Parte III: IA e Incerteza.

Prof. Fabiano Araujo Soares, Dr. / FGA 0221 - Inteligência Artificial

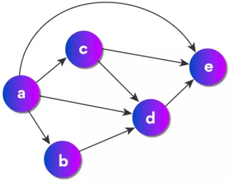
Universidade de Brasília

2025

UnB - FCTE
Faculdade de Ciências e Tecnologias em Engenharia

conhecimento em movimento
sociedade em transformação

Redes Bayesianas



Como um sistema de IA pode tomar decisões confiáveis em ambientes incertos mesmo quando as evidências são imprecisas ou parciais?

Redes Bayesianas são a resposta: elas combinam conhecimento causal com dados observados para calcular racionalmente o que é mais provável, permitindo que máquinas "raciocinem sob incerteza".

Motivação: Por que Redes Bayesianas?

- Em IA e Engenharia, poucas situações têm informação completa ou livre de ruído: incerteza é a regra.
- Modelos como o Naive Bayes simplificam suposições (independência total entre variáveis) — o que raramente é realista.
- Precisamos representar relações de dependência e independência condicional de forma compacta e interpretável.
- Redes Bayesianas oferecem um grafo probabilístico que combina intuição visual e base matemática sólida.
- Permitem atualizar crenças à medida que novas evidências chegam (raciocínio probabilístico dinâmico).

Usos marcantes:

- Diagnóstico médico assistido por computador (p.ex., avaliação de sintomas para doenças raras);
- Planejamento e tomada de decisão em sistemas autônomos (robótica, veículos inteligentes);
- Análise de causas em engenharia de falhas e segurança;
- Sistemas de recomendação sensíveis a contexto.

Redes Bayesianas: Vantagem chave

Representação **compacta e interpretável** de distribuições conjuntas de alta dimensionalidade.

Definição e Estrutura

O que é uma Rede Bayesiana?

- Um grafo dirigido acíclico (DAG) em que cada nó representa uma **variável aleatória** (discreta ou contínua).
- As arestas (setas) indicam dependências diretas: se há uma seta de A para B, então A é um *pai* de B e influencia sua distribuição.
- A ausência de ciclos garante que não haja dependências circulares — o que é essencial para a consistência probabilística.
- Cada variável tem uma **Tabela de Probabilidade Condicional (CPT)** que especifica $P(X_i \mid Pais(X_i))$.
- A estrutura do grafo codifica independências condicionais: duas variáveis são independentes dado um conjunto de evidências se forem d-separadas no grafo.

Princípio fundamental:

A distribuição conjunta de todas as variáveis $X_1, X_2, \ldots, X_n$ pode ser decomposta como:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pais(X_i))$$

onde $Pais(X_i)$ são os nós que têm arestas apontando para X_i.

Fatoração da Distribuição Conjunta (cont.)

Exemplo: Diagnóstico médico

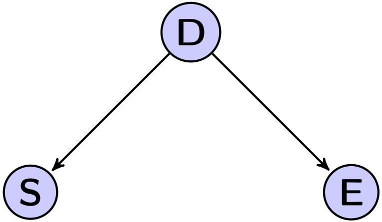
Variáveis: D (Doença), S (Sintoma), E (Exame).
Estrutura: $D \to S$, $D \to E$

$$P(D, S, E) = P(D) \cdot P(S \mid D) \cdot P(E \mid D)$$

- Sem a rede, precisaríamos de $2^3 - 1 = 7$ parâmetros (para 3 variáveis binárias).
- Com a rede, apenas: 1 (para D) + 2 (para $S \mid D$) + 2 (para $E \mid D$) $= 5$.

A fatoração reduz complexidade e facilita inferência!

Interpretação do Diagrama: Diagnóstico Médico



Implicação probabilística

A estrutura implica:

$$P(S, E \mid D) = P(S \mid D) \cdot P(E \mid D)$$

Essa independência condicional é o que permite a fatoração eficiente da distribuição conjunta.

- **Nó** D **(Doença)**: variável raiz, representa a causa subjacente.
- **Nó** S **(Sintoma)**: depende de D — pacientes com a doença têm maior probabilidade de apresentar o sintoma.
- **Nó** E **(Exame)**: depende de D — o resultado do exame é influenciado pela presença da doença.
- **Ausência de aresta entre** S **e** E: dado D, o sintoma e o exame são condicionalmente independentes.
  - Ou seja: se sabemos o estado da doença, saber o sintoma não muda nossa crença sobre o exame (e vice-versa).

