho não pode ser previsto com 100% de certeza …
  - Diagnóstico médico e de falhas
  - Previsão do comportamento de um cliente
  - Qual o risco de ataque cardíaco para uma pessoa que faz boa alimentação, pratica desporto e não fuma? É certo que não o terá?
- A modelação da incerteza pode basear-se em probabilidade, conjuntos difusos, factores de certeza, crença, etc.
- Vamos tratar de um classificador probabilístico designado por Rede Bayesiana

6. Redes Bayesianas

2. Probabilidade - Revisão

- Ideia base de um classificador Bayesiano:
  - Estimar a probabilidade de uma dada classificação ou **hipótese** H, conhecidos os valores dos atributos de um novo exemplo
  - Formalmente, calcular: $p(H_i| \mathbf{X}_{ij})$, com H_i = Hipótese de ordem i, $\mathbf{X}_j$ = vector de atributos $(x_{i1}, x_{i2}\ldots x_{in})$ e $p(H|X)$ = probabilidade *a posteriori* ou probabilidade condicionada (p de H suposto **X**)
- Para isso bastaria:
  - Dispor de um dataset de treino composto pelas diversas combinações possíveis de **todos** os atributos com **todas** as classes, i.e.:
  - Para cada combinação possível dos atributos, contar as ocorrências de cada classe, H_i e as respectivas ocorrências da combinação $\mathbf{X}_{ij}$
  - Calcular as frequências relativas H_i/X_{ij}

6. Redes Bayesianas

	Atributos			Classe
ID	Febre	Tosse	Dores	Diag
1	B	B	B	Tuberculose
2	B	B	A	Gripe
3	B	A	B	Tuberculose
4	B	A	A	Gripe
5	A	B	B	Gripe
6	A	B	A	Gripe
7	A	A	B	Gripe
8	A	A	A	Gripe
9	B	B	B	Tuberculose
10	B	B	A	Gripe
11	B	A	B	Gripe
12	B	A	A	Gripe
13	A	B	B	Tuberculose
14	A	B	A	Gripe
15	A	A	B	Gripe
16	A	A	A	Gripe

- Fórmula de Bayes
  - Relaciona probabilidade *a posteriori* com probabilidade *a priori*

$$p(H \mid X) = \frac{p(X \mid H).p(H)}{p(X)}$$

a priori p(Febre=B|Tuberculose)=#{1,3,9}/#{1,3,9,13}=3/4=0,75

a posteriori p(Tuberculose|Febre=B)=#{1,3,9}/#{1,2,3,4,9,10,11,12}=3/8=0,375

Fórmula de Bayes p(B|A)=p(A|B)*p(B)/p(A)

p(Tuberculose|Febre=B)=p(Febre=B|Tuberculose)*p(Tuberculose)/p(Febre=Baixa)

p(Tuberculose|Febre=B)=0,75*(4/16)/(8/16)=0,375

6. Redes Bayesianas

	Atributos			Classe
ID	**Febre**	**Tosse**	**Dores**	**Diagnóstico**
1	B	B	B	Tuberculose
2	B	B	A	Gripe
3	B	A	B	Tuberculose
4	B	A	A	Gripe
5	A	B	B	Gripe
6	A	B	A	Gripe
7	A	A	B	Gripe
8	A	A	A	Gripe
9	B	B	B	Tuberculose
10	B	B	A	Gripe
11	B	A	B	Gripe
12	B	A	A	Gripe
13	A	B	B	Tuberculose
14	A	B	A	Gripe
15	A	A	B	Gripe
16	A	A	A	Gripe
17	B	B	A	Gripe
18	B	A	B	Gripe
19	B	A	A	Gripe
20	A	B	B	Tuberculose
21	A	B	A	Gripe
22	A	A	B	Gripe
23	A	A	A	Gripe

