INF623

2024/1

UFV

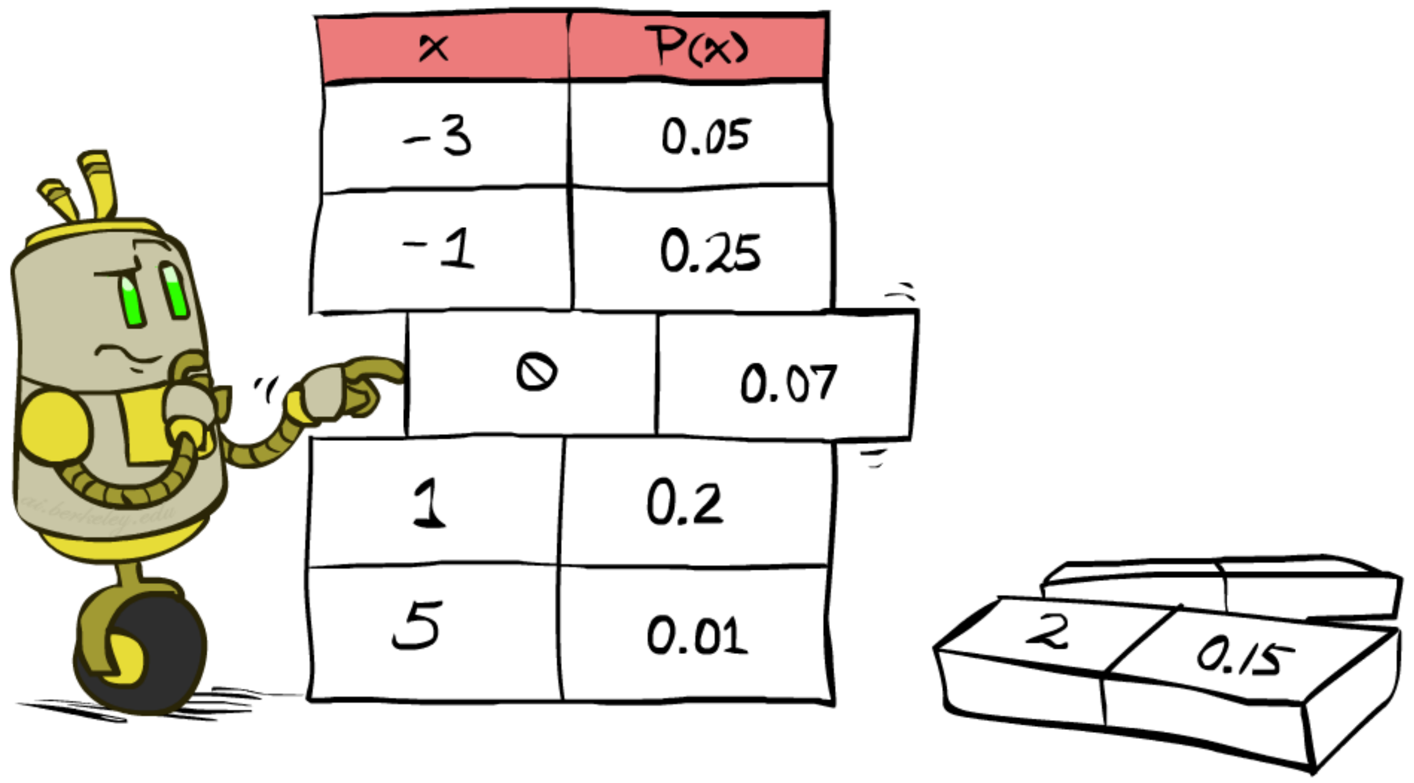
Inteligência Artificial

A15: Raciocínio Probabilístico II

Plano de aula

- Teorema de Bayes
- Independência
- Independência condicional
- Redes Bayesianas

Na última aula



Problemas da inferência por enumeração:

- Complexidade de tempo $O(d^n)$
- Complexidade de espaço para armazenar a distribuição conjunta $O(d^n)$
- Obter $O(d^n)$ exemplos para estimar as entradas da distribuição conjunta

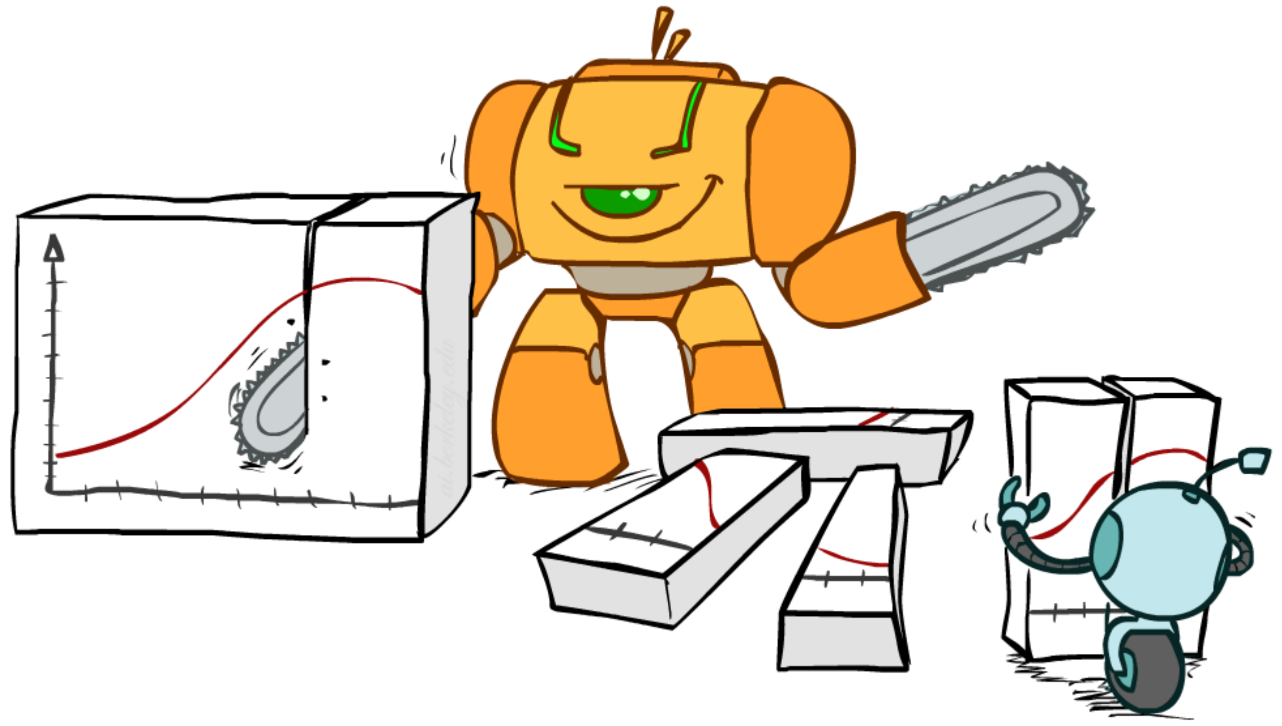
Teorema de Bayes

Regra do produto

$$P(a, b) = P(a \mid b)P(b) = P(b \mid a)P(a)$$

Dividindo por $P(b)$

$$P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)}$$



Qual a utilidade do Teorema de Bayes?

- Podemos calcular uma probabilidade condicional a partir da sua inversa
- Frequentemente, uma probabilidade condicional é complicada mas a outra é simples
- Descreve uma regra de atualização de $P(a)$ anterior para o $P(a \mid b)$ posterior

Inferência com o Teorema de Bayes

O Teorema de Bayes é muito utilizado em situações de causa/efeito:

$$P(causa \mid efeito) = \frac{P(efeito \mid causa)P(causa)}{P(efeito)}$$

Exemplo: M: Meningite, S: torcicolo

$$P(m) = 0{,}0001$$
$$P(s \mid m) = 0{,}8$$
$$P(s \mid \neg m) = 0{,}01$$

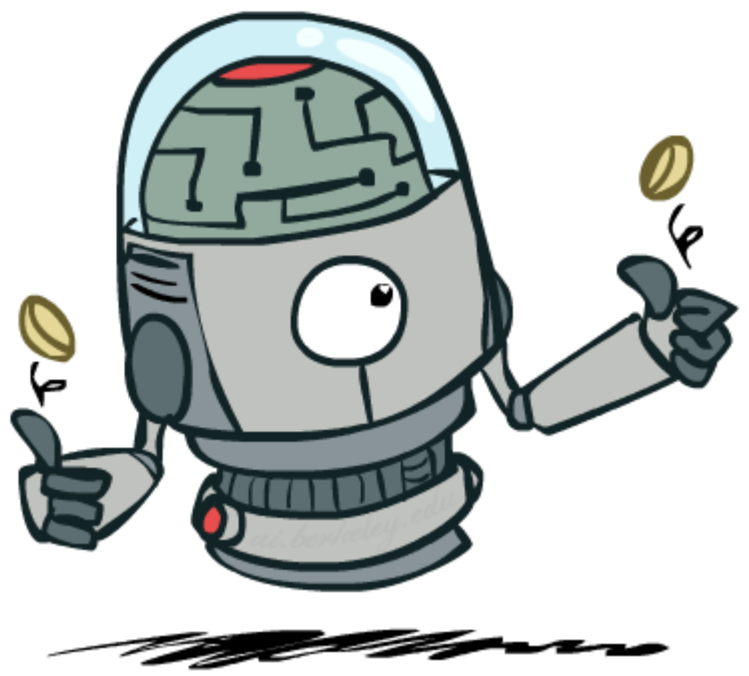
$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{P(s \mid m)P(m)}{P(s \mid m)P(m) + P(s \mid \neg m)P(\neg m)} = 0{,}007944$$

Independência

Duas variáveis X e Y são independentes ($X \perp\!\!\!\perp Y$) se e somente se:

$X \perp\!\!\!\perp Y \Leftrightarrow P(X, Y) = P(X)P(Y)$ → A distribuição conjunta é fatorada em distribuições mais simples

$X \perp\!\!\!\perp Y \Leftrightarrow P(X \mid Y) = P(X)$ → O resultado de X independente do resultado de Y



Exemplos:

- Independentes → lançar duas moedas
- Dependentes → se estiver nublado pela manhã, é mais provável que chova durante o dia

- Em geral, variáveis aleatórias não são independentes. No entanto, muitas vezes assumimos independência para simplificar o modelo de probabilidades.