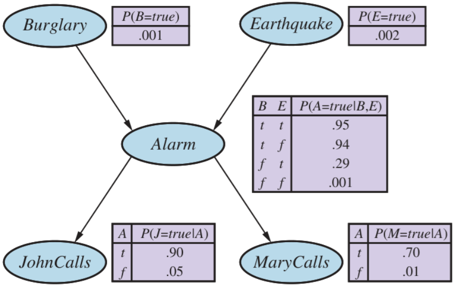
Exemplo: Alarme de Roubo (Russell & Norvig)

- Um morador de LA (Bruce) comprou um novo sistema antirroubo que detecta tentativas de invasão de forma bastante eficaz mas dispara as vezes quando terremotos ocorrem;
- Bruce tem dois vizinhos Mary e John que prometeram telefonar para ele caso o alarme acione;
- John quase sempre liga quando escuta o alarme, mas às vezes confunde o telefone tocando com o alarme e liga também;
- Mary, por outro lado, gosta de música alta e muitas vezes não escuta o alarme.

- Dada a evidência de quem ligou ou não, gostaríamos de estimar a probabilidade de um roubo.

Exemplo: Alarme de Roubo (Russell & Norvig)

- Variáveis: Roubo (B), Terremoto (E), Alarme (A), JohnLiga (J), MaryLiga (M)
- CPTs baseadas em conhecimento causal



Fatoração:

$$P(B, E, A, J, M) = P(B)P(E)P(A \mid B, E)P(J \mid A)P(M \mid A)$$

Independência Condicional

- Em uma Rede Bayesiana, a estrutura do grafo codifica relações de dependência e independência entre variáveis.
- Regra central: Cada nó (variável) é condicionalmente independente de todos os seus não-descendentes, dadas as variáveis que são seus pais.
- Formalmente: Se X é um nó e $Pa(X)$ seus pais, então X é independente dos não-descendentes dados $Pa(X)$.
- Isso simplifica a descrição da distribuição conjunta e reduz o número de parâmetros.

Independência Condicional

- Em uma Rede Bayesiana, a estrutura do grafo codifica relações de dependência e independência entre variáveis.
- Regra central: Cada nó (variável) é condicionalmente independente de todos os seus não-descendentes, dadas as variáveis que são seus pais.
- Formalmente: Se X é um nó e $Pa(X)$ seus pais, então X é independente dos não-descendentes dados $Pa(X)$.
- Isso simplifica a descrição da distribuição conjunta e reduz o número de parâmetros.

Exemplo:

$$P(J, M \mid A) = P(J \mid A) \cdot P(M \mid A)$$

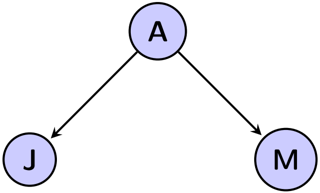
Dado o alarme A, João (J) e Maria (M) tornam-se independentes.

Independência Condicional

- Estrutura do grafo codifica independências: cada variável é independente dos não-descendentes dado seus pais.
- Formalmente: se X é um nó e $Pa(X)$ seus pais, então $X \perp\!\!\!\perp$ não-descendentes dado $Pa(X)$ (X é condicionalmente independente de seus não-decendentes dado $Pa(X)$).
- Facilita a fatoração da distribuição conjunta.

Exemplo:

$$P(J, M \mid A) = P(J \mid A) \cdot P(M \mid A)$$



Bloco (Cobertura) de Markov

- O **bloco de Markov** de um nó X em uma Rede Bayesiana é formado por:
  - Os pais de X
  - Os filhos de X
  - Os demais pais dos filhos de X
- Se condicionarmos sobre o bloco de Markov de X, então X se torna independente do restante da rede.
- Usado para construir algoritmos eficientes de inferência e de amostragem (Gibbs Sampling, etc).

Bloco (Cobertura) de Markov

- O **bloco de Markov** de um nó X em uma Rede Bayesiana é formado por:
  - Os pais de X
  - Os filhos de X
  - Os demais pais dos filhos de X
- Se condicionarmos sobre o bloco de Markov de X, então X se torna independente do restante da rede.
- Usado para construir algoritmos eficientes de inferência e de amostragem (Gibbs Sampling, etc).

Exemplo:

- No problema do alarme, o bloco de Markov do nó A é $\{R, T, J, M\}$ (Roubo, Terremoto, JoãoLiga, MariaLiga).
- Condicionando $(A \mid R, T, J, M)$, A é independente dos outros nós da rede.