- Exemplo igual ao anterior mas com mais tuplos

Para 1 Atributo		
a priori	p(Febre=B|Tuberculose)=3/5=	0,6
a posteriori	p(Tuberculose|Febre=B)=3/11=	0,272727
Fórmula de Bayes	p(B|A)=p(A|B)*p(B)/p(A)	
	p(Tuberculose|Febre=B)=p(Febre=B|Tuberculose)*p(Tuberculose)/p(Febre=Baixa)	
	p(Tuberculose|Febre=B)=0,6*(5/23)/(11/23)=	0,272727
Para os 3 Atributos		
a priori	p(B,B,B|Tuberculose)=2/5=	0,4
a posteriori	p(Tuberculose|B,B,B)=2/2=	1
Fórmula de Bayes	p(B|**A**)=p(**A**|B)*p(B)/p(**A**)	
	p(Tuberculose|B,B,B)=p(B,B,B|Tuberculose)*p(Tuberculose)/p(B,B,B)	
	p(Tuberculose|B,B,B)=0,4*(5/23)/(2/23)=	1

6. Redes Bayesianas

- O exemplo anterior mostra que
  - Para calcular directamente P(H|X) das tabelas, é necessário conhecer um número enorme de exemplos, dado que têm de se considerar:
  - Todas as combinações possíveis dos N atributos
  - Multiplicar essas combinações por todas as classes
  - Ter um número “elevado” de instâncias de cada combinação de atributos / classe para se poder aproximar a probabilidade através das frequências relativas calculadas
- Exemplo:
  - 5 atributos, cada um com 3 valores, 4 classes: $3^5*4=972$. Se se pretendesse uma média de 100 exemplos por cada combinação de atributos / classe, ter-se-ia necessidade de cerca de 972 * 100 = 97200 exemplos !

6. Redes Bayesianas

- Em termos práticos é pois necessário procurar uma abordagem que permita reduzir o número de exemplos de treino necessários:
  - Pode conseguir-se partindo da Fórmula de Bayes, já apresentada:

$$\boxed{p(H \mid X) = \frac{p(X \mid H).p(H)}{p(X)}}$$

  - Contudo, aparentemente nada se ganha ao exprimir p(H|X) em função de p(X|H): o número de exemplos necessário seria do mesmo modo muito elevado!
  - E com a agravante de terem de se calcular ainda p(H) e p(X)

6. Redes Bayesianas

Independência Condicional

- Exemplo: a habilidade de leitura e o comprimento do braço são variáveis aleatórias condicionalmente independentes. Porquê?
  - Estatisticamente poderá verificar-se que quanto maior o comprimento do braço, C, maior a habilidade de leitura, L
  - Contudo, esta relação é explicada atendendo à idade, I: quanto maior a idade, maior o comprimento do braço e a habilidade de leitura. Fixando I, a correlação entre L e C desaparece
  - Ou seja, L é afinal condicionalmente **dependente** de I
  - E, além disso, L é condicionalmente **independente** de C
- Formalmente, sendo X, Y e Z variáveis aleatórias, X é independente de Y se

$$\boxed{p(X \mid Y, Z) = p(X \mid Z)}$$

6. Redes Bayesianas

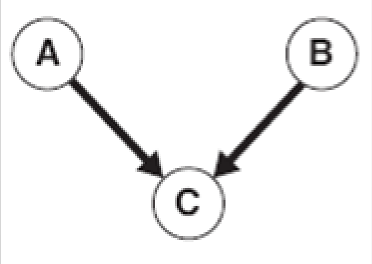
3. Redes Bayesianas

- Este classificador não exige a independência condicional de todos os atributos dentro de uma mesma classe
- Pelo contrário, este modelo permite traduzir dependências condicionais, mas apenas as "importantes"
  - Situa-se assim entre as exigências da Fórmula de Bayes original (dependência condicional de todos os atributos, da classe) e as do classificador Naïve Bayes (independência condicional de todos os atributos, da classe)
- O aspecto geral de uma rede Bayesiana é o de um grafo acíclico e dirigido
  - As dependências são traduzidas por setas, entre atributos
  - Os atributos constituem nós e contêm tabelas de probabilidade condicionada que os associam ao(s) nó(s) pai
  - O grafo não pode conter quaisquer ciclos fechados