Exemplo 1: dependência

Temperatura T e clima C

$\neq$

$P(T, C)$

T	C	P
quente	ensolarado	0,4
quente	chuvoso	0,1
frio	ensolarado	0,2
frio	chuvoso	0,3

$P(T)$

T	P
quente	0,5
Frio	0,5

$P(C)$

T	P
ensolarado	0,6
chuvoso	0,4

$P(T, C) = P(T)P(C)$

T	C	P
quente	ensolarado	0,3
quente	chuvoso	0,2
frio	ensolarado	0,3
frio	chuvoso	0,2

Exemplo 2: independência

Lançamento de n moedas justas

$P(X_1)$

X1	P
cara	0,5
coroa	0,5

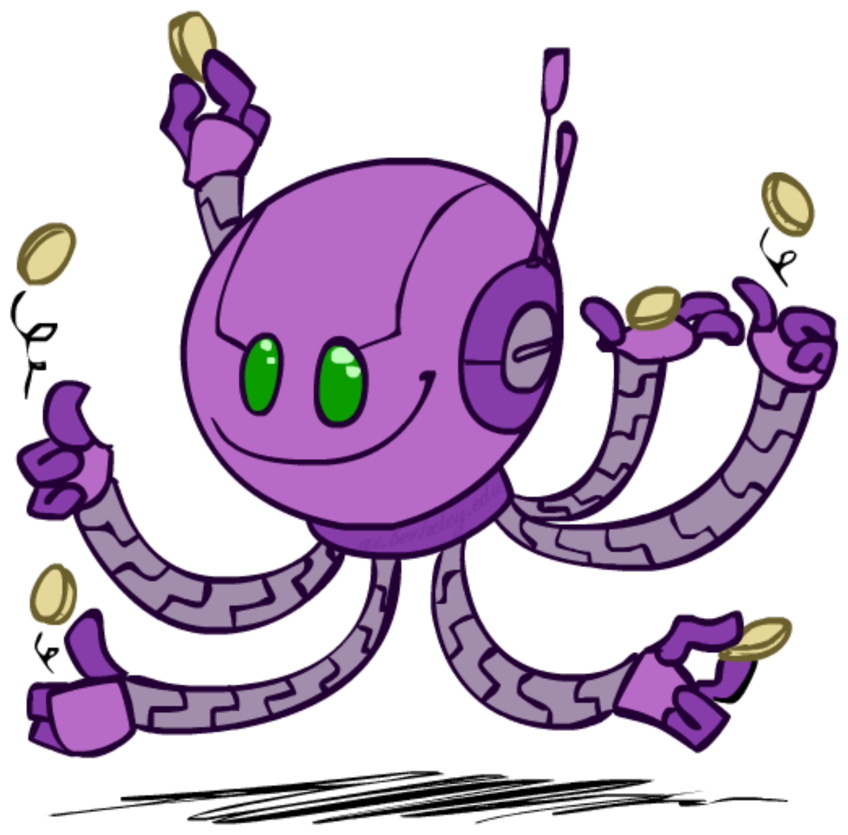
$P(X_2)$

X2	P
cara	0,5
coroa	0,5

$\cdots$

$P(X_3)$

Xn	P
cara	0,5
coroa	0,5



$P(X_1, X_2, \ldots, X_n)$

2^n

Independência condicional

Duas variáveis X e Y são condicionalmente independentes dado Z ($X \perp\!\!\!\perp Y \mid Z$) se e somente se:

$$\forall x, y, z : P(x, y \mid z) = P(x \mid z)P(y \mid z)$$

Ou, de maneira equivalente:

$$\forall x, y, z : P(x \mid y, z) = P(x \mid z)$$

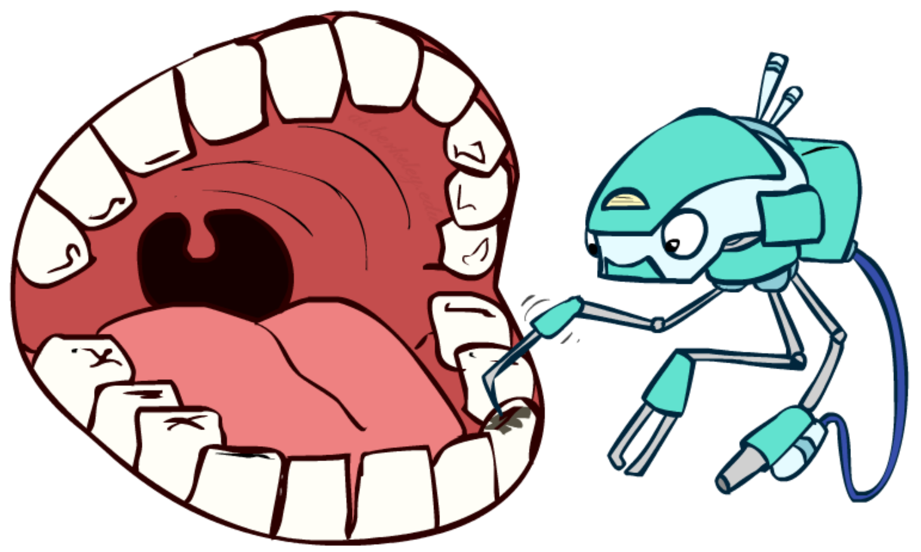
- Independência (absoluta) raramente acontece, por quê?
- Independência condicional é uma suposição mais fraca que independência (absoluta) e é a nossa forma mais básica e robusta de conhecimento sobre ambientes incertos.

Exemplo 3: independencia condicional

$P(\text{DorDente}, \text{Cavidade}, \text{EstileteAgarrar})$

Se eu tenho uma cavidade no dente, a probabilidade do estilete agarrar não depende de eu estar com dor de dente:

- $P(\text{EstileteAgarrar} \mid \text{DorDente}, \text{Cavidade}) = P(\text{EstileteAgarrar} \mid \text{Cavidade})$



A mesma independência vale se eu não tenho uma cavidade:

- $P(\text{EstileteAgarrar} \mid \text{DorDente}, \neg\text{Cavidade}) = P(\text{EstileteAgarrar} \mid \neg\text{Cavidade})$

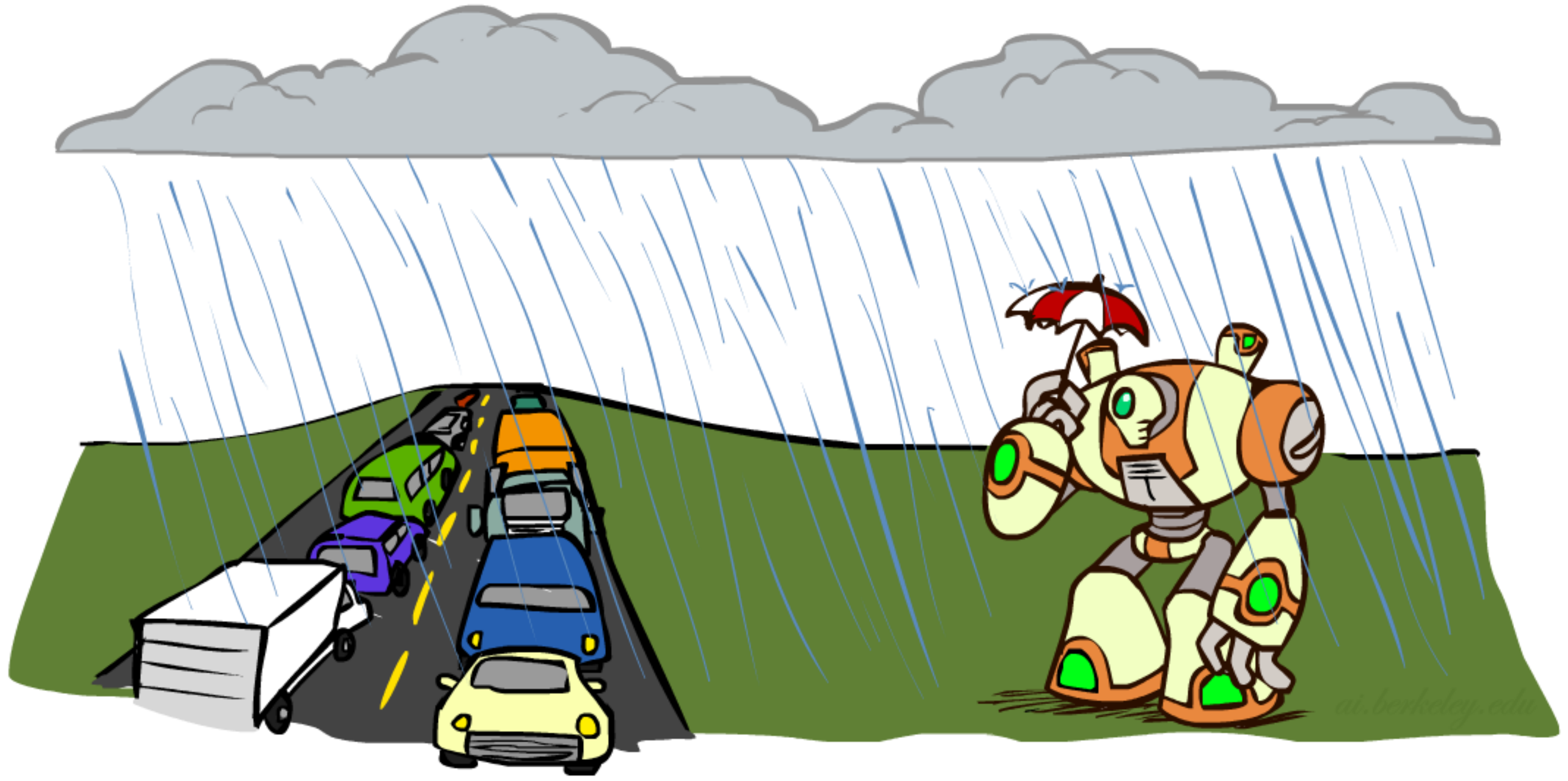
EstileteAgarrar é *condicionalmente independente* de DorDente dado Cavidade