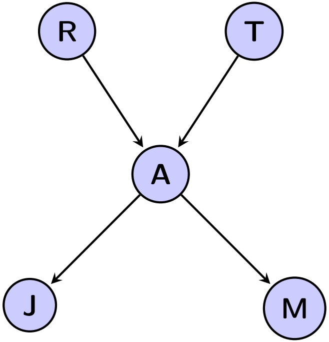
Bloco de Markov (Cobertura de Markov)

- O bloco de Markov de um nó X é o conjunto formado por seus pais, filhos e os outros pais dos seus filhos.
- Condicionando no bloco de Markov, X é independente do restante da rede.
- Isso é fundamental para algoritmos de inferência local (ex: Gibbs Sampling).

Exemplo no problema do alarme: O bloco de Markov do nó A é:

$$\{R, T, J, M\}$$

onde R=Roubo, T=Terremoto, J=João Liga, M=Maria Liga.



d-separação

- A **d-separação** é um critério gráfico que determina se dois conjuntos de variáveis são condicionalmente independentes em um DAG.
- Um conjunto Z d-separa X de Y se todos os caminhos entre X e Y são bloqueados por Z, segundo regras específicas:
  - Caminhos de "caso causal"($\rightarrow$) ou "caso comum"($\leftarrow$) são bloqueados se o nó intermediário está em Z.
  - Caminhos "caso v-estrutural"($\rightarrow V \leftarrow$) só ficam bloqueados se nem o nó v-estrutural nem seus descendentes estão em Z.
- d-separação permite identificar independências e orientar inferências eficientes.

d-separação

- A **d-separação** é um critério gráfico que determina se dois conjuntos de variáveis são condicionalmente independentes em um DAG.
- Um conjunto Z d-separa X de Y se todos os caminhos entre X e Y são bloqueados por Z, segundo regras específicas:
  - Caminhos de "caso causal"($\rightarrow$) ou "caso comum"($\leftarrow$) são bloqueados se o nó intermediário está em Z.
  - Caminhos "caso v-estrutural"($\rightarrow V \leftarrow$) só ficam bloqueados se nem o nó v-estrutural nem seus descendentes estão em Z.
- d-separação permite identificar independências e orientar inferências eficientes.

Exemplo:

- No alarme: J (JoãoLiga) e M (MariaLiga) são d-separados se condicionarmos em A (Alarme).
- Portanto, $P(J, M \mid A) = P(J \mid A) \cdot P(M \mid A)$.

d-separação

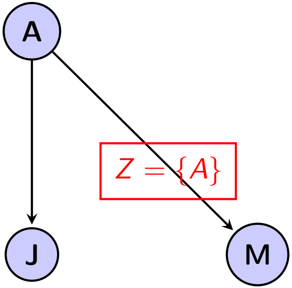
- d-separação é um critério para determinar independência condicional em grafos dirigidos acíclicos.
- Um caminho é bloqueado por um conjunto de nós Z se atende às regras:
  - **Caminhos em cadeia** $(X \to Y \to Z)$ e **forks** $(X \leftarrow Y \to Z)$ são bloqueados se o nó intermediário está em Z.
  - **V-estruturas** $(X \to Y \leftarrow Z)$ são bloqueadas **somente se** Y **não está em** Z e **nem nenhum de seus descendentes**.
- Se todos os caminhos entre X e Y são bloqueados por Z, então $X \perp\!\!\!\perp Y \mid Z$.

UnB

UnB - FCTE
Faculdade de Ciências e Tecnologias em Engenharia
conhecimento em **movimento**
sociedade em **transformação**

Exemplo: No problema do alarme, J e M são d-separados dado A. Isso implica:

$$P(J, M \mid A) = P(J \mid A) \cdot P(M \mid A)$$



Resumindo: Bloco de Markov e d-separação

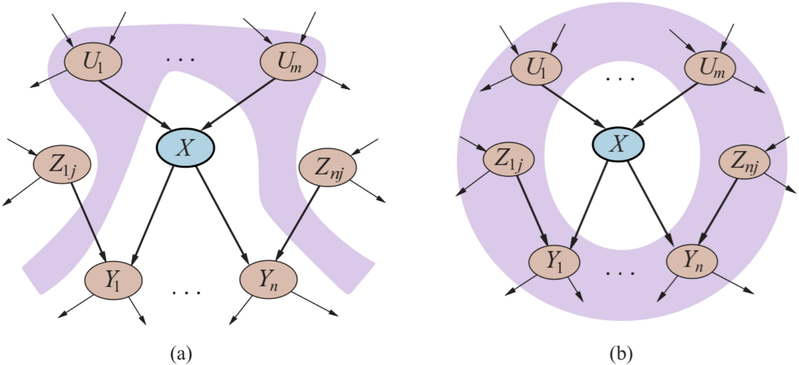


Figure 13.4 (a) A node X is conditionally independent of its non-descendants (e.g., the Z_{ij}s) given its parents (the U_is shown in the gray area). (b) A node X is conditionally independent of all other nodes in the network given its Markov blanket (the gray area).