6. Redes Bayesianas

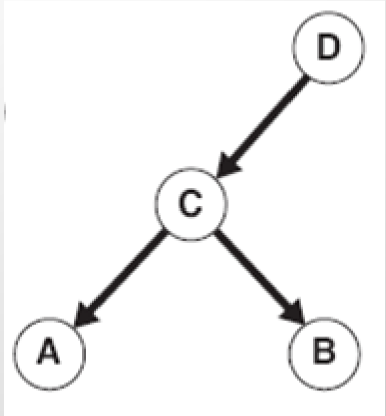
Modelo

- Sejam 3 variáveis aleatórias A, B e C
  - Se A for independente de B, e A e B influenciarem, ambas, C, a rede correspondente terá o seguinte aspecto:



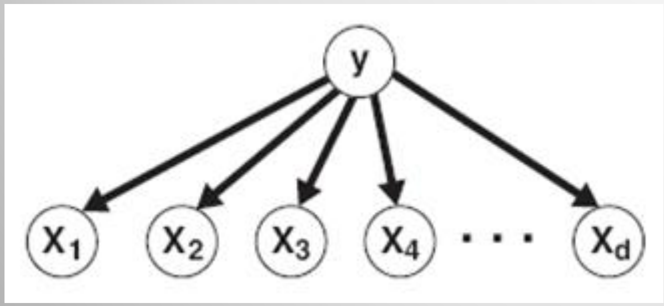
  - A e B são **pais** de C
  - C é **filho** de A e B

- Na rede seguinte:
  - D é um **antepassado** de B
  - B é um **descendente** de D



6. Redes Bayesianas

- A cada nó são assinaladas as seguintes tabelas de probabilidade:
  - Se um nó não tem pais, a tabela é do tipo p(X)
  - Se um nó tem apenas 1 pai, a tabela é do tipo p(X|Y)
  - Se um nó tem vários pais, a tabela é do tipo $p(X|Y_1, Y_2 \ldots Y_n)$
- De acordo com este modelo, a independência condicional entre atributos feita pelo Naïve Bayes Classifier pode ser assim visualizada:



- Cada nó X_i terá apenas uma tabela de probabilidade condicionada do tipo $p(X_i|Y)$, dado que cada atributo depende apenas de Y
- Foram exactamente estas frequências que se calcularam anteriormente e que figuram no produto $\prod_{i=1}^{d} P(X_i \mid Y = y)$