Declarações equivalentes:

- $P(\text{DorDente} \mid \text{EstileteAgarrar}, \text{Cavidade}) = P(\text{DorDente} \mid \text{Cavidade})$
- $P(\text{DorDente}, \text{EstileteAgarrar} \mid \text{Cavidade}) = P(\text{DorDente} \mid \text{Cavidade}) P(\text{EstileteAgarrar} \mid \text{Cavidade})$

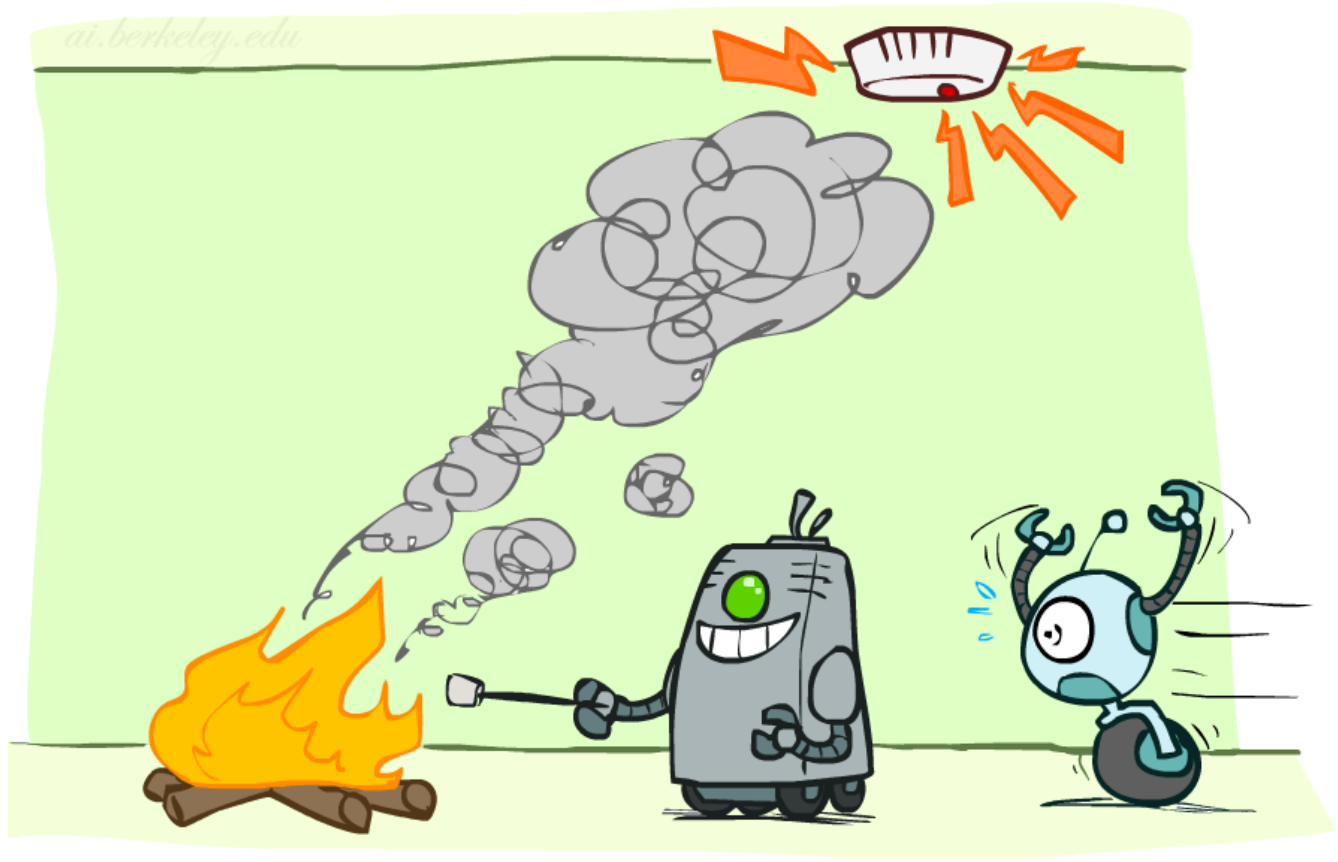
Exercício 1

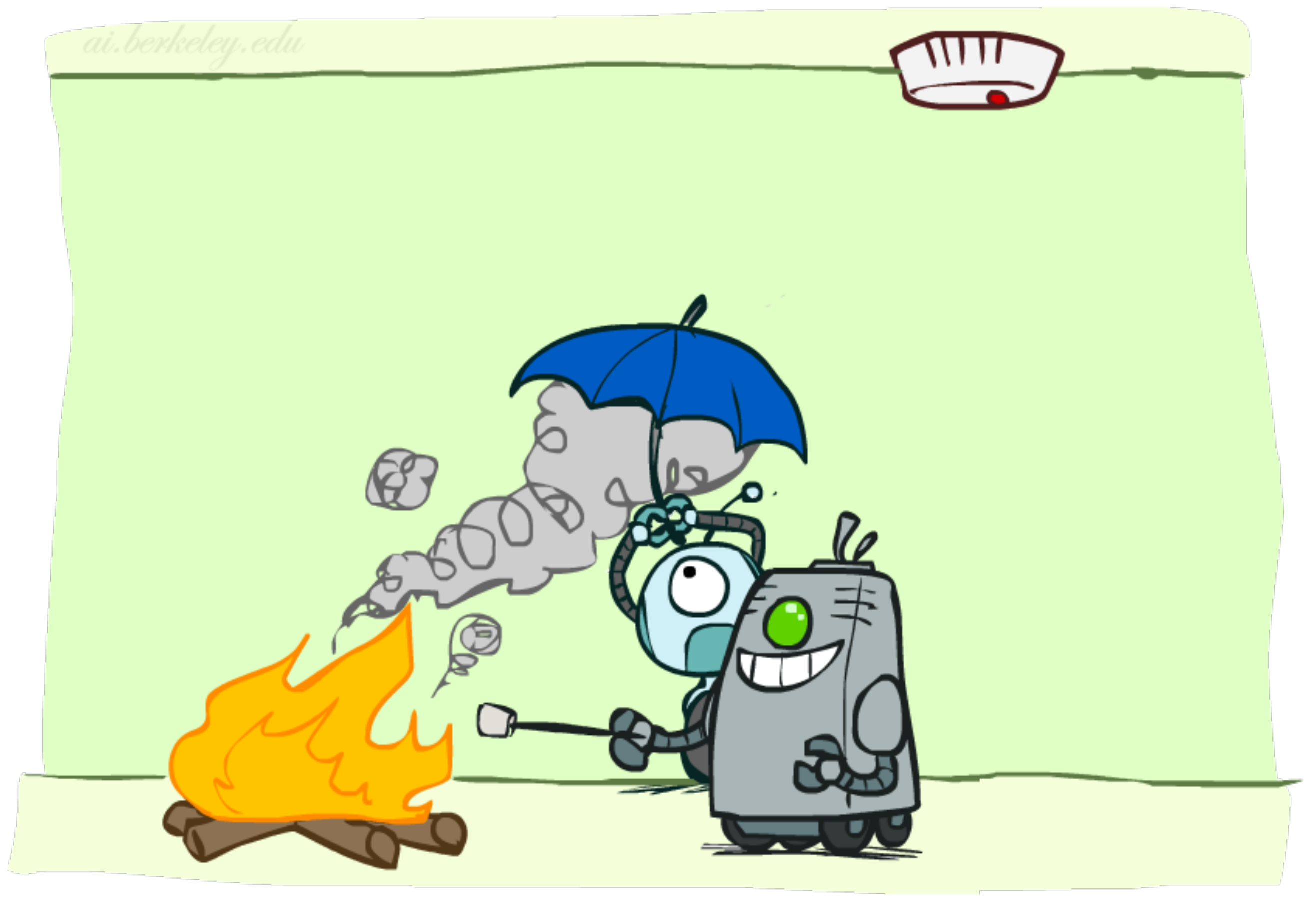
- Trânsito
- Guarda-chuvas
- Chuva



Exercício 2

- Fogo
- Fumaça
- Alarme



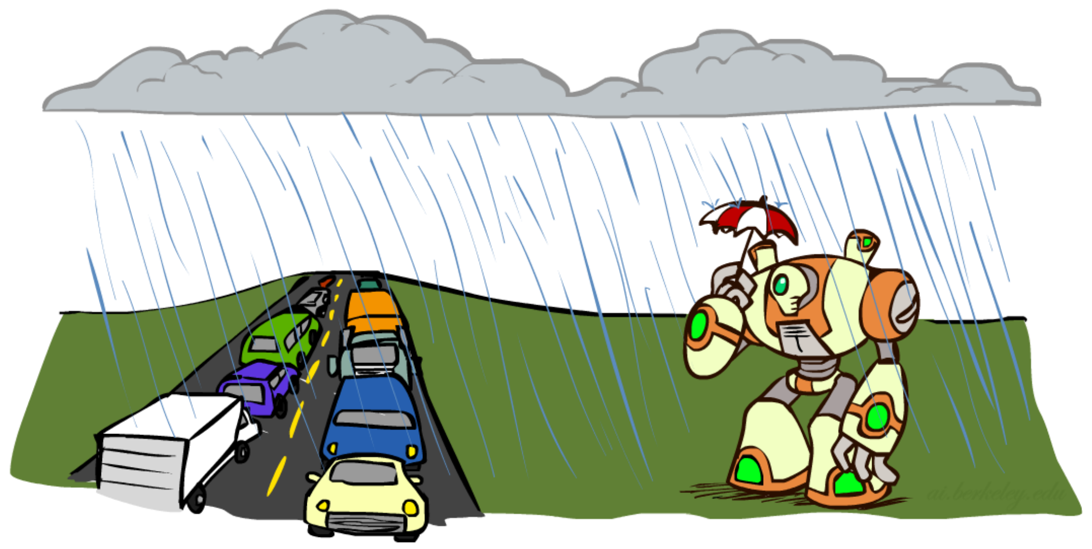


Independência condicional e a regra da cadeia