Redes Bayesianas com variáveis contínuas

1. **Discretização**: Dividir os valores possíveis em intervalos fixos, por exemplo, faixas de temperatura ($< 0^\circ C$, $0^\circ C - 100^\circ C$, e $> 100^\circ C$);
2. **Distribuições probabilísticas**: Outra abordagem é definir uma variável contínua usando uma das famílias padrão de funções de densidade de probabilidade, como a distribuição Gaussiana $N(x; \sigma^2)$;
3. **Não paramétrica**: Define a distribuição condicional implicitamente com uma coleção de instâncias, cada uma contendo valores das variáveis genitoras e filho.

Redes Bayesianas com variáveis contínuas — Exemplo

- Redes Bayesianas que possuem variáveis discretas e contínuas são chamadas de **Redes Bayesianas Hibridas**;

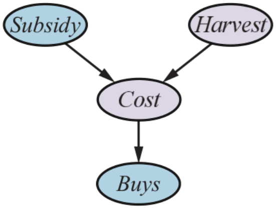


Figure 13.6 A simple network with discrete variables ($Subsidy$ and $Buys$) and continuous variables ($Harvest$ and $Cost$).

Redes Bayesianas com variáveis contínuas

- Para lidar com redes hibridas, precisamos aprender a converter variáveis genitoras contínuas em variáveis filhos discretas e vice versa;
- Para trazer a informação de genitores discretos em filhos contínuos podemos utilizar estratégias como **distribuição condicional linear-gaussiana** em que o filho tem uma distribuição gaussiana cuja média μ varia linearmente com o valor do genitor e cujo desvio padrão σ é fixo.
- Para trazer informações de genitores contínuos em filhos discretos podemos usar funções de limiarização como a integral da distribuição normal até o valor x;

Inferência: O que é e por que importa?

- O objetivo da inferência é calcular **probabilidades *a posteriori*** de variáveis de interesse, dadas evidências observadas.
- Exemplo: dado que Maria ligou ($M = V$), qual é a probabilidade de haver um roubo ($R = V$)?

$$P(R = V \mid M = V)$$

- Isso envolve atualizar crenças usando a regra de Bayes e a estrutura da rede.
- Tipos comuns de consultas:
  - **Diagnóstico**: $P(\text{causa} \mid \text{efeito})$
  - **Predição**: $P(\text{efeito} \mid \text{causa})$
  - **Explicação mais provável (MAP)**: qual configuração de variáveis explica melhor as evidências?
- A estrutura do grafo permite computar essas probabilidades de forma eficiente — sem precisar da tabela completa.

Inferência Exata: Explorando a Estrutura

- **Eliminação de Variáveis**:
  - Algoritmo que calcula $P(X \mid E)$ somando variáveis não relevantes na ordem certa.
  - Usa a fatoração da rede: $P(X_1, \ldots, X_n) = \prod_i P(X_i \mid \text{Pais}(X_i))$.
  - Ordem de eliminação afeta eficiência — ideal: minimizar tamanho dos fatores intermediários.
- **Propagação em Árvore de Cliques (Clique Tree)**:
  - Transforma a rede em uma árvore de cliques (super nós com variáveis agrupadas).
  - Permite inferência eficiente via troca de mensagens entre cliques.
  - Ideal para múltiplas consultas: a estrutura é pré-compilada.
- **Complexidade**: exponencial na *treewidth* da rede, mas viável para redes esparsas ou com estrutura de árvore.

Inferência Exata: Exemplo

- Vamos retornar ao exemplo do **alarme contra roubo**. Poderíamos perguntar a probabilidade de um roubo dado que **John ligou** e **Mary ligou**:

  $$P(Burglary \mid JohnCalls = true, MaryCalls = true) = \langle 0.284, 0.716 \rangle$$

- uma consulta do tipo $P(X \mid \boldsymbol{e})$ pode ser respondida como:

  $$P(X \mid \boldsymbol{e}) = \alpha P(X \wedge \boldsymbol{e}) = \alpha \sum_{\boldsymbol{y}} P(X \wedge \boldsymbol{e} \wedge \boldsymbol{y})$$

  onde **y** são todas as combinações possíveis de valores das variáveis não observadas **Y**.
- Uma rede Bayesiana fornece uma representação da distribuição conjunta completa.
- Temos que os termos $P(x, \boldsymbol{e}, \boldsymbol{y})$ na distribuição conjunta podem ser escritos como produtos de probabilidades condicionais da rede.

