

# Aula 9 – Tratamento de Incerteza

22705/1001336 - Inteligência Artificial  
2019/1 - Turma A  
Prof. Dr. Murilo Naldi

[naldi@dc.ufscar.br](mailto:naldi@dc.ufscar.br)

# Agradecimentos

- Parte do material utilizado nesta aula foi cedido pelos professores André C.P.L.F de Carvalho e Ricardo J.G.B. Campello e, por esse motivo, o crédito deste material é deles
- Parte do material utilizado nesta aula foi disponibilizado por M. Kumar no endereço:
  - [www-users.cs.umn.edu/~kumar/dmbook/index.php](http://www-users.cs.umn.edu/~kumar/dmbook/index.php)
- Agradecimentos a Intel Software e a Intel IA Academy pelo material disponibilizado e recursos didáticos

Copyright © 2017, Intel Corporation. All rights reserved.



# O que é incerteza?

- Agentes lógicos admitem que proposições sejam verdadeiras, falsas ou desconhecidas
- Quase nunca é possível conhecer a verdade sobre todo o ambiente.
- É preciso trabalhar com incerteza.
- Exemplo:
- O objetivo “ir para o aeroporto em 40 minutos” pode não ser alcançado caso o agente fique sem combustível ou encontre tráfego.

# Tratamento do incerto

- Utilizar LPO para tratar incerteza é falho devido à quantidade de causas (conhecidas e desconhecidas) possíveis para um determinado fenômeno
- Podemos então utilizar a teoria de probabilidade, para expressar grau de crença
- Grau de crença são atribuído por meio de números entre 0 (falsa) e 1 (verdadeira)

# Incerteza e Probabilidade

- Variável aleatória
  - Representa “algo” (objetos, fatos, ações, crenças,...)
    - Seu valor é inicialmente desconhecido
  - Elas podem possuir domínios:
    - Booleano: como Feliz com domínio <verdadeiro, falso>
    - Discreto: como Clima com domínio <ensolarado, nublado, chuvoso>
    - Contínuo: como Peso com domínio nos números reais positivos

# Variável aleatória

- Geralmente, escrevemos o nome da variável aleatória com a primeira letra maiúscula quando não sabemos o seu valor
- Adicionalmente, escrevemos os valores (estados) que podem ser assumidos por estas variáveis em letras minúsculas
- Exemplo:

*Cárie* = verdadeiro ou *cárie*

*Cárie* = falso ou  $\neg$ cárie

# Axiomas de probabilidade

- Todas as probabilidades estão entre 0 e 1

$$0 \leq P(a) \leq 1$$

- Proposições verdadeiras tem probabilidade 1 e as falsas possuem 0

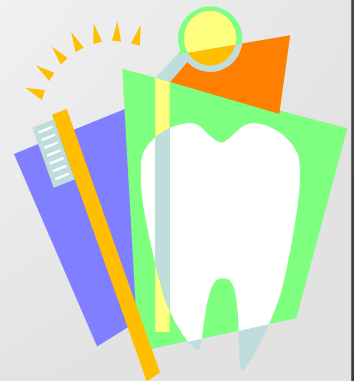
$$P(\text{verdadeiro}) = 1 \text{ e } P(\text{falso}) = 0$$

- A probabilidade de uma disjunção é:

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$

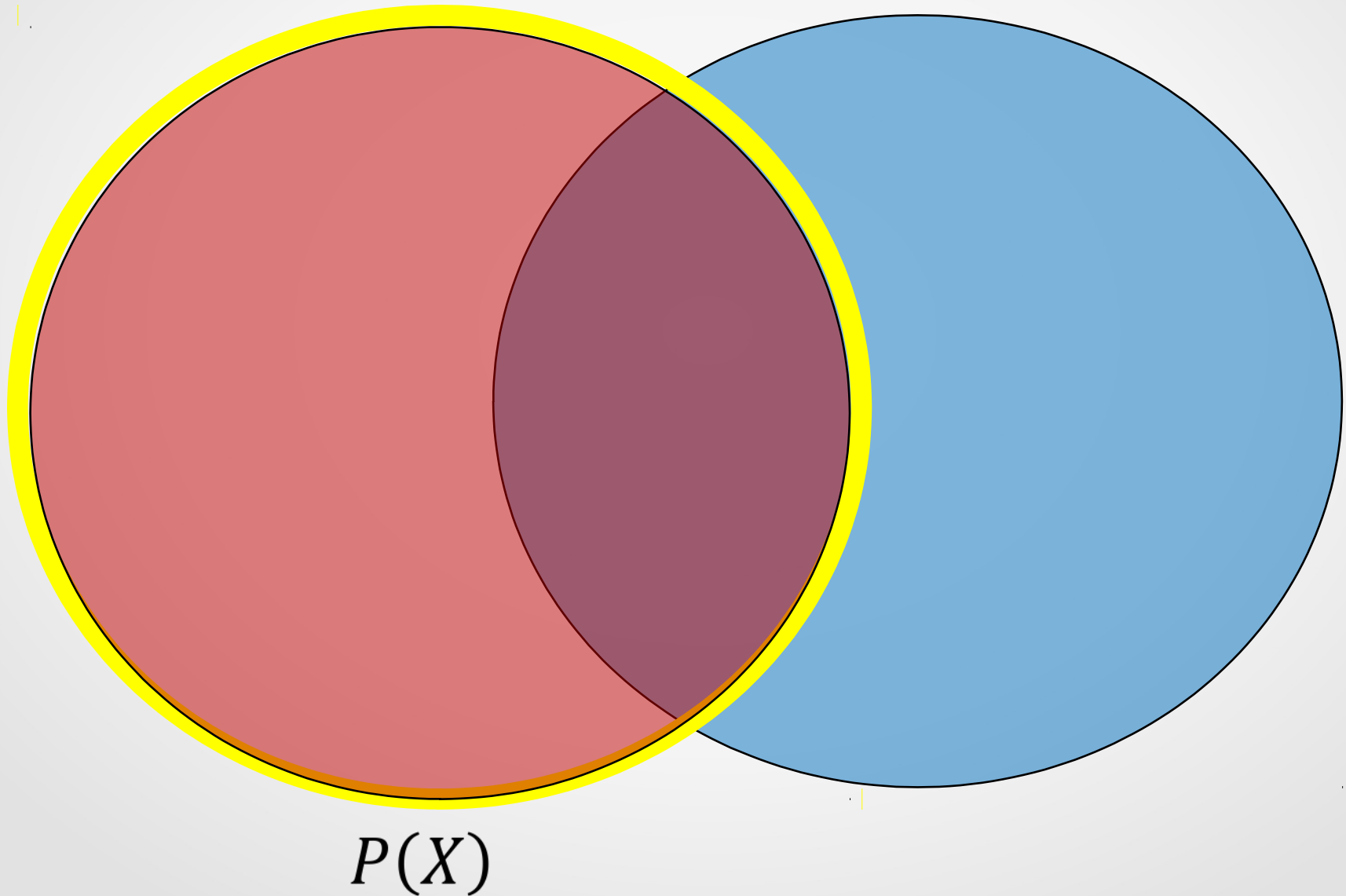
# Probabilidade *a priori*

- Ou probabilidade *incondicional*
- Consiste na probabilidade de uma proposição ocorrer na ausência de quaisquer outras informações
- Pode ser *simples* ou *conjunta*
- Exemplos:
  - $P(\text{Clima} = \text{ensolarado}) = 0,5$
  - $P(\text{cárie} \wedge \text{dor}) = 0,1$