- Regra da cadeia: $P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1, X_2)\ldots$

- Aplicação no Exercício 1 (slide 11):

$$P(\text{Trânsito, Chuva, GuardaChuva}) = P(\text{Chuva})P(\text{Trânsito}|\text{Chuva})P(\text{GuardaChuva} \mid \text{Chuva, Trânsito})$$

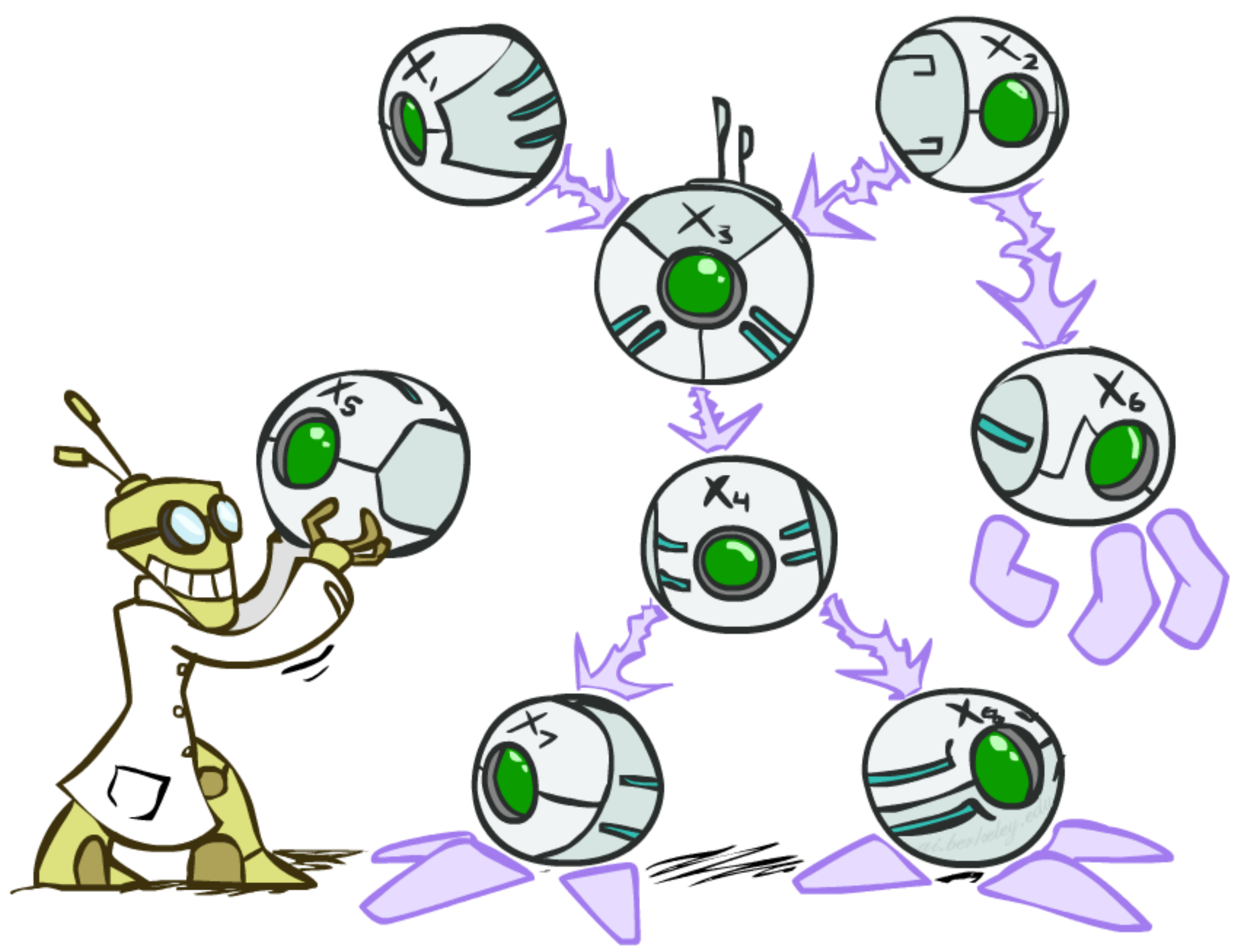


- Supondo independência condicional

$$P(\text{Trânsito, Chuva, GuardaChuva}) = P(\text{Chuva})P(\text{Trânsito}|\text{Chuva})P(\text{GuardaChuva} \mid \text{Chuva})$$

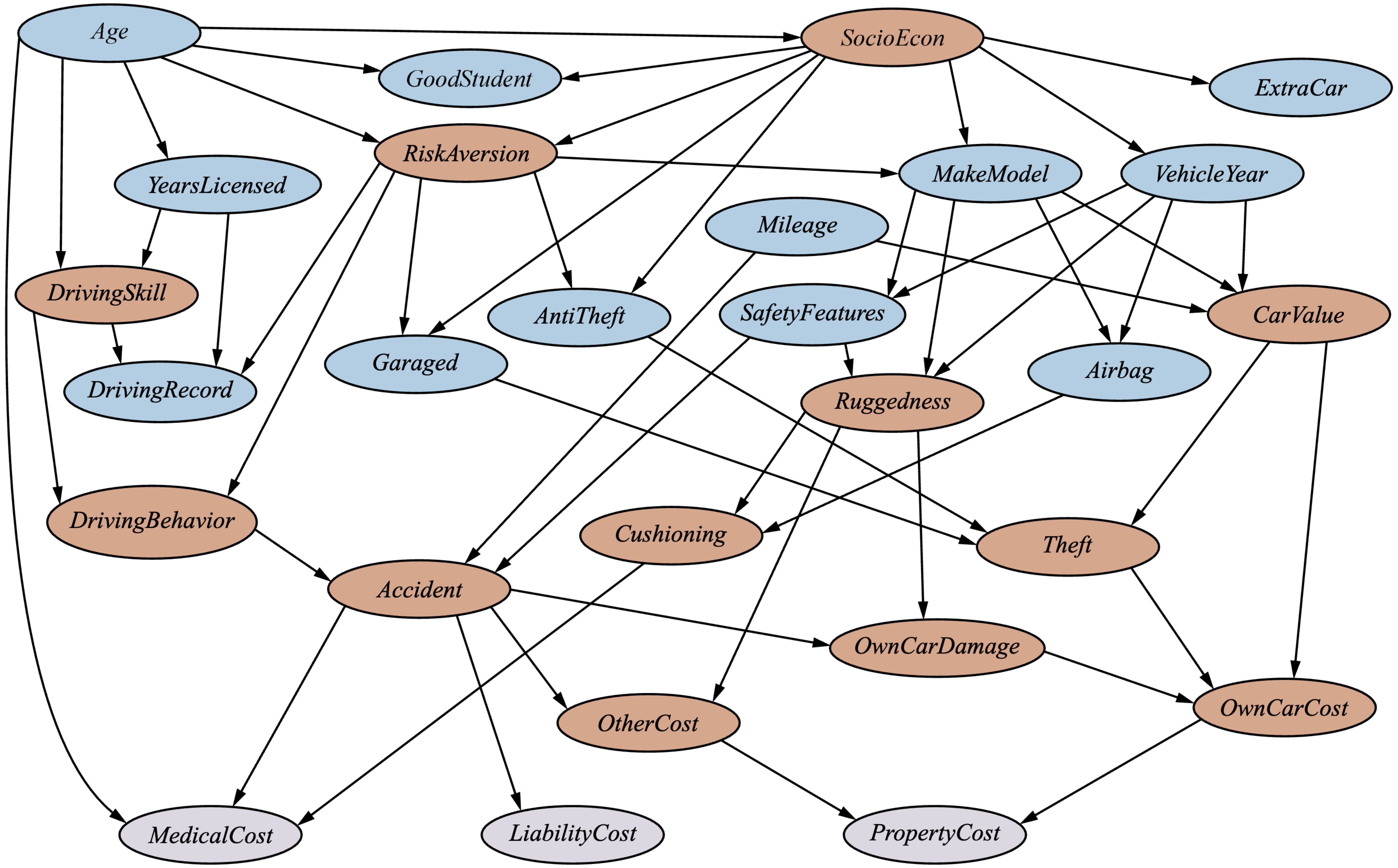
Redes Bayesianas nos ajudam a expressar suposições de independência condicional!