Inferência Exata: Exemplo

- Portanto, uma consulta pode ser respondida usando uma rede Bayesiana calculando somas de produtos de probabilidades condicionais da rede.
- Por exemplo, considere a cunsulta:

  $$P(Burglary \mid JohnCalls = true, MaryCalls = true).$$

- As variáveis ocultas para esta consulta são **Earthquake** (e) e **Alarm** (a). Temos então: $P(B \mid j, m) = \alpha P(B, j, m) = \alpha \sum_e \sum_a \mathbf{P}(B, j, m, e, a)$

Inferência Exata: Exemplo

- A semântica das redes de Bayes nos dá então uma expressão em termos das entradas CPT. Para simplificar, fazemos isso apenas para $Burglary = true$:

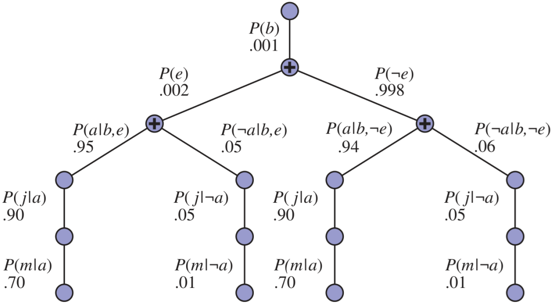
$$P(B \mid j, m) = \alpha \sum_e \sum_a P(b)P(e)P(a \mid b, e)P(j \mid a)P(m \mid a).$$

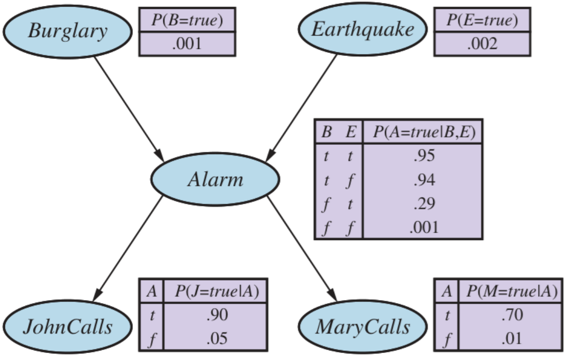
- Para simplificar o esforço computacional, podemos tirar do somatório as variáveis que não dependem das variáveis do somatório em questão:

$$P(B \mid j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e)P(j \mid a)P(m \mid a).$$

- Esta expressão pode ser avaliada percorrendo as variáveis em ordem, multiplicando entradas CPT à medida que avançamos.
- Para cada soma, também precisamos fazer um loop sobre os valores possíveis da variável.

Inferência Exata: Graficamente





Inferência Exata: Exemplo

- Dessa forma obtemos

  $$P(B \mid j, m) = \alpha 0.00059224.$$

- E ainda, considerando a probabilidade de não-roubo:

  $$P(B \mid j, m) = \alpha \langle 0.00059224, 0.0014919 \rangle \approx \langle 0.284, 0.716 \rangle.$$

- Ou seja, a chance de um roubo, dadas as ligações de ambos os vizinhos, é de cerca de 28%.

Inferência Exata: Conclusões

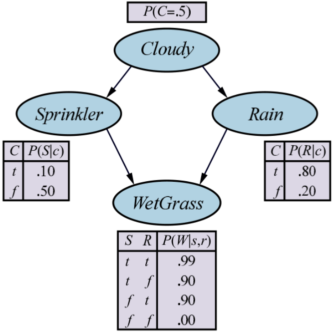
- Infelizmente a complexidade para uma rede Bayesiana com n variáveis ainda é alta $O(2^n)$;
- A título de comparação, uma rede como a do seguro de caro precisaria de 227 milhões de operações aritméticas para uma consulta;
- No entanto, se olharmos para a arvore anterior, podemos ver que existem expressões repetidas, como os produtos $P(j \mid a)P(m \mid a)$ e $P(j \mid \neg a)P(m \mid \neg a)$;
- A chave para uma inferência eficiente em redes Bayesianas é evitar tais cálculos repetidos.

- Para grandes redes, calcular a inferência exata é intratável.

Inferência Aproximada: Quando a exata é inviável

- Para redes grandes ou com alta conectividade, métodos exatos tornam-se impraticáveis.
- **Amostragem (Sampling)**:
  - Gera amostras da distribuição conjunta (ex: *Prior Sampling*, *Rejection Sampling*).
  - Estima probabilidades por frequência relativa.
- **MCMC (Markov Chain Monte Carlo)**:
  - Ex: *Gibbs Sampling* — atualiza uma variável por vez, condicionada ao resto.
  - Converge para a distribuição correta, mesmo sem conhecer a normalização.
  - Muito usado em redes com muitos ciclos ou evidências complexas.
- **Vantagem**: escalabilidade.
- **Desvantagem**: resultados são estimativas, com erro que diminui com mais amostras.