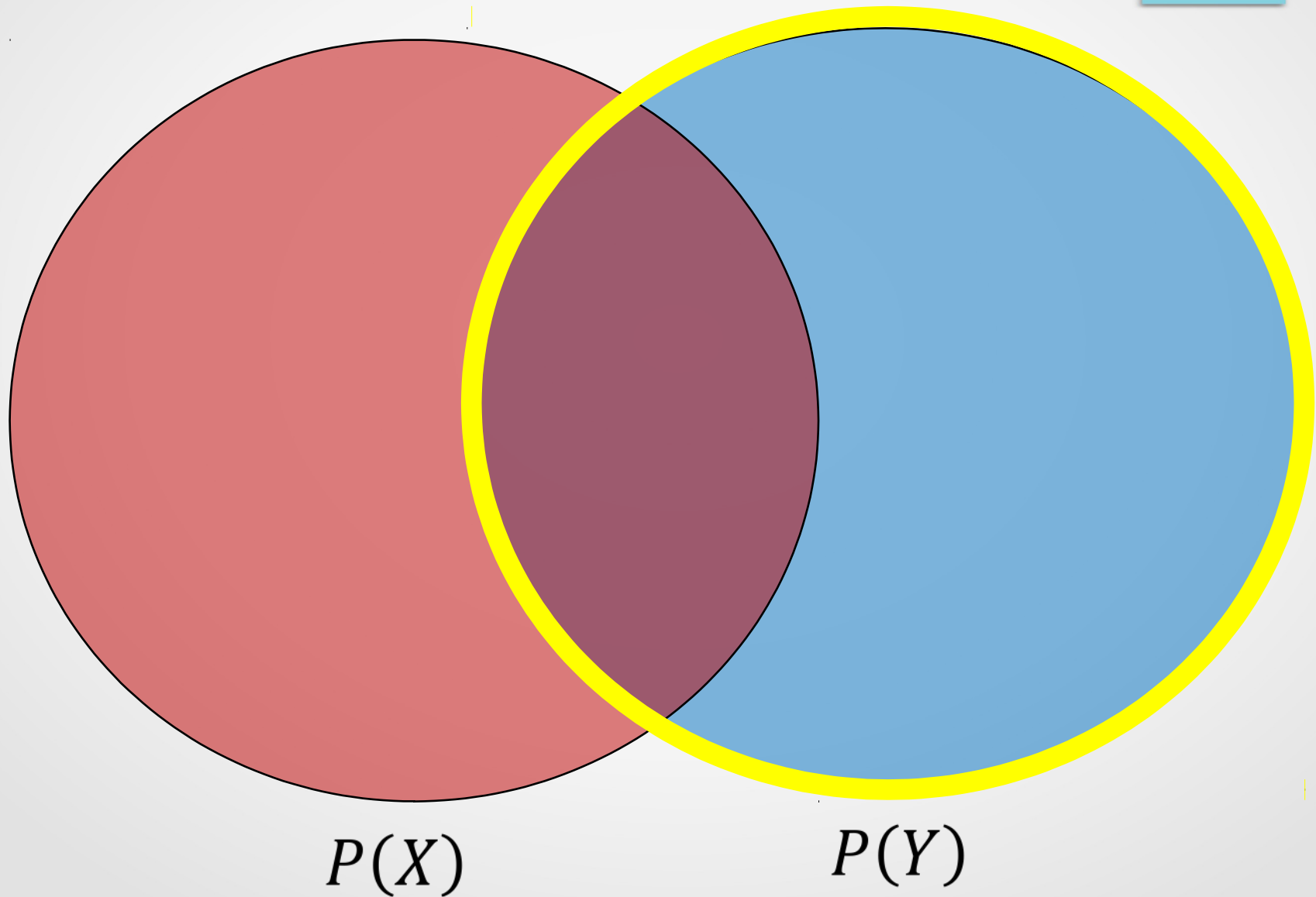




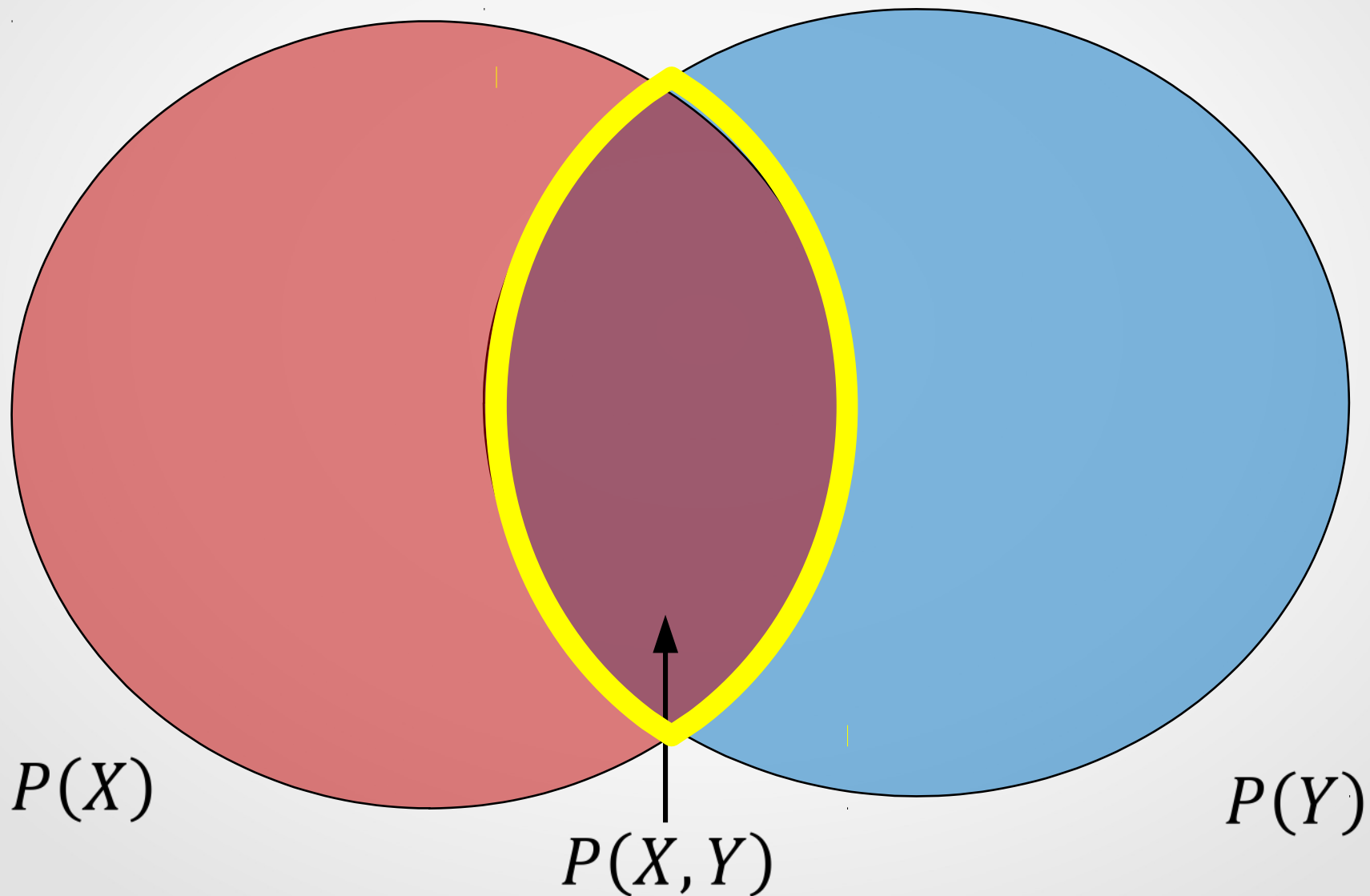
# Probabilidade *a priori*



# Probabilidade a *priori*



# Probabilidade *a priori* conjunta



# Distribuição de probabilidade

- São as probabilidades *a priori* de uma determinada variável aleatória
- Exemplo:
  - $P(\text{Clima}) = \langle 0,5, 0,2, 0,3 \rangle$   
significa que *Clima* possui probabilidade *a priori* 0,5 de ser *ensolarado*, 0,2 de estar *nublado* e 0,3 de estar *chuvoso*.



# Distribuição de probabilidade

- A distribuição de probabilidade conjunta é quando temos a probabilidade a priori de cada combinação de um conjunto de variáveis aleatórias
- Quando essas variáveis definem o ambiente, então ela é chamada de distribuição de probabilidade conjunta total

# Inferência

- Podemos utilizar a distribuição de probabilidade conjunta total como um modelo do ambiente

	<i>atrasado</i>		<i>¬atrasado</i>	
	<i>gasolina</i>	<i>¬gasolina</i>	<i>gasolina</i>	<i>¬gasolina</i>
<i>engarrafamento</i>	0,19	0,14	0,13	0,01
<i>¬engarrafamento</i>	0,06	0,07	0,25	0,15

# Marginalização

- Ou totalização consiste em calcular a probabilidade de um acontecimento a partir da base de conhecimento

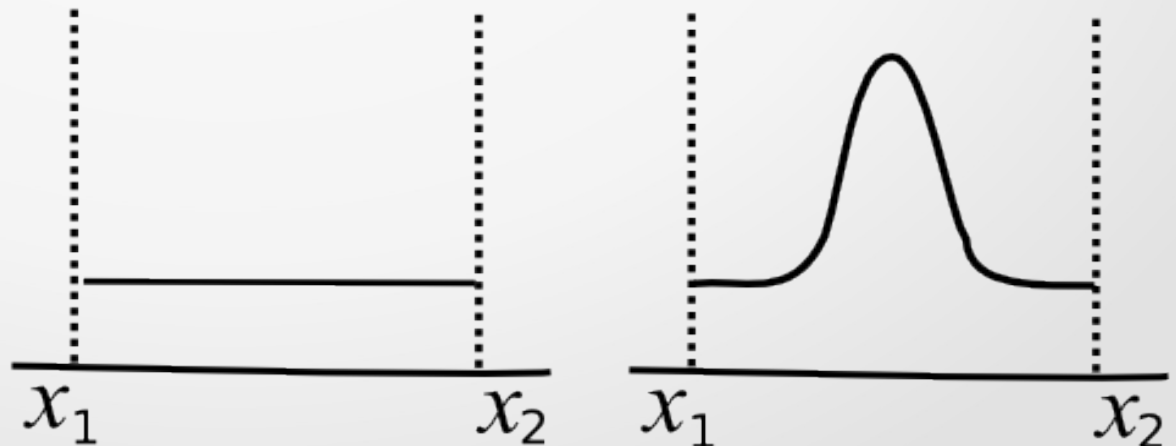
$$P(Y) = \sum_{z \in Z} P(Y, z)$$

- Exemplo: a probabilidade de ocorrer um *engarrafamento* é de 0,47 e não ocorrer é 0,53