6. Redes Bayesianas

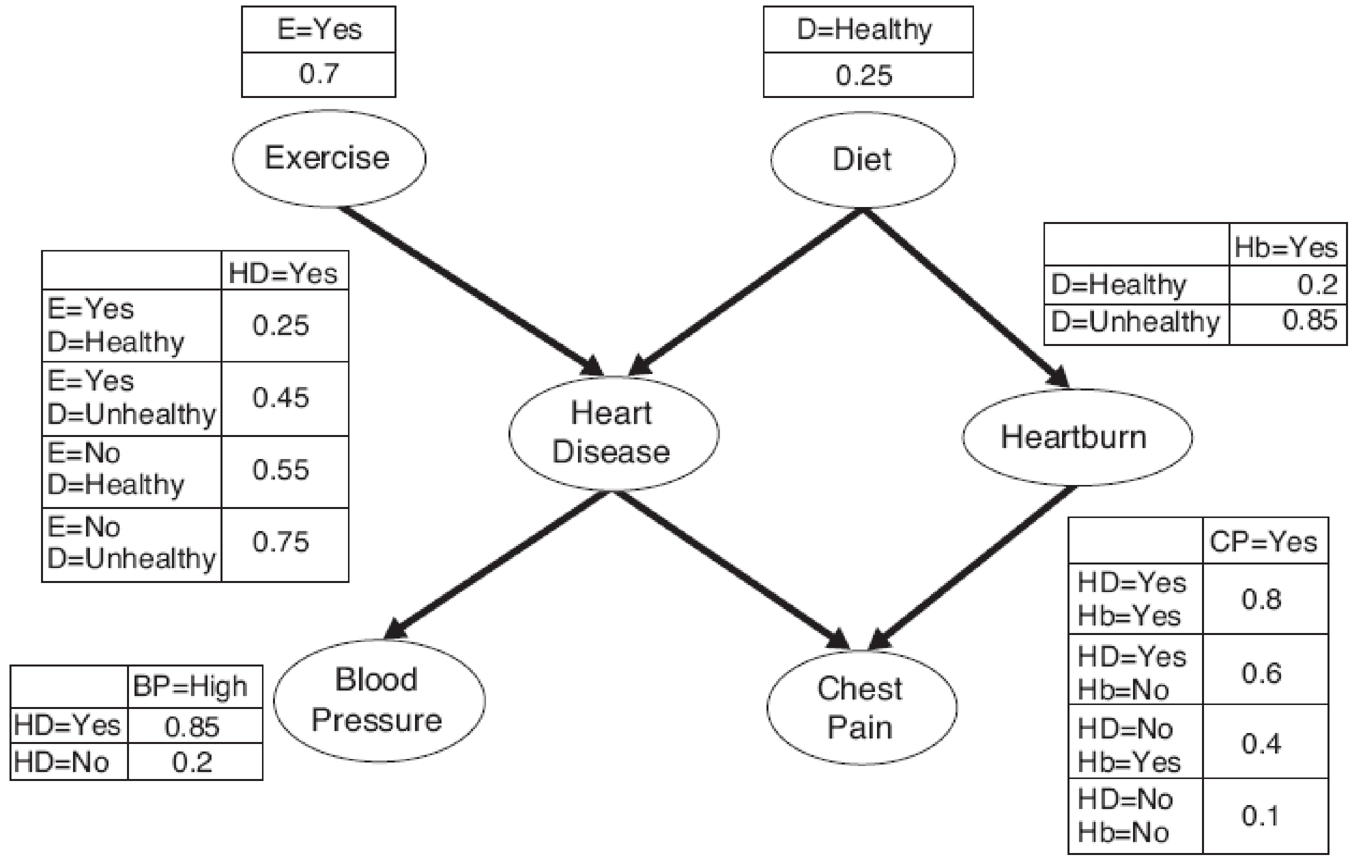


Figure 5.13. A Bayesian belief network for detecting heart disease and heartburn in patients.

6. Redes Bayesianas

- Desenhar uma Rede Bayesiana envolve 2 fases:
  - Desenhar a rede
  - Estimar as probabilidades condicionais envolvidas (através de históricos ou opinião de peritos)
- Para desenhar a rede há duas opções
  1. Captar as relações de causa-efeito e traçar as setas das causas para os efeitos
  2. Proceder sistematicamente (ver rede anterior):
     - *Ordenar as variáveis. Por exemplo: E,D,HD,Hb,CP,BP*
     - *P(D|E) pode simplificar-se para p(D)*
     - *P(HD|E,D) não pode simplificar-se*
     - *P(Hb|HD,E,D) pode simplificar-se para p(Hb|D)*
     - *P(CP|Hb,HD,E,D) pode simplificar-se para p(CP|Hb,HD)*
     - *P(BP|CP,Hb,HD,E,D) pode simplificar-se para p(BP|HD)*
  - Esta abordagem pode gerar modelos diferentes consoante a ordenação original das variáveis, mas garante ausência de ciclos
  - Dificuldade: para *d* atributos há *d!* grafos distintos

6. Redes Bayesianas

Cálculos com Redes Bayesianas

- Uma Rede Bayesiana permite 3 tipos de cálculos base:
  1. Probabilidade conjunta de qualquer acontecimento

$$\boxed{P(x_1, x_2 \ldots x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i))}$$

  2. Inferência:
     a. Causal (das causas para os efeitos)
     b. Diagnóstico (dos efeitos para as causas)

e $$\boxed{p(x_1 \mid x_2 \ldots x_n) = \frac{p(x_1, x_2 \ldots x_n)}{p(x_2 \ldots x_n)}} \longrightarrow \left[\begin{array}{l} p(A \mid B) = \dfrac{p(A, B)}{p(B)} \\ p(A \mid B, C, D) = \dfrac{p(A, B, C, D)}{p(B, C, D)} \end{array}\right.$$