Inferência Aproximada: Exemplo — Método de amostragem direta



function PRIOR-SAMPLE(bn) **returns** an event sampled from the prior specified by bn
  inputs: bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $\mathbf{x} \leftarrow$ an event with n elements
  for each variable X_i **in** $X_1, \ldots, X_n$ **do**
    $\mathbf{x}[i] \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
  return x

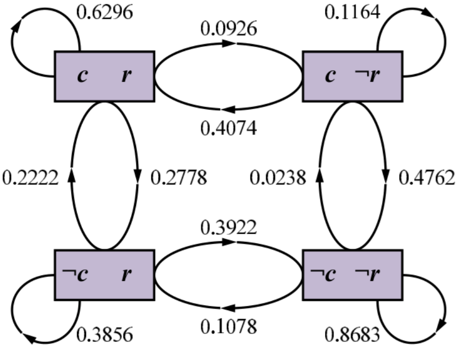
Inferência Aproximada: Exemplo — Método de amostragem direta

1. Amostra de $P(Cloudy) = \langle 0.5, 0.5 \rangle$, valor é **true**.
2. Amostra de $P(Sprinkler \mid Cloudy = true) = \langle 0.1, 0.9 \rangle$, valor é **false**.
3. Amostra de $P(Rain \mid Cloudy = true) = \langle 0.8, 0.2 \rangle$, valor é **true**.
4. Amostra de $P(WetGrass \mid Sprinkler = false, Rain = true) = \langle 0.9, 0.1 \rangle$, valor é **true**.

- Nesse caso, a amostra retornada pelo algoritmo seria **[true,false,true,true]**.

Inferência Aproximada: Exemplo — Cadeia de Markov Monte Carlo

- **Cadeia de Markov Monte Carlo** (MCMC): Em vez de gerar cada amostra do zero, os algoritmos MCMC gerar uma amostra fazendo uma alteração aleatória na amostra anterior.



Aprendizado de Parâmetros

- Dada a **estrutura** da rede, o objetivo é estimar as **Tabelas de Probabilidade Condicional (CPTs)** a partir de dados.
- **Dados completos** (todas as variáveis observadas):
  - Estimativa direta por **contagem de frequências**.
  - Ex: $P(S = V \mid D = V) = \frac{\text{número de casos com } S{=}V, D{=}V}{\text{número de casos com } D{=}V}$
  - Pode-se usar **apriori de Dirichlet** (suavização de Laplace) para evitar probabilidades zero.
- **Dados incompletos** (variáveis ocultas ou faltantes):
  - Não é possível contar diretamente.
  - Usa-se o algoritmo **EM (Expectation-Maximization)**:
    - **E-step**: Estima valores esperados das variáveis ocultas dadas as evidências e os parâmetros atuais.
    - **M-step**: Atualiza os parâmetros maximizando a verossimilhança esperada.
  - Itera até convergência.

Algoritmo EM: E-step e M-step Explicados

Objetivo: Estimar parâmetros quando há **variáveis ocultas** (ex: D não observada).

Exemplo: Rede $D \to S$, com dados apenas de S:

Caso	1	2	3	4	5
S	V	V	F	V	F

Parâmetros a estimar:

- $\theta = P(D = V)$
- $\theta_1 = P(S = V \mid D = V)$
- $\theta_0 = P(S = V \mid D = F)$

Importante: θ_1 e θ_0 são independentes — não precisam somar 1!

Inicialização:

- $\theta = 0.5$, $\theta_1 = 0.7$, $\theta_0 = 0.3$

E-step: Estimando Valores Esperados

Para cada caso, calculamos a **probabilidade esperada** de $D = V$ dado S, usando Bayes:

$$P(D = V \mid S) = \frac{P(S \mid D = V)P(D = V)}{P(S)}$$

Exemplo para Caso 1 ($S = V$):

$$P(D = V \mid S = V) = \frac{0.7 \times 0.5}{0.7 \times 0.5 + 0.3 \times 0.5} = \frac{0.35}{0.5} = 0.7$$

Repetimos para todos os casos $\rightarrow$ obtemos valores esperados de D.