Redes bayesianas



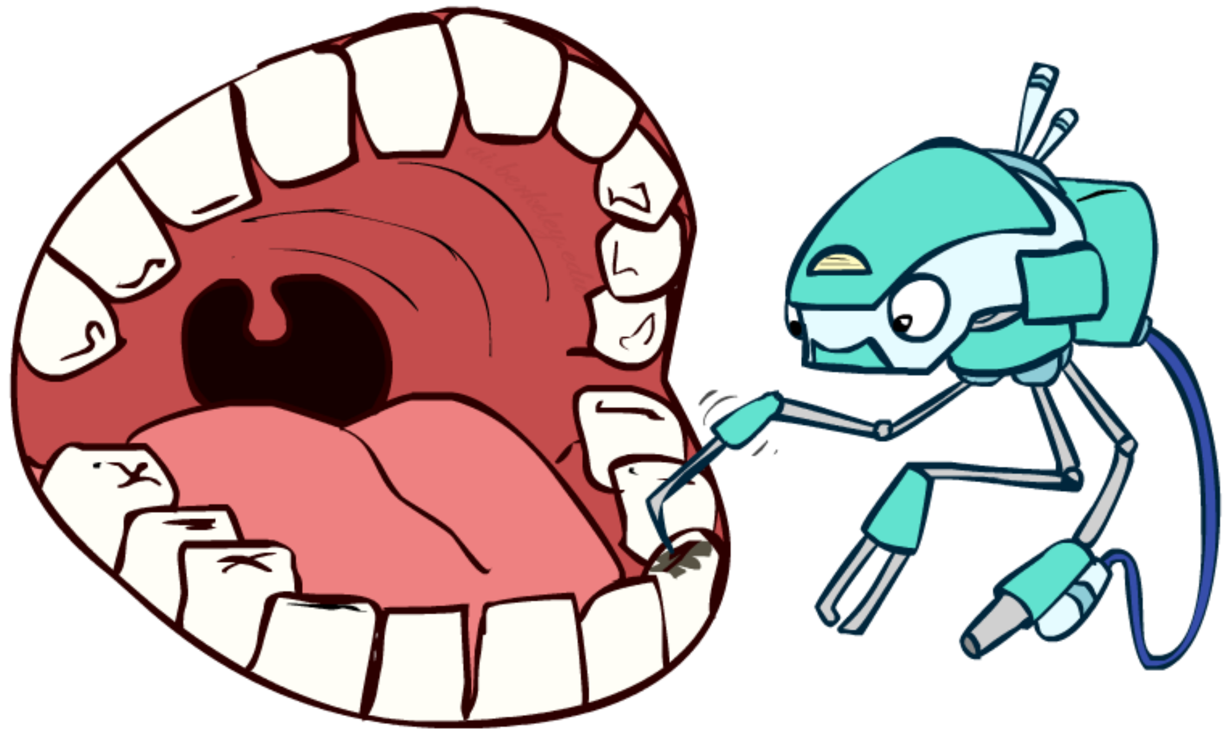
- Problemas em usar distribuições conjuntas $P(X_1, X_2, \ldots, X_n)$ como modelo de probabilidades:
  - A não ser que o problema tenha poucas variáveis, a distribuição conjunta é muito grande para representar explicitamente
  - Quanto maior o número de variáveis, mais difícil é estimar os valores das variáveis empiricamente
- **Redes bayesianas** são uma técnica para descrever distribuições conjuntas (modelos probabilísticos) complexa(o)s em termos de distribuições locais mais simples:
  - Também chamadas de modelos gráficos
  - Descreve como variáveis interagem localmente
  - Interações locais são combinadas para descrever interações globais

Exemplo 4: solicitações de seguro de carro



Redes bayesianas: sintaxe

- **Vértices** representam variáveis (com seus domínios)
  - Podem estar atribuídas (observadas) ou não
- **Arestas** representam interações entre variáveis
  - Similar às restrições dos PSRs
  - Indicam "influência direta" entre variáveis
  - Formalmente codificam independencia condicional

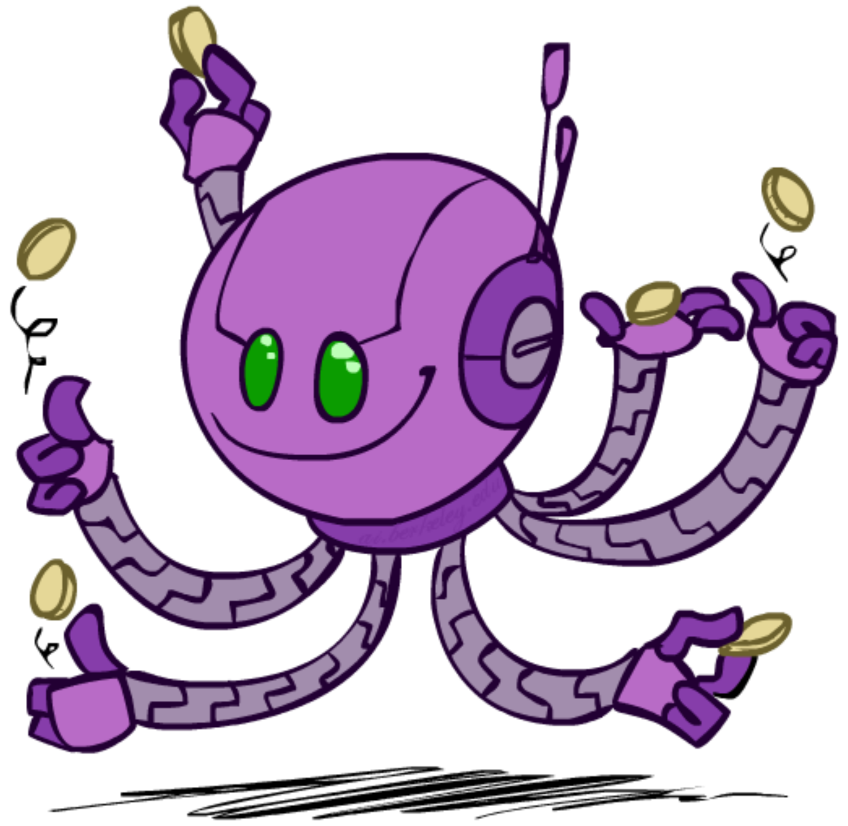


```mermaid
graph TD
    Cavidade --> DorDente
    Cavidade --> EstileteAgarrar
```

Exemplo 5: redes bayesianas

Lançamento de n moedas justas

$$X_1 \qquad X_2 \qquad \cdots \qquad X_n$$



Nenhuma interação entre variáveis: **independência absoluta!**

Exemplo 6: redes bayesianas

Ocorrência de trânsito veicular em cidades

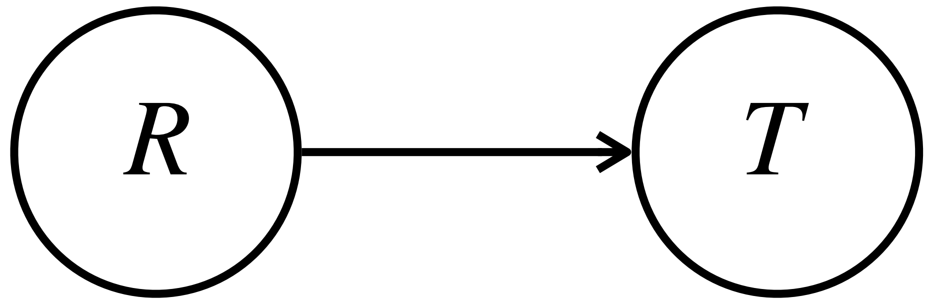
Variáveis:

- R: está chovendo
- T: ocorrência de trânsito

Modelo 1: assumindo independência

Modelo 2: chuva causa trânsito

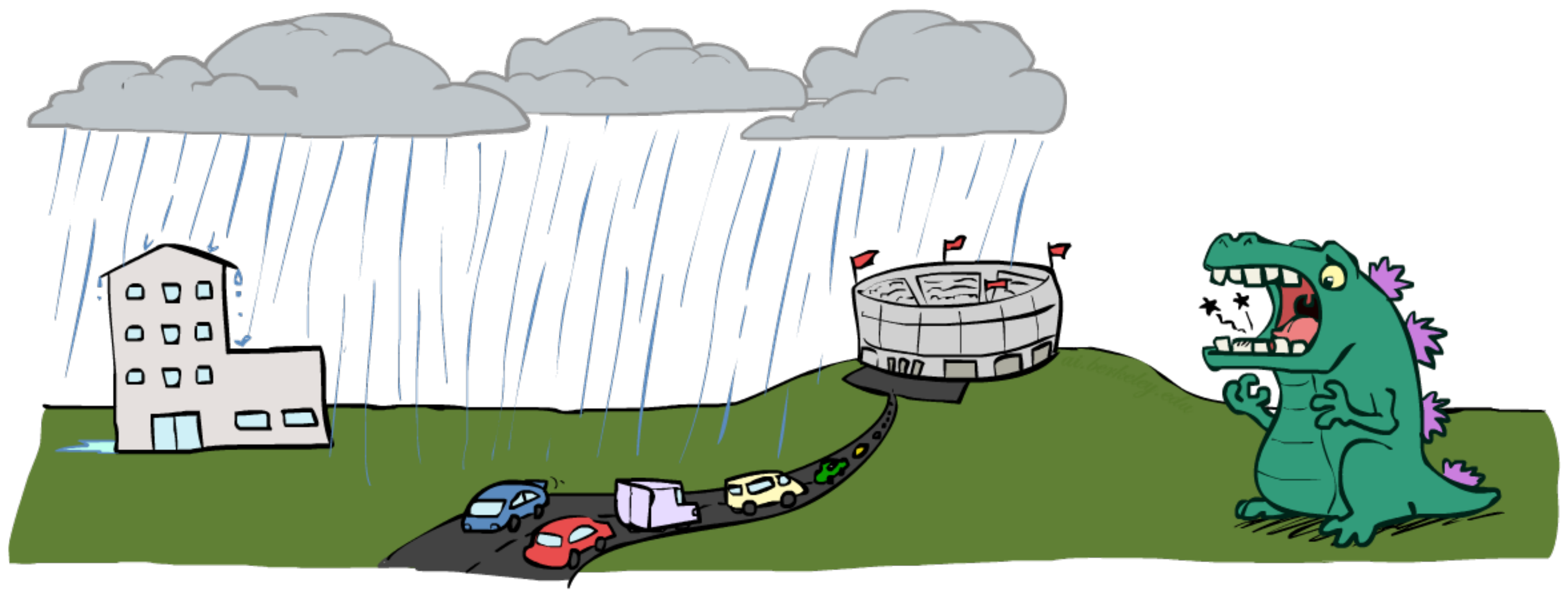
Porquê um agente que usa o **modelo 2** é melhor?



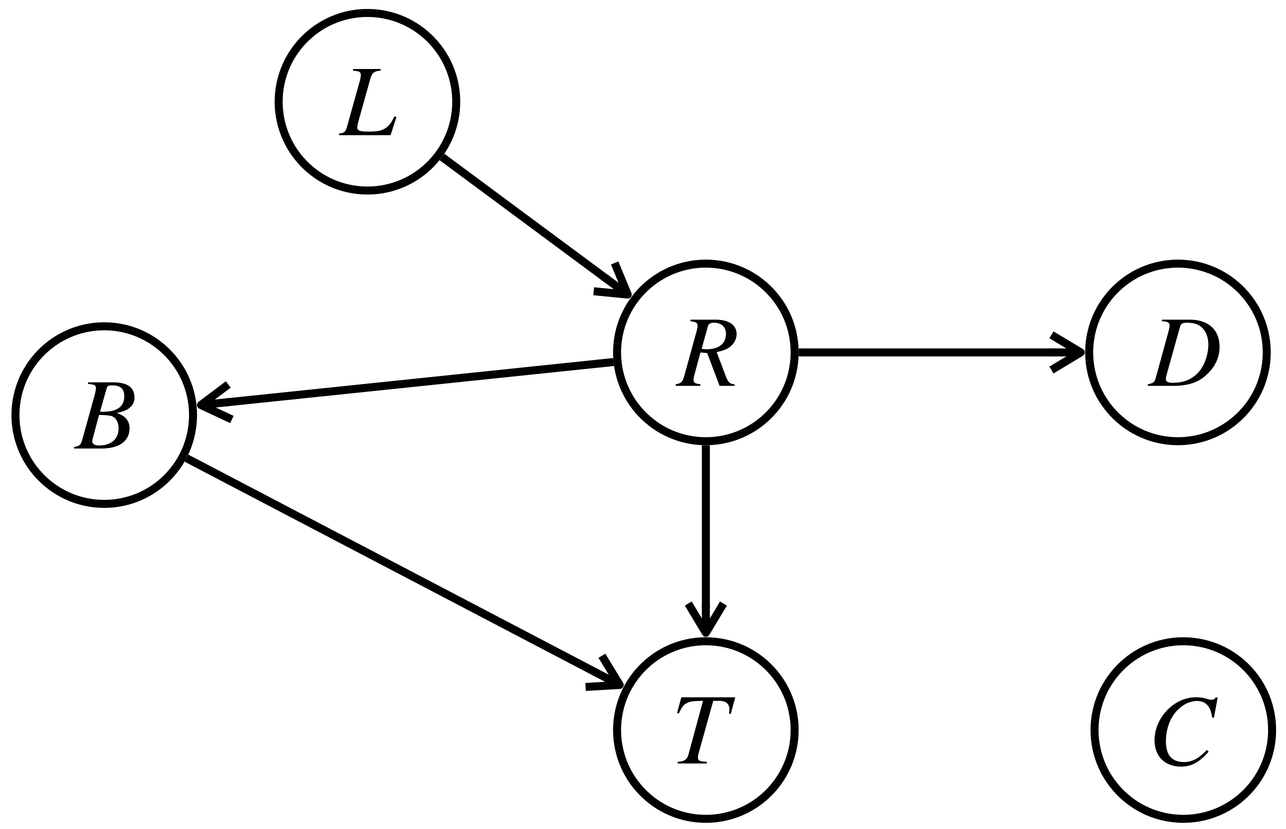
Exemplo 7: redes bayesianas

Construir uma rede bayesiana para modelar a ocorrência de trânsito considerando as seguintes variáveis aleatórias:

- T: ocorrência de trânsito
- R: está chovendo
- L: baixa pressão atmosférica
- D: vazamento no teto
- B: jogo de futebol
- C: cavidade



Não existe um único modelo correto. O modelo depende de como acreditamos que as variáveis influenciam umas as outras!