- Existem ainda outros tipos de cálculos baseados nestes, nomeadamente Intercausal (estudo da influência de novas causas sobre um efeito comum) e Misto (mistura de várias destas formas num só cálculo)

6. Redes Bayesianas

- Note-se que, por exemplo

$$p(A|B,C,D)=\frac{p(A,B,C,D)}{p(B,C,D)}=\frac{p(A,B,C,D)}{p\left(\bar{A}B,C,D\right)+p(A,B,C,D)}$$

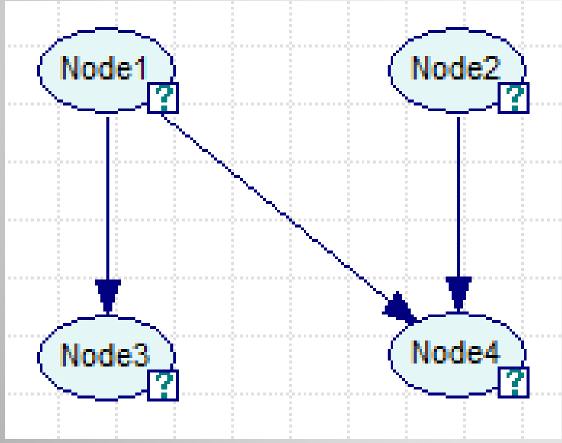
- Isto significa que, conhecidas as probabilidades conjuntas relativas a uma rede, é pode aplicar-se a equação acima para realizar uma inferência de diagnóstico
- Por outro lado, como vimos, para calcular estas probabilidades conjuntas basta aplicar

$$P(x_1, x_2 ... x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i))$$

- que permite calculá-las a partir das probabilidades condicionadas que figuram nos nós da rede

6. Redes Bayesianas

- Uma rede Bayesiana permite:
  - Reduzir muito o número de probabilidades que é necessário conhecer inicialmente para implementar a rede
  - Que essas probabilidades sejam do tipo condicionado e não conjuntas (importante no caso de a rede ser modelizada por recurso a opinião de peritos e não por análise de dados)
- Exemplo



Supondo que as 4 variáveis são binárias, teremos de conhecer, para implementar a rede, 1+1+2+4=8 probab. condicionadas

O número de probabilidades conjuntas é, no entanto, $2^4=16$ (correspondente às combinações de 0000 a 1111) ou seja, o dobro das condicionadas atrás referidas

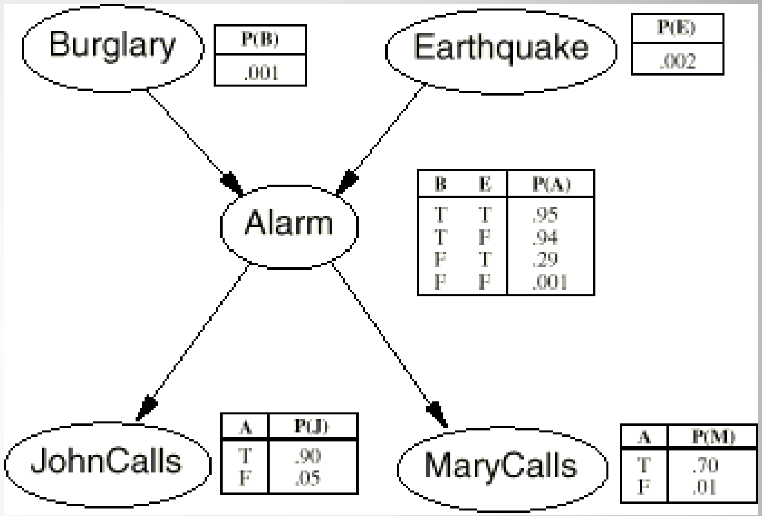
6. Redes Bayesianas

- Suponha-se que temos 20 variáveis booleanas tais que cada uma é influenciada, no máximo, por 5 outras:
  - Rede Bayesiana:
    - Para cada nó, a tabela conterá as combinações binárias de 5 variáveis, num total de 25=32 linhas.
    - Como são 20 nós, a rede necessitará da especificação de 20x32=640 probabilidades.
  - Probabilidade Conjunta:
    - Uma tabela de probabilidade conjunta teria de conter 2^{20}= 1.048.576 probabilidades
- Ora, para realizar inferências com redes Bayesianas estas probabilidades conjuntas
  - Podem ser calculadas somente à medida que são necessárias
  - O seu valor é determinado a partir das condicionadas