Resultado do E-step: uma tabela com $E[D = V \mid S]$ para cada caso:

Caso	1	2	3	4	5
S	V	V	F	V	F
$E[D = V]$	0.7	0.7	0.3	0.7	0.3

M-step: Maximizando a Verossimilhança Esperada

Com os valores esperados do E-step, atualizamos os parâmetros:

1. $P(D = V)$:

$$\theta = \frac{\sum E[D = V]}{N} = \frac{0.7 + 0.7 + 0.3 + 0.7 + 0.3}{5} = 0.54$$

2. $P(S = V \mid D = V)$:

$$\theta_1 = \frac{\sum_{S=V} E[D = V]}{\sum E[D = V]} = \frac{0.7 + 0.7 + 0.7}{0.7 + 0.7 + 0.3 + 0.7 + 0.3} = \frac{2.1}{2.7} \approx 0.78$$

3. $P(S = V \mid D = F)$:

$$\theta_0 = \frac{\sum_{S=V} E[D = F]}{\sum E[D = F]} = \frac{(1 - 0.7) + (1 - 0.7) + (1 - 0.7)}{(1 - 0.7) + (1 - 0.7) + (1 - 0.3) + (1 - 0.7) + (1 - 0.3)} = \frac{0.9}{2.3} \approx 0.39$$

Próxima iteração: usar novos $\theta, \theta_1, \theta_0$ no E-step. **Convergência**: quando os parâmetros estabilizam.

Aprendizado de Estrutura

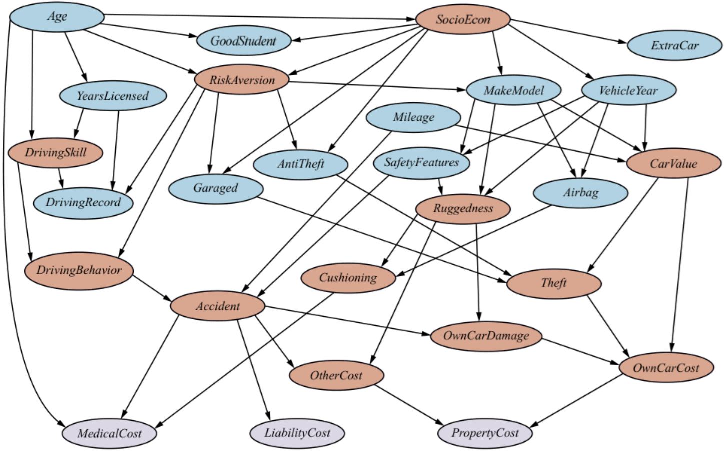
- Objetivo: descobrir a **estrutura do grafo** (quem influencia quem) a partir de dados.
- É um problema difícil: número de DAGs cresce exponencialmente com o número de variáveis.
- **Abordagem comum**: busca heurística no espaço de estruturas.
  - Começa com uma estrutura inicial (ex: vazia).
  - Aplica operações: adicionar, remover ou inverter arestas.
  - Usa uma **função de pontuação** para avaliar cada estrutura.
- **Critérios de pontuação**:
  - **MDL (Minimum Description Length)**: favorece modelos que comprimem bem os dados.
  - **AIC (Akaike Information Criterion)**: equilibra verossimilhança e número de parâmetros.
  - **BIC (Bayesian Information Criterion)**: similar ao AIC, mas com penalização mais forte para complexidade.
- **Algoritmo K2**: busca eficiente com ordem de nós pré-definida e prior uniforme.

Integração com Conhecimento de Domínio

- O aprendizado não precisa ser puramente baseado em dados.
- É comum **incorporar conhecimento de especialistas**:
  - Fixar certas arestas (ex: sabemos que $D \rightarrow S$).
  - Restringir a ordem dos nós (para algoritmos como K2).
  - Definir subestruturas conhecidas (ex: módulos funcionais).
- **Vantagens**:
  - Reduz espaço de busca.
  - Melhora interpretabilidade.
  - Corrige viés de dados escassos ou ruidosos.
- **Desafios**:
  - Conflito entre dados e conhecimento.
  - Dificuldade de elicitação de conhecimento.
- Em muitos casos, usa-se uma abordagem **híbrida**: estrutura parcialmente definida por especialistas, parâmetros aprendidos dos dados.

Rede Bayesiana: Exemplo Real

- Aqui está um exemplo de uma rede Bayesiana para decisão sobre seguro de automóvel.



Aplicações em Saúde e Diagnóstico

- **Diagnóstico médico**: modelar incerteza entre doenças, sintomas e exames.
  - Exemplo clássico: **Pathfinder** — sistema para diagnóstico de doenças linfáticas.
  - Usava uma rede com 60+ variáveis e alcançou desempenho superior ao de especialistas.
- **Sistemas de apoio à decisão clínica**:
  - Ex: ajuste de **dosagem de insulina** em diabéticos.
  - Incorpora variáveis como glicemia, refeições, atividade física e histórico.
  - Atualiza recomendações com base em evidências observadas.
- **Vantagem**: interpretabilidade — médicos entendem o raciocínio probabilístico.