<i>engarrafamento</i>	0,19	0,14	0,13	0,01
-----------------------	------	------	------	------

# Funções de densidade de probabilidade

- Variáveis contínuas não podem ser representadas em uma tabela pois seus valores são infinitos
- Para isso, utilizamos funções de densidade de probabilidade
- Exemplos:
  - uniforme
  - normal





# Exemplo

- Podemos definir que uma variável aleatória  $X$  que define a temperatura em uma manhã de outubro em São Carlos está *distribuída uniformemente* entre 18 e 26 °C

$$P(X=x) = U[18,26](x)$$

- Qual a probabilidade de fazer 20,5°C? Não é possível calcular a probabilidade em um ponto (zero). Então, usamos uma pequena região em torno dele.

$$\lim_{dx \rightarrow 0} P(20,5 \leq X \leq 20,5 + dx) / dx = 0,125$$

# Probabilidade condicional

- Conhecimento a respeito de variáveis aleatórias podem influenciar na probabilidade de outras
- Por exemplo, saber que um paciente possui dor de dente aumenta a chance deste paciente ter cárie

$$P(\text{cárie}|\text{dordedente}) = 0,9$$

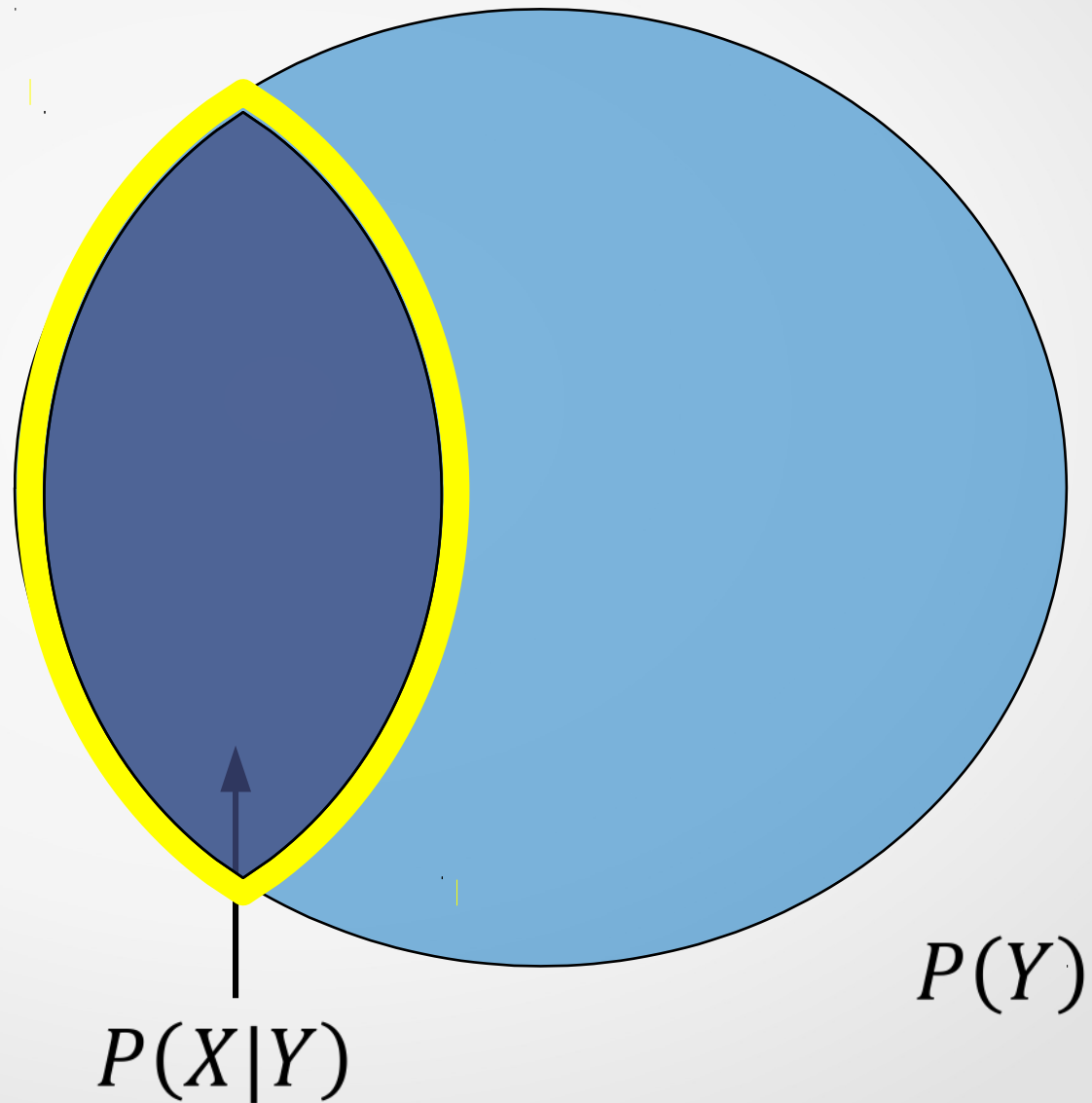
- Chamamos isso de probabilidade condicional

# Probabilidade condicional

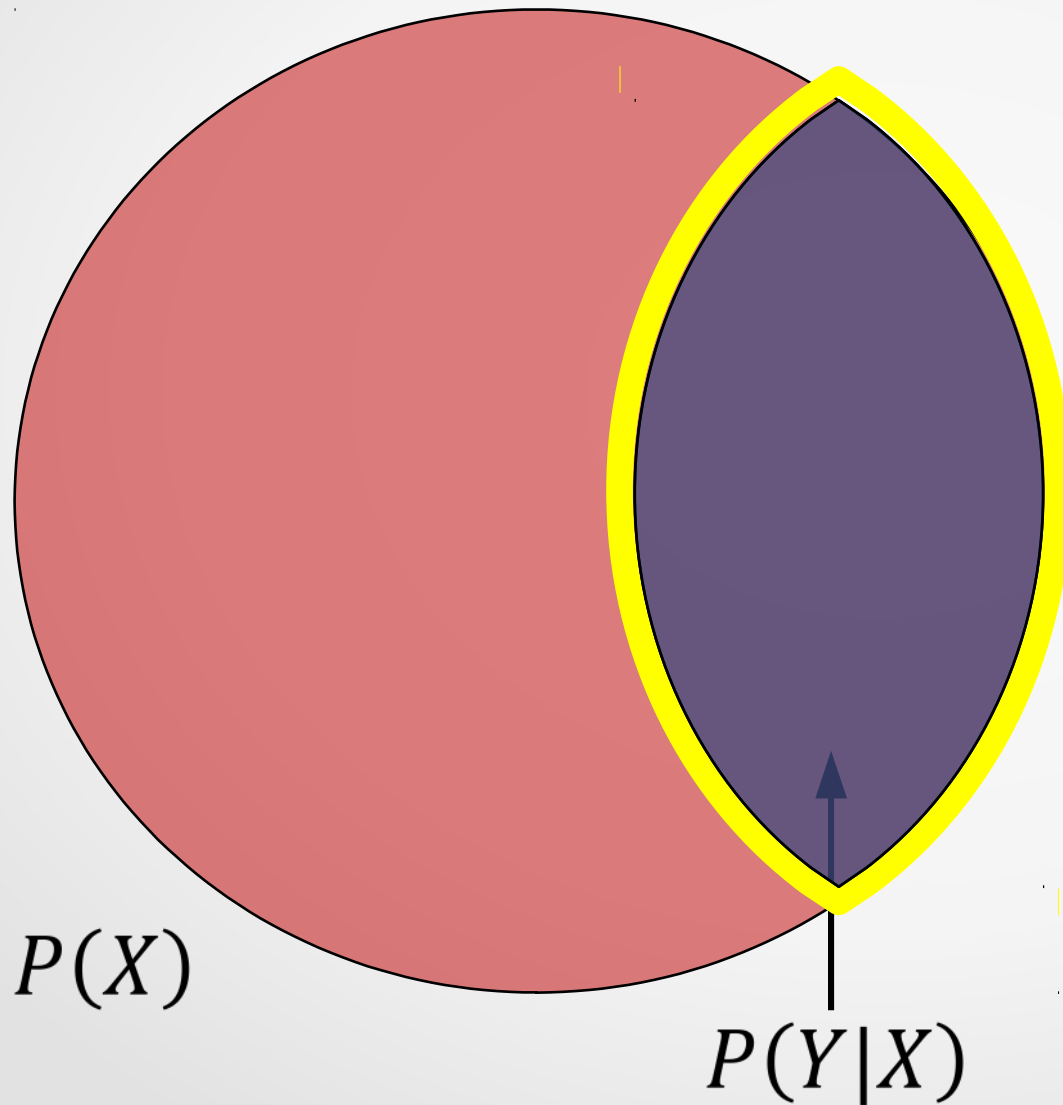
- Probabilidades condicionais podem ser calculadas a partir de probabilidades incondicionais

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

# Probabilidade Condicional



# Probabilidade Condicional



# Condicionamento

- Outra forma de obter a probabilidade *a priori* de uma variável é a partir de variáveis condicionais.
- Esse processo é conhecido como condicionamento.

$$P(Y) = \sum_z P(Y|z) P(z)$$

# Independência

- Ou independência absoluta
- Uma variável é independente de outra quando sua probabilidade condicional em relação à última é igual a sua probabilidade *a priori*, ou seja:

$$P(A|B) = P(A)$$

- Exemplo:
  - Imagine que adicionamos uma quarta variável a nossa base de conhecimento, a variável booleana *Apaixonado*

<i>apaixonado</i>	<i>atrasado</i>		$\neg$ <i>atrasado</i>	
	<i>gasolina</i>	$\neg$ <i>gasolina</i>	<i>gasolina</i>	$\neg$ <i>gasolina</i>
<i>engarrafamento</i>	0,145	0,07	0,06	0,005
$\neg$ <i>engarrafamento</i>	0,03	0,035	0,125	0,075

$\neg$ <i>apaixonado</i>	<i>atrasado</i>		$\neg$ <i>atrasado</i>	
	<i>gasolina</i>	$\neg$ <i>gasolina</i>	<i>gasolina</i>	$\neg$ <i>gasolina</i>
<i>engarrafamento</i>	0,145	0,07	0,06	0,005
$\neg$ <i>engarrafamento</i>	0,03	0,035	0,125	0,075



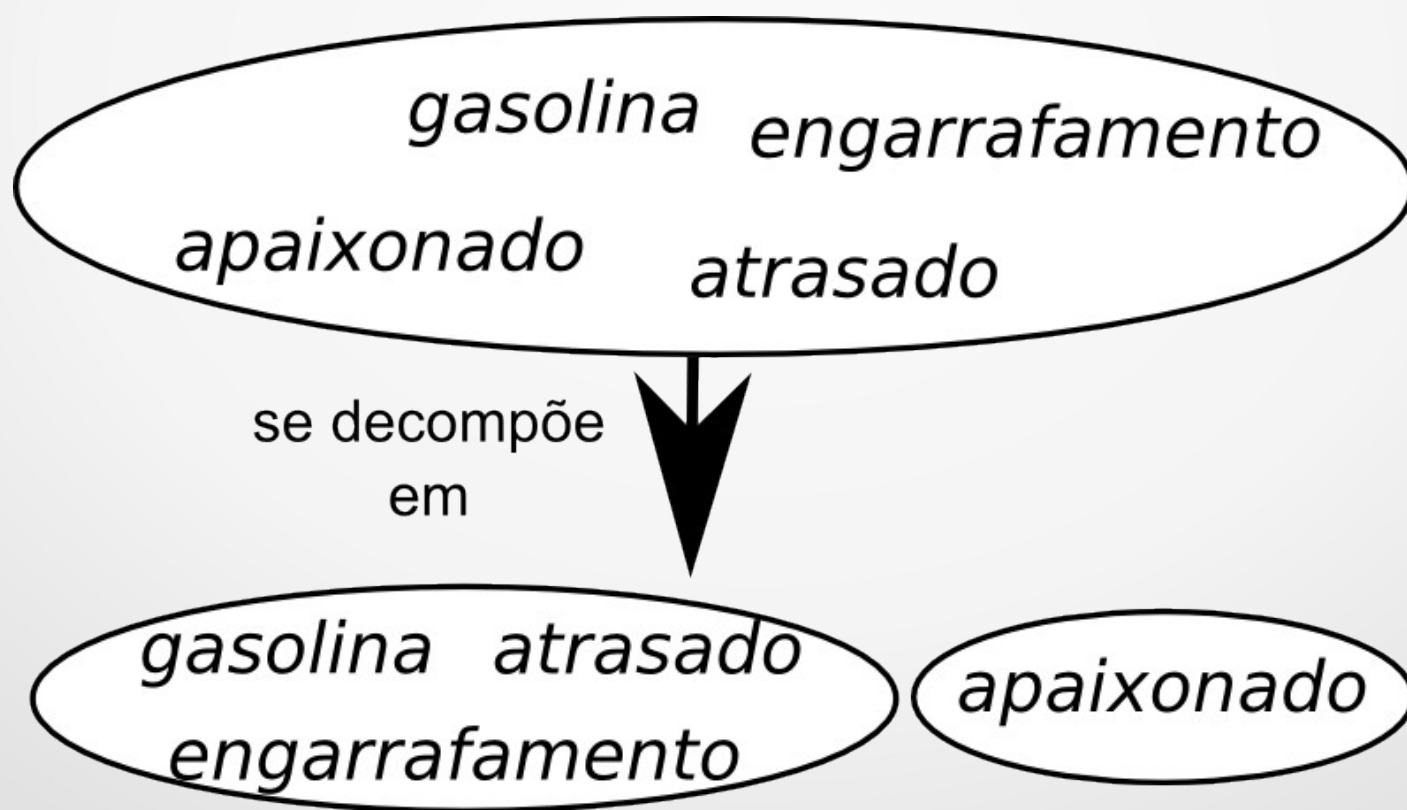
# Independência

- Se somarmos as probabilidades das duas tabelas, vemos que existe a probabilidade de 0,50 do agente estar apaixonado
- Verificando a independência de Engarrafamento e Apaixonado temos:

$$P(\text{engarrafamento}|\text{apaixonado}) = \frac{P(\text{engarrafamento} \wedge \text{apaixonado})}{P(\text{apaixonado})}$$

# Fatoração

- Consiste em reduzir a quantidade de informações necessárias para especificar a distribuição do conjunto total



# Fatoração

- A fatoração é uma característica muito importante.
  - O número de valores de probabilidades condicionais é combinatorial em relação ao número de variáveis dependentes
  - Independência e fatoração permitem ganho computacional, pois tratam variáveis independentemente.

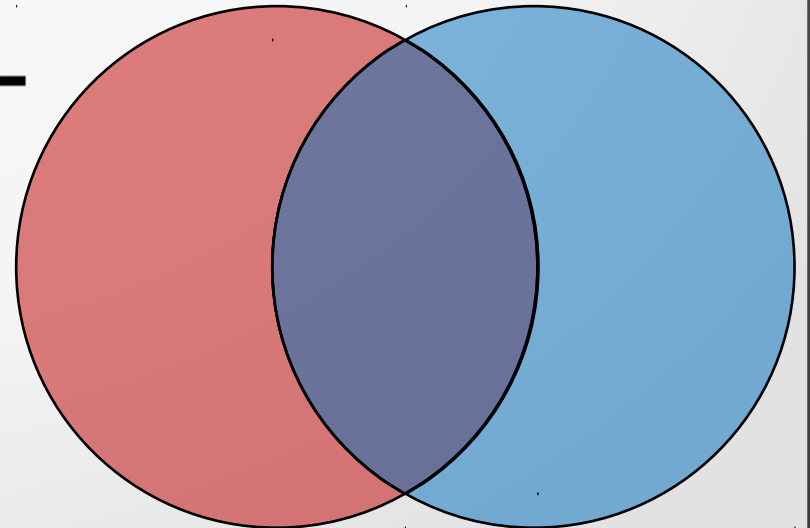
# Relações entre probabilidades

- Considerando as relações entre probabilidades

$$P(X, Y) = P(Y|X) * P(X) = P(X|Y) * P(Y)$$

- Se invertemos as probabilidades condicionais, temos que:

$$P(Y|X) = \frac{P(X|Y) * P(Y)}{P(X)}$$



# Regra de Bayes

- Para variáveis multivaloradas temos que a regra de Bayes pode ser escrita como:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

- A regra de Bayes é a base de todos os sistemas modernos de inferência probabilística.

# Regra de Bayes

- Ou em sua forma condicionada com uma evidência prática  $e$

$$P(Y|X, e) = \frac{P(X|Y, e)P(Y|e)}{P(X|e)}$$

# Aplicação

- Muito utilizado na prática quando se tem um efeito conhecido e é desejado descobrir a causa.

$$P(\text{causa} | \text{efeito}) = \frac{P(\text{efeito} | \text{causa}) P(\text{causa})}{P(\text{efeito})}$$

## Exemplo de uso

- Um agente de diagnóstico sabe que meningite causa rigidez no pescoço 50% das vezes.
- Ele também sabe que a probabilidade *a priori* de um paciente ter meningite é de  $1/50.000$  e de qualquer paciente ter rigidez no pescoço é  $1/20$ .
- Qual a probabilidade que um paciente tenha rigidez no pescoço causada por meningite?



## Exemplo de uso

- Sendo  $s$  a proposição que o paciente tem rigidez no pescoço e  $m$  a proposição de que o paciente possui meningite, temos:
  - $P(s|m) = 0,5$
  - $P(m) = 1/50.000$
  - $P(s) = 1/20$
  - $P(m|s) = P(s|m)P(m)/P(s) =$   
 $0,5 \times 1/50.000 \times 20 = 0,0002$

# Combinação de evidências

- Vimos que a regra de Bayes é útil para responder consultas condicionadas sobre uma única peça de evidência.
- Contudo, e se existirem mais de uma variável de evidência?
- É preciso conhecer a dependência entre cada dupla de variáveis!

# Combinação de evidências

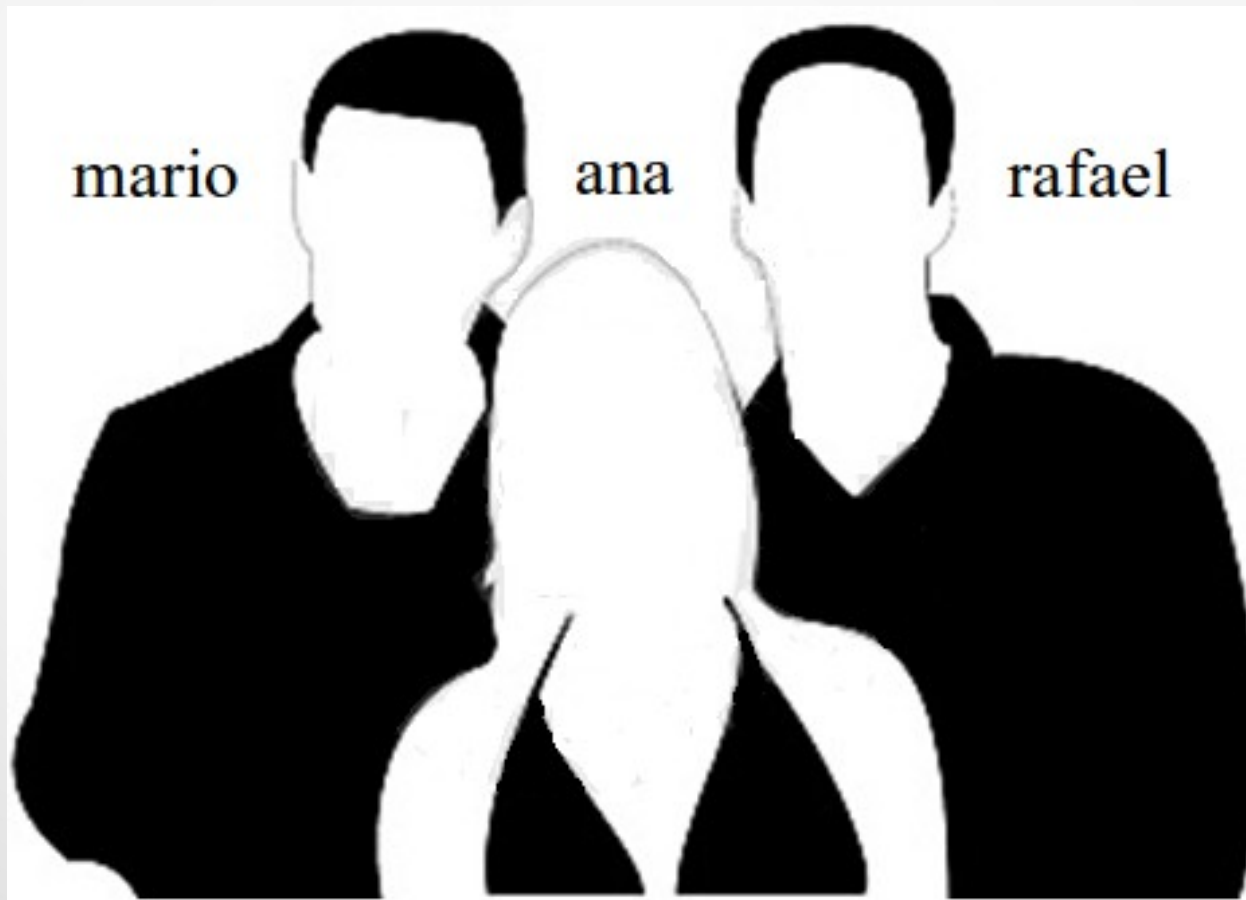
- Se houver  $n$  variáveis de evidência, o número de combinações possíveis de variáveis aleatória será  $2^n$ .
- Portanto, é preciso encontrar asserções sobre o domínio que permitam simplificar as expressões.
- Usar um método aproximado!

# Combinação de evidências

- Utilizar independência pode ser a solução.
  - Porém, pode ser difícil encontrar!
- Duas variáveis podem ser independentes quando condicionadas aos valores de uma terceira.
  - Esse é o conceito de independência condicional!
  - Menos raro que absoluta.

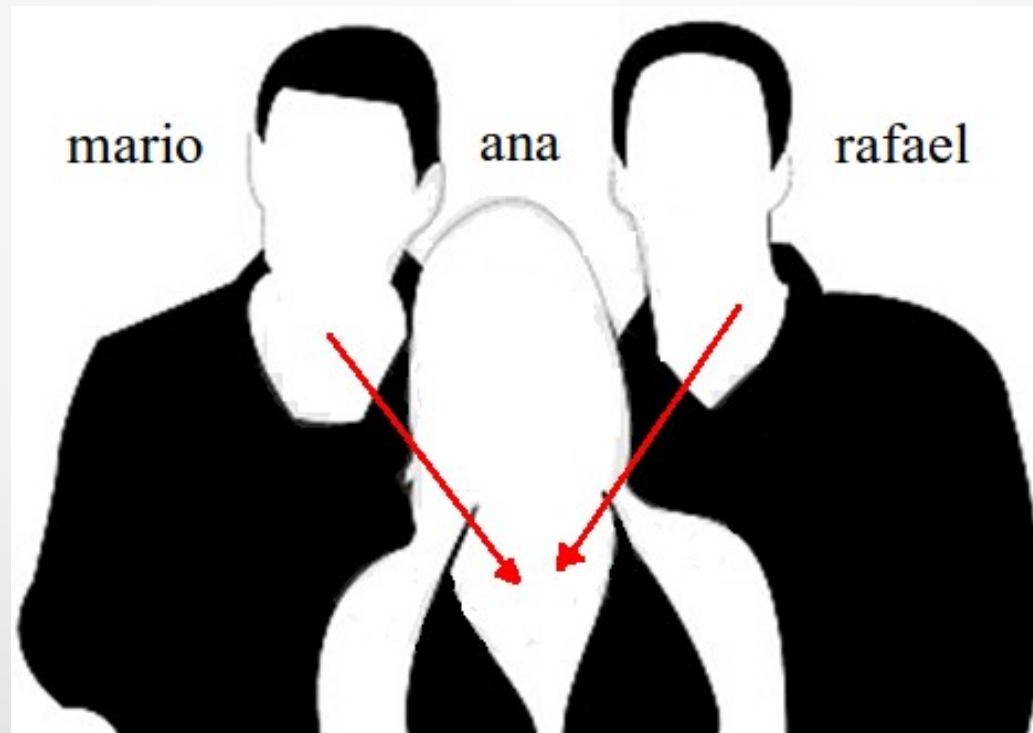
# Independência condicional

- Considere a probabilidade de três pessoas irem no TUSCA.



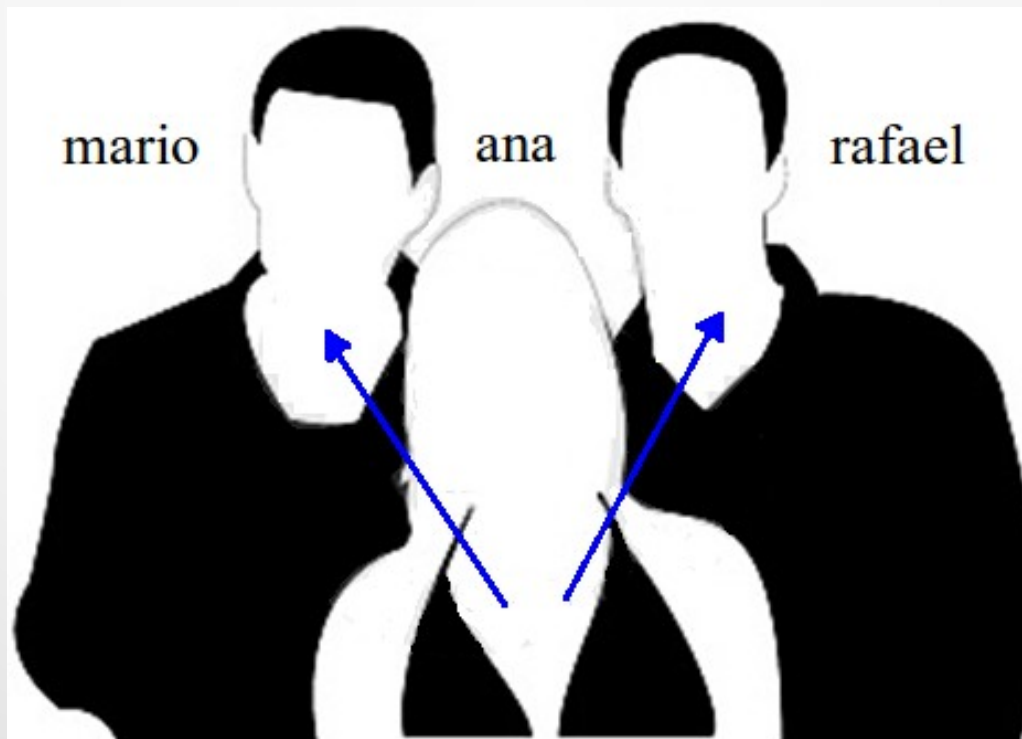
# Independência condicional

- Ana paquera Mario e Rafael. Portanto, a probabilidade de Mario ou Rafael ir no TUSCA (ou deixar de ir) influencia na probabilidade de Ana ir.



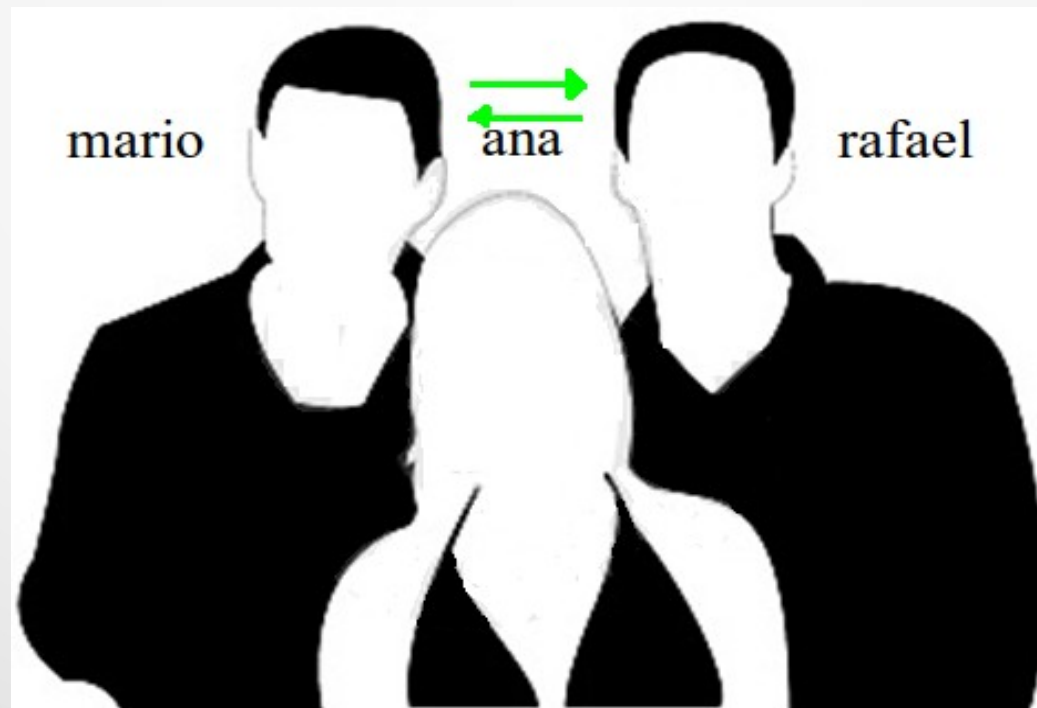
# Independência condicional

- Em contrapartida, Mario e Rafael possuem interesse por Ana e a probabilidade de Ana ir (ou deixar de ir) influencia na decisão dos dois.



# Independência condicional