6. Redes Bayesianas

- **Exemplo 1:** Cálculo de Probabilidades Conjuntas

Uma casa possui um alarme que toca quando há assaltos mas, por vezes, também quando há um tremor de terra. Quando toca, os vizinhos João e Maria telefonam ao dono, segundo as probabilidades assinaladas na rede Bayesiana da figura. Calcular a probabilidade de John e Mary telefonarem ambos, o alarme tocar e não ocorrer nenhum roubo nem tremor de terra:



$$P(x_1, x_2 \ldots x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i))$$

P(J , M , A , ¬B , ¬E)=

= P(J|A) x P(M|A) x P(A |¬B ,¬E) x P(¬B) x P(¬E) =

= 0.90 x 0.70 x 0.001 x 0.999 x 0.998 =

= 0.00062

6. Redes Bayesianas

- **Exemplo 2:** Inferência Causal
  - Um indivíduo X pratica exercício.
  - De acordo com a rede Bayesiana da figura, deverá ser classificado como candidato ou não candidato a uma doença de coração?

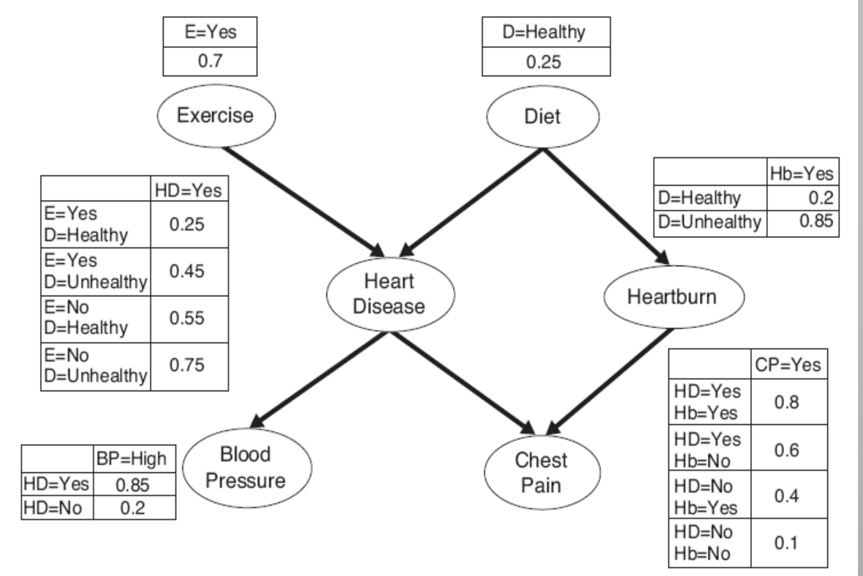


Figure 5.13. A Bayesian belief network for detecting heart disease and heartburn in patients.

6. Redes Bayesianas

1. A rede especifica que p(Exercise=Yes)=0,7.
2. Por se tratar de um nó sem pai, esta probabilidade significa que, retirada uma amostra da população, 70% dos indivíduos praticam exercício.
3. Se nada for dito em contrário, é com esta probabilidade que se lida no cálculo de outras probabilidades da rede.
4. Porém, neste caso o enunciado especifica o valor do atributo Exercise ao afirmar que X pratica exercício: ou seja, p(Exercise=Yes)=1 (logo, p(Exercise=No)=0)
5. Contudo, para dieta nada se sabe (o enunciado é omisso) e portanto X pode ou não ter uma dieta saudável
6. O problema consiste em calcular p(HD=Yes), sabendo que para X p(Exercise=Yes)=1 e nada sabendo de p(Diet=Healhty), pelo que se assumem as probabilidades dadas na tabela de Diet. Ou seja:

D=Healthy
0.25

	HD=Yes
E=Yes D=Healthy	0.25
E=Yes D=Unhealthy	0.45
E=No D=Healthy	0.55
E=No D=Unhealthy	0.75

p(HD=Yes|(Exercise=yes,Diet=Healthy V Exercise=yes,Diet=Unhealthy))=

=p(HD=Yes|(E=Yes,D=Yes).p(E=Yes).p(D=Yes)+p(HD=Yes|(E=Yes,D=No).p(E=Yes).p(D=No)=

=(0,25x1x0,25)+(0,45x1x(1-0,25))=0,4

6. Redes Bayesianas

Falta agora determinar a probabilidade de NÃO ter doença do coração:

p(HD=No|(Exercise=yes,Diet=Healthy V Exercise=yes,Diet=Unhealthy))=

=p(HD=No|(E=Yes,D=Yes).p(E=Yes).p(D=Yes)+p(HD=No|(E=Yes,D=No).p(E=Yes).p(D=No)=

=((1-0,25) x1x0,25)+((1-0,45)x1x(1-0,25))=0,6

- Como 0,6 > 0,4 o indivíduo X é classificado como **não candidato** a doença do coração
- Notar que o segundo cálculo é o complementar do primeiro, pelo que poderia ter-se feito assim:

p(HD=No|(Exercise=yes,Diet=Healthy V Exercise=yes,Diet=Unhealthy))=

=1-p(HD=Yes|(Exercise=yes,Diet=Healthy V Exercise=yes,Diet=Unhealthy))=

=1-0,4=

=0,6

6. Redes Bayesianas

➢ **Exemplo 3:** Inferência Causal

- Sem qualquer informação adicional, e no cenário descrito por esta rede, inferir qual a probabilidade de uma pessoa ter uma doença do coração (HD)

p(HD=Yes)=

=p(HD=Yes|E=Yes,D=Yes).p(E=Yes).p(D=Yes)+p(HD=Yes|E=No,D=Yes).p(E=No).p(D=Yes)+

+p(HD=Yes|E=No,D=No).p(E=No).p(D=No)+p(HD=Yes|E=No,D=No).p(E=No).p(D=No)=

=(0,25x0,7x0,25)+(0,45x0,7x(1-0,25))+(0,75x(1-0,7)x(1-0,25)+(0,55x(1-0,7)x0,25)=

=(0,25x0,7x0,25)+(0,45x0,7x0,75)+(0,75x0,3x0,75)+(0,55x0,3x0,25)=

=0,49

p(HD=No)=1-p(HD=Yes)=
=1-0,49=
=0,51

Como 0,51 > 0,49, na ausência de informação adicional há uma ligeira probabilidade a favor de não se contrair doença do coração (HD)

6. Redes Bayesianas

➢ **Exemplo 4:** Inferência Causal

- Qual a probabilidade de uma pessoa ter tensão alta (BP=H) sabido que pratica exercício e tem dieta saudável ?

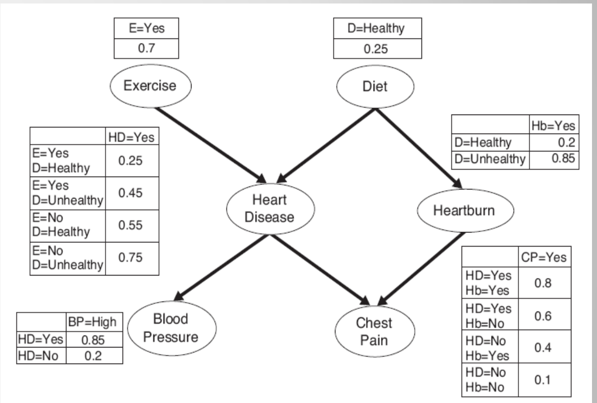


Figure 5.13. A Bayesian belief network for detecting heart disease and heartburn in patients.