Aplicações em Engenharia e Automação

- **Controle de veículos autônomos**:
  - Fusão de sensores (câmeras, LIDAR, radar) com incerteza.
  - Inferência sobre estado do ambiente (ex: presença de pedestre) dado ruído nos dados.
  - Decisões robustas mesmo com informações incompletas.
- **Modelagem de risco operacional** (bancos e indústria):
  - Avaliação de falhas em processos, fraudes, falhas de sistemas.
  - Combina dados históricos com conhecimento de especialistas.
  - Ex: estimar probabilidade de falha em um sistema crítico.
- **Manutenção preditiva**:
  - Prever falhas em equipamentos com base em sensores e histórico.
  - Reduz custos e paradas não planejadas.

Aplicações em Educação e Sistemas Inteligentes

- **Tutores inteligentes** e **sistemas adaptativos de ensino**:
  - Modelam o conhecimento do aluno como variáveis ocultas.
  - Inferem o nível de domínio com base em respostas a exercícios.
  - Ex: se o aluno erra uma questão, a rede atualiza a crença sobre seus pontos fracos.
- **Personalização do aprendizado**:
  - Seleciona próximos tópicos ou exercícios com base na inferência.
  - Adapta ritmo e estilo conforme o perfil do aluno.
- **Vantagem**: combina modelagem cognitiva com incerteza — ideal para ambientes dinâmicos.
- **Exemplo real**: sistemas como *Bayesian Knowledge Tracing (BKT)* usados em plataformas educacionais.

Vantagens das Redes Bayesianas

- **Representação compacta de dependências probabilísticas**:
  - Reduz exponencialmente o número de parâmetros via independência condicional.
  - Ex: 10 variáveis binárias — sem estrutura: $2^{10} - 1 = 1023$ parâmetros; com rede esparsa: ~ 50 parâmetros.
- **Combinação natural de conhecimento *a priori* e aprendizado**:
  - Especialistas definem estrutura causal; dados refinam parâmetros.
  - Ex: médico define que Doença $\rightarrow$ Sintoma; dados estimam $P(\text{Sintoma} \mid \text{Doença})$.
- **Capacidade de inferência bidirecional e explicável**:
  - Diagnóstico: $P(\text{causa} \mid \text{efeito})$ e predição: $P(\text{efeito} \mid \text{causa})$.
  - **Transparência**: cada decisão é justificada por cálculos probabilísticos interpretáveis.
  - Crucial em aplicações críticas (saúde, segurança).
- **Robustez à incerteza**: lida naturalmente com dados incompletos, ruído e ambiguidade.

Limitações e Desafios

- **Dificuldade na elicitação de CPTs para muitos nós/pais**:
  - Um nó com k pais binários requer 2^k probabilidades condicionais.
  - Ex: se X tem 5 pais binários, são $2^5 = 32$ valores para $P(X \mid \text{pais})$.
  - Cresce exponencialmente — inviável para especialistas especificarem manualmente.
- **Inferência exata inviável para topologias densas**:
  - Complexidade é exponencial na *treewidth* da rede.
  - Redes com muitos ciclos ou alta conectividade exigem métodos aproximados.
  - Trade-off: precisão vs. eficiência computacional.
- **Estrutura causal difícil de descobrir automaticamente**:
  - Dados apenas observacionais não distinguem causalidade de correlação.
  - Grafos equivalentes (mesmo ajuste aos dados, estruturas diferentes).
  - Necessita conhecimento de domínio ou dados experimentais.
- **Suposição de independência condicional** pode ser restritiva em domínios complexos.

Redes Bayesianas vs. Outras Abordagens

Critério	Redes Bayesianas	Redes Neurais	Árvores de Decisão
Interpretabilidade	Alta	Baixa	Alta
Robustez à incerteza	Alta	Média	Baixa
Eficiência com poucos dados	Alta	Baixa	Média
Escalabilidade	Limitada	Alta	Média
Conhecimento prévio	Facilita	Dificulta	Média

Conclusão: Redes Bayesianas são ideais quando *interpretabilidade*, *robustez à incerteza* e *integração de conhecimento prévio* são prioritárias.

Referências

- Russell, S., Norvig, P. Artificial Intelligence: A Modern Approach, Cap. 14
- Heckerman, D. "A tutorial on learning with Bayesian networks", (arXiv:2002.00269)
- Pearl, J. "Probabilistic Reasoning in Intelligent Systems", 1988
- Slide decks: Poupart (UW), Conati (UBC), Lecture121.pdf

Obrigado!

E-mail: fabianosoares@unb.br

LinkedIn: https://www.linkedin.com/in/fabiano-soares-06b6a821a/

Site do curso: https://www.fabianosoares.eng.br/fga0221-inteligência-artificial