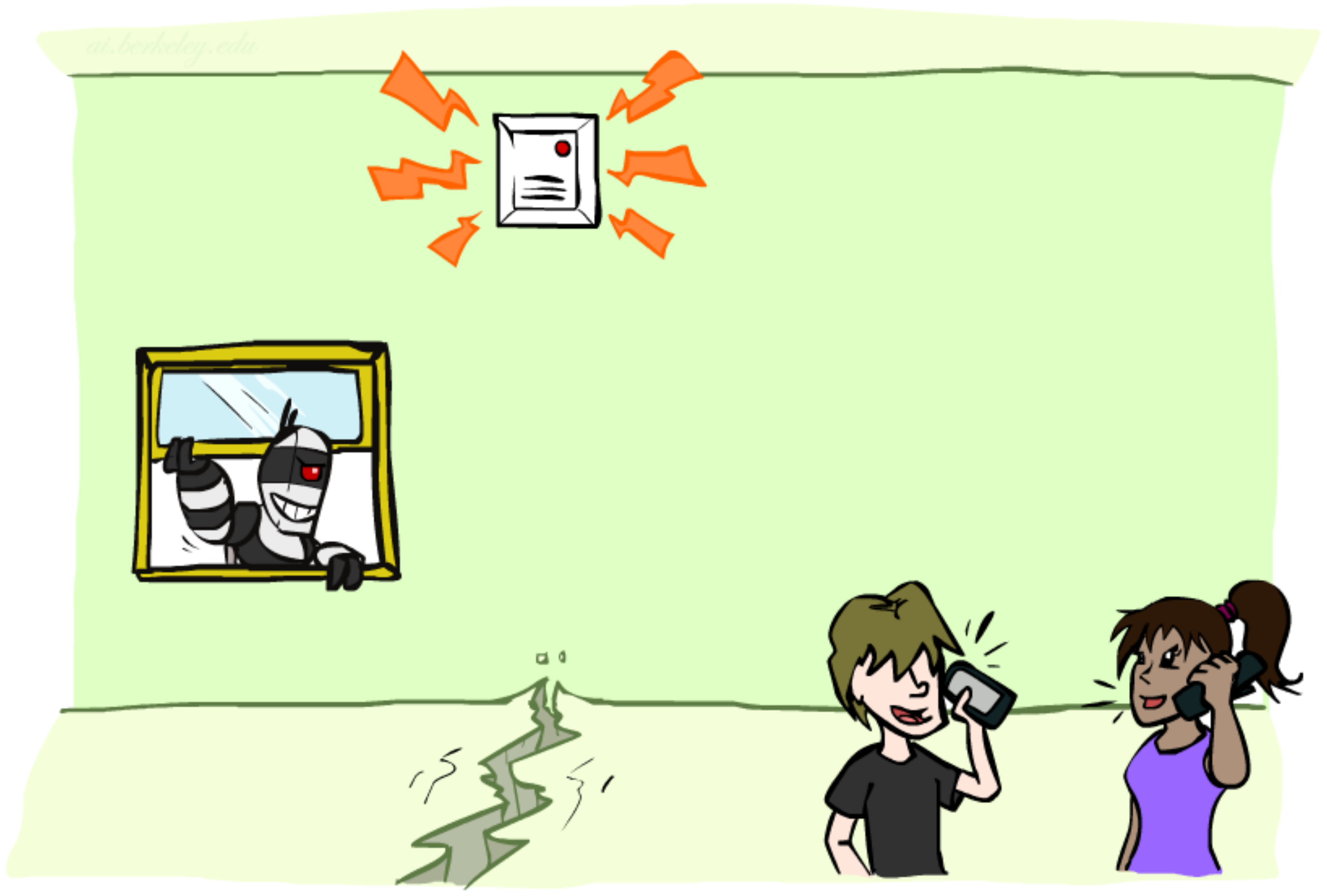


UFV

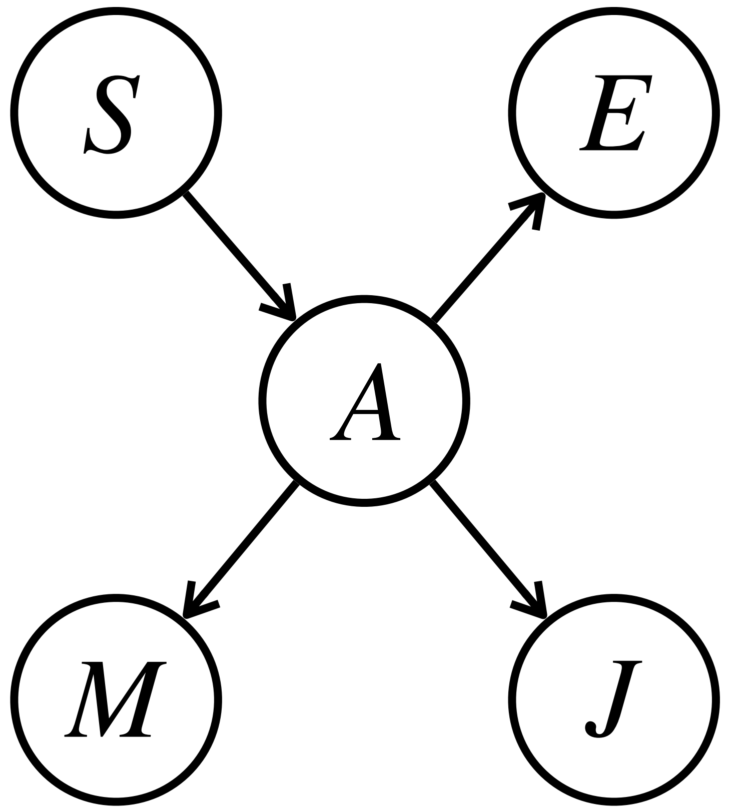
Exemplo 8: redes bayesianas

Alarme para detecção contra assaltos:

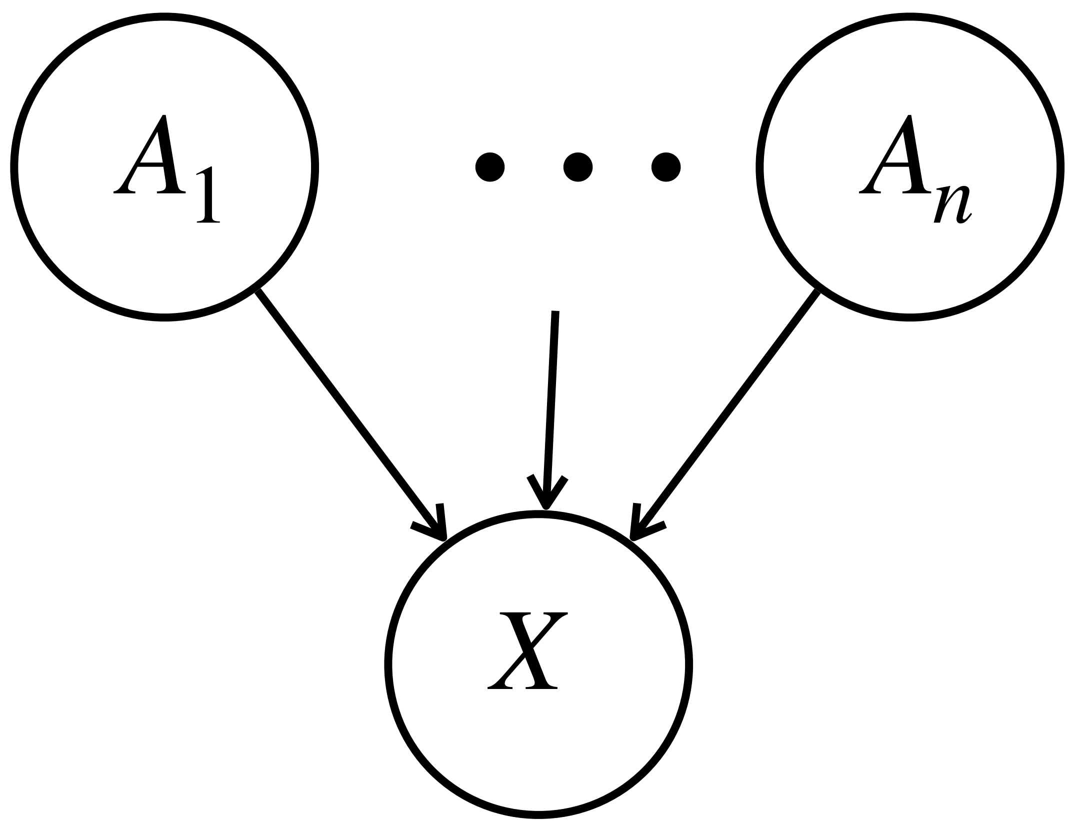
- A: alarme
- S: assalto
- M: ligação de Maria
- J: ligação de João
- E: terremotos



Não existe um único modelo correto. O modelo depende de como acreditamos que as variáveis influenciam umas as outras!



Redes bayesianas: semântica



- Uma rede bayesiana é um grafo acíclico dirigido com:
  - Um vértice para cada variável aleatória
  - Uma aresta entre cada par de variáveis dependentes
  - Uma distribuição condicional (tabela) para cada vértice:

$$P(X \mid a_1, \ldots, a_n)$$

- Uma rede bayesiana representa uma distribuição conjunta implicitamente:

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

UFV

Exemplo 5: redes bayesianas

Lançamento de n moedas justas

X_1 $\quad X_2 \quad \cdots \quad X_n$

$P(X_1)$

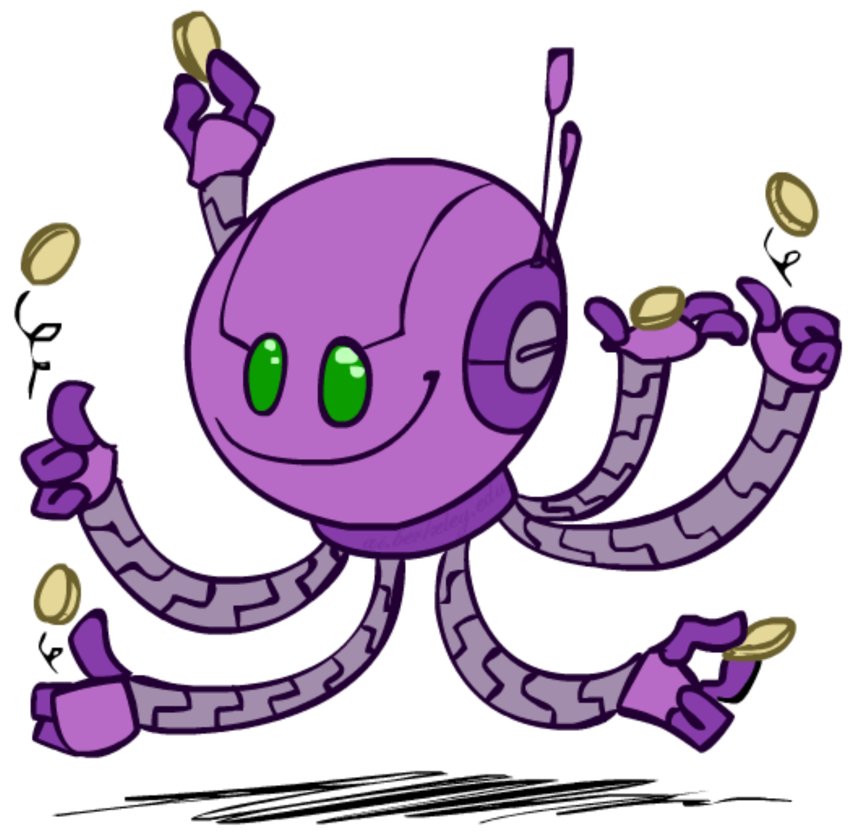
X1	P
Cara	1/2
Coroa	1/2

$P(X_2)$

X2	P
Cara	1/2
Coroa	1/2

$P(X_n)$

Xn	P
Cara	1/2
Coroa	1/2



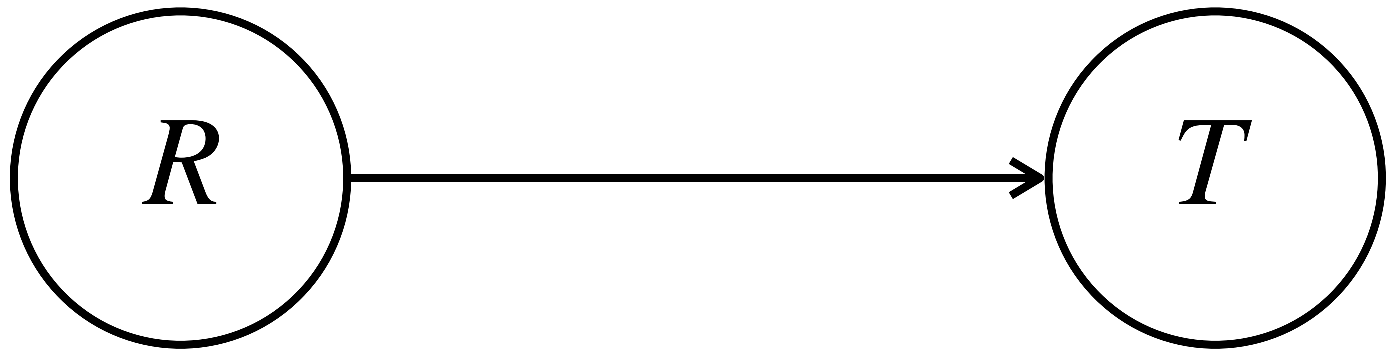
$$P(h, h, t, h) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 0{,}0625$$

Exemplo 6: redes bayesianas

Ocorrência de trânsito veicular em cidades

Variáveis:

- R: está chovendo
- T: ocorrência de trânsito



$P(R)$

R	P
+chuva	1/2
-chuva	1/2

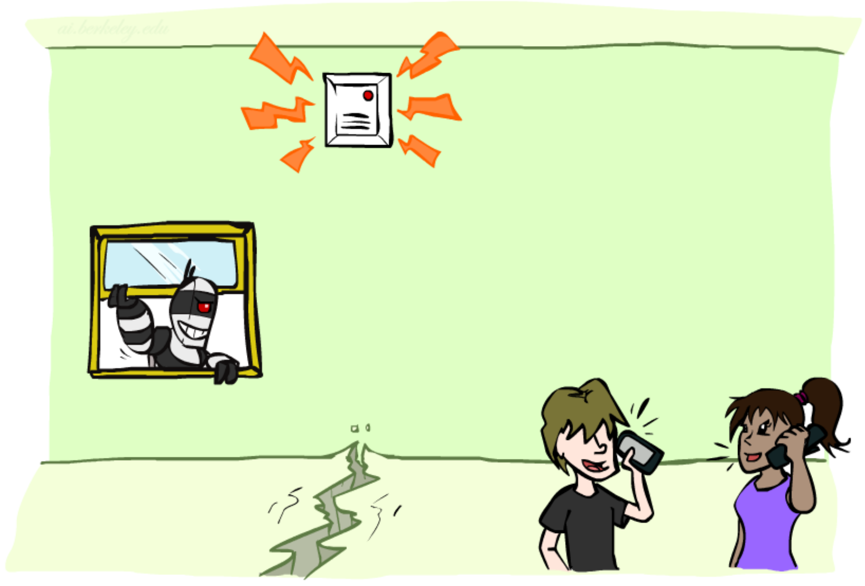
$P(T \mid R)$

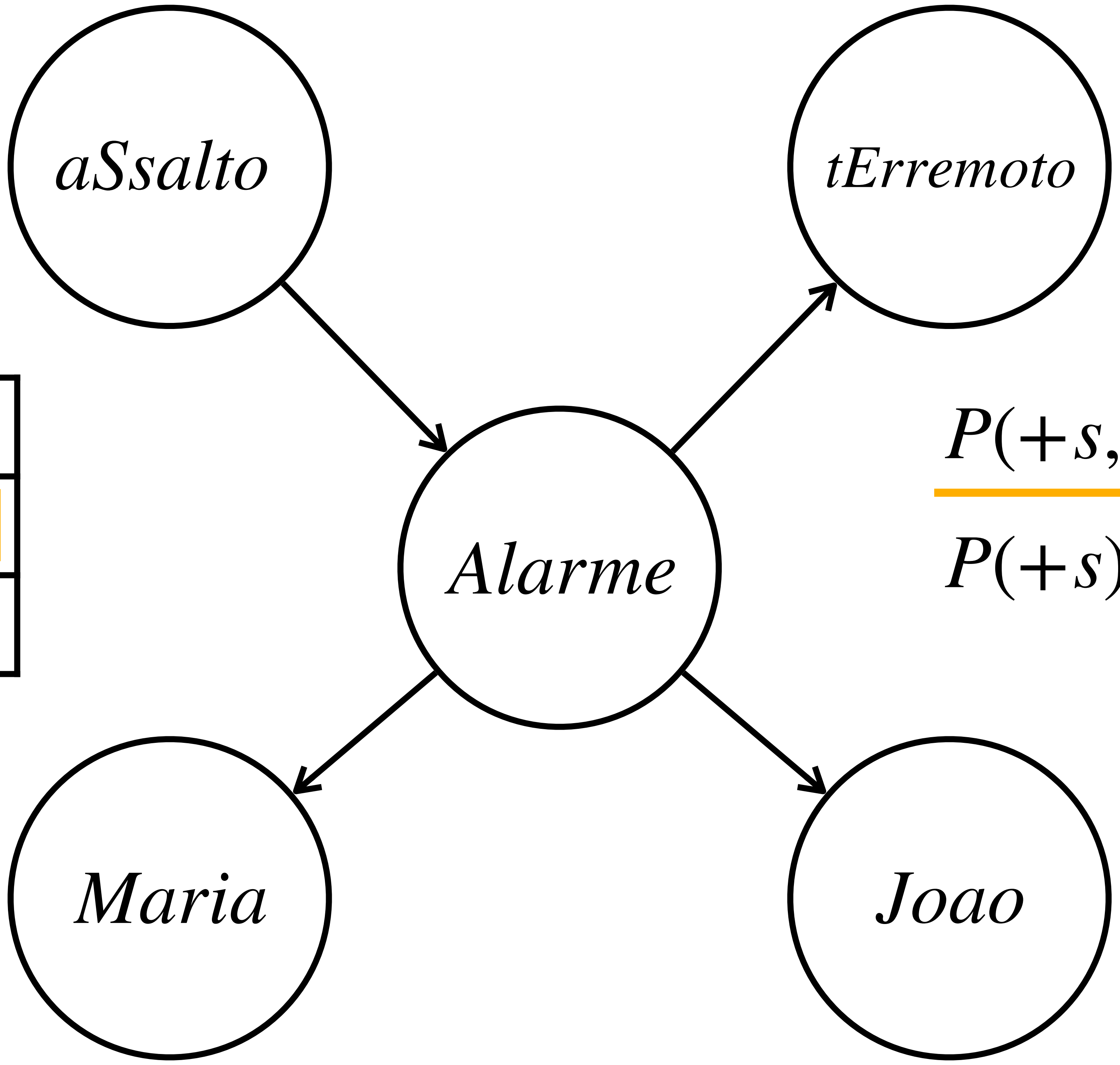
R	T	P
+chuva	+trânsito	3/4
+chuva	-trânsito	1/4
-chuva	+trânsito	1/2
-chuva	-trânsito	1/2

$$P(+chuva, -transito) = P(+chuva)P(-transito \mid +chuva) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$$

Exemplo 8: redes bayesianas

- A: alarme
- S: assalto
- M: maria
- J: joão
- E: terremotos





E	P(E)
+e	0,002
-e	0,998

S	P(S)
+s	0,001
-s	0,999

$$\underline{P(+s, -e, +a, +m, +j)} =$$

$$P(+s)\,P(-e)P(+a\,|-e, +s)P(+a\,|+m)P(+a\,|+j)$$

S	E	A	P(A|B,E)
+s	+e	+a	0,95
+s	+e	-a	0,05
+s	-e	+a	0,94
+s	-e	-a	0,06
-s	+e	+a	0,29
-s	+e	-a	0,71
-s	-e	+a	0,001
-s	-e	-a	0,999

A	M	P(M|A)
+a	+m	0,7
+a	-m	0,3
-a	+m	0,01
-a	-m	0,99

A	J	P(J|A)
+a	+j	0,9
+a	-j	0,1
-a	+j	0,05
-a	-j	0,95

Inferência por enumeração

- A: alarme
- S: assalto
- M: maria
- J: joão
- E: terremotos

Dada uma rede bayesiana, queremos calcular a probabilidade condicional $P(Q \mid e_1, ..., e_k)$, onde:

- Q é uma variável aleatória definindo uma consulta
- $E_1, ..., E_k = e_1, ..., e_k$ é um conjunto de variáveis aleatórias definindo a evidência
- Além disso, $H_1, ..., H_r$ é o conjunto de variáveis "escondidas" do modelo, não inclusas na consulta nem na evidência

Inferência por enumeração:

1) **Selecionar** as entradas consistentes com a evidência
2) **Somar** as variáveis em H para obter $P(Q, e_1, ..., e_k)$

$$P(Q, e_1, ..., e_k) = \sum_{h_1, ..., h_r} P(Q, h_1, ..., h_r, e_1, ..., e_k)$$

3) **Normalizar** $P(Q, e_1, ..., e_k)$ para obter $P(Q \mid e_1, ..., e_k)$

$$P(Q \mid e_1, ..., e_k) = \frac{P(Q, e_1, ..., e_k)}{\sum_q P(Q, e_1, ..., e_k)}$$

Exemplo (alarme):

$$P(S \mid +j, +m)$$

$$P(S, +j, +m) = \sum_{e,a} P(S, e, a, +j, +m)$$

$$P(S, +j, +m) = \sum_{e,a} P(S)P(e)P(a \mid S, e)P(+j \mid s)P(+m \mid s)$$

$$P(S \mid +j, +m) = \frac{P(S, +j, +m)}{\sum_s P(S, +j, +m)}$$

UFV

Próxima aula

A16: Raciocínio Probabilístico III

Processos (cadeias) de Markov, modelos de transição, inferência, amostragem, distribuição estacionária