$$P(x_1, x_2 \ldots x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i))$$

p(BP=Yes, E=Yes, D=Yes)=

=p(BP=H|HD=Yes,E=YES,D=Yes ∨ HD=No,E=YES,D=Yes)=

= p(BP=H|HD=Yes,E=YES,D=Yes)+ p(BP=H|HD=No,E=YES,D=Yes)=

=p(BP=H|HD=Yes).p(HD=Yes|E=Yes,D=Yes).p(E=Yes).p(D=Yes)+

+p(BP=H|HD=No).p(HD=No|E=Yes,D=Yes).p(E=Yes).p(D=Yes)=

=(0,85x0,25x1x1)+(0,2x(1-0,25)x1x1)=

=0,3625

6. Redes Bayesianas

➢ **Exemplo 5:** Inferência de Diagnóstico

Qual a probabilidade de doença do coração para um doente que tem tensão arterial elevada, não faz exercício nem tem uma dieta saudável?

$$p(x_1 \mid x_2 \ldots x_n) = \frac{p(x_1, x_2 \ldots x_n)}{p(x_2 \ldots x_n)}$$

$$P(x_1, x_2 \ldots x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i))$$

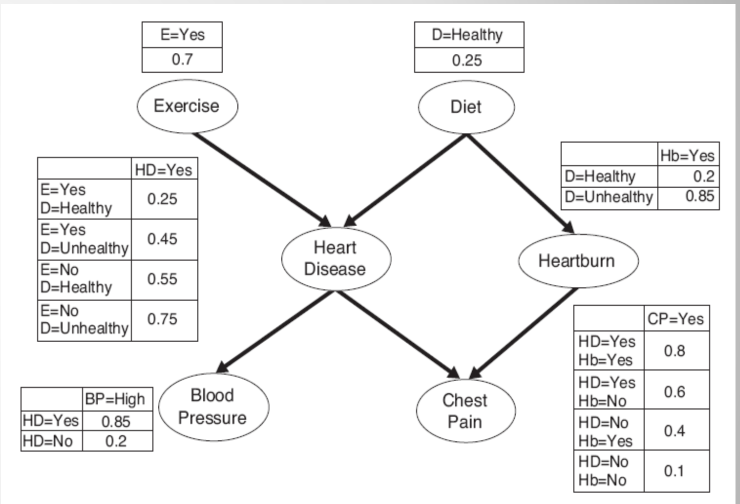


Figure 5.13. A Bayesian belief network for detecting heart disease and heartburn in patients.

$$p\left(DC \mid TA, \bar{E}, \bar{D}\right) = \frac{p\left(DC, TA, \bar{E}, \bar{D}\right)}{p\left(TA, \bar{E}, \bar{D}\right)} =$$

$$= \frac{p(TA \mid DC).p\left(DC \mid \bar{E}, \bar{D}\right).p\left(\bar{E}\right).p\left(\bar{D}\right)}{p(TA \mid DC).p\left(DC \mid \bar{E}, \bar{D}\right).p\left(\bar{E}\right).p\left(\bar{D}\right) + p\left(TA \mid \overline{DC}\right).p\left(\overline{DC} \mid \bar{E}, \bar{D}\right).p\left(\bar{E}\right).p\left(\bar{D}\right)} =$$

$$= \frac{0{,}85 \times 0{,}75 \times 1 \times 1}{0{,}85 \times 0{,}75 \times 1 \times 1 + 0{,}2 \times (1 - 0{,}75) \times 1 \times 1} = 0{,}9272$$

6. Redes Bayesianas

Redes Bayesianas - Resumo

- Este classificador:
  - Apresenta uma solução gráfica para a representação de conhecimento / probabilidades / dependência
  - Pode representar ligações causais
  - Construir uma rede real pode consumir muito tempo e esforço
    - Contudo, a sua manutenção é simples
  - É adequado a domínios com dados incompletos
  - É robusto em relação a *overfitting*

6. Redes Bayesianas

Redes Bayesianas

FIM

Naïve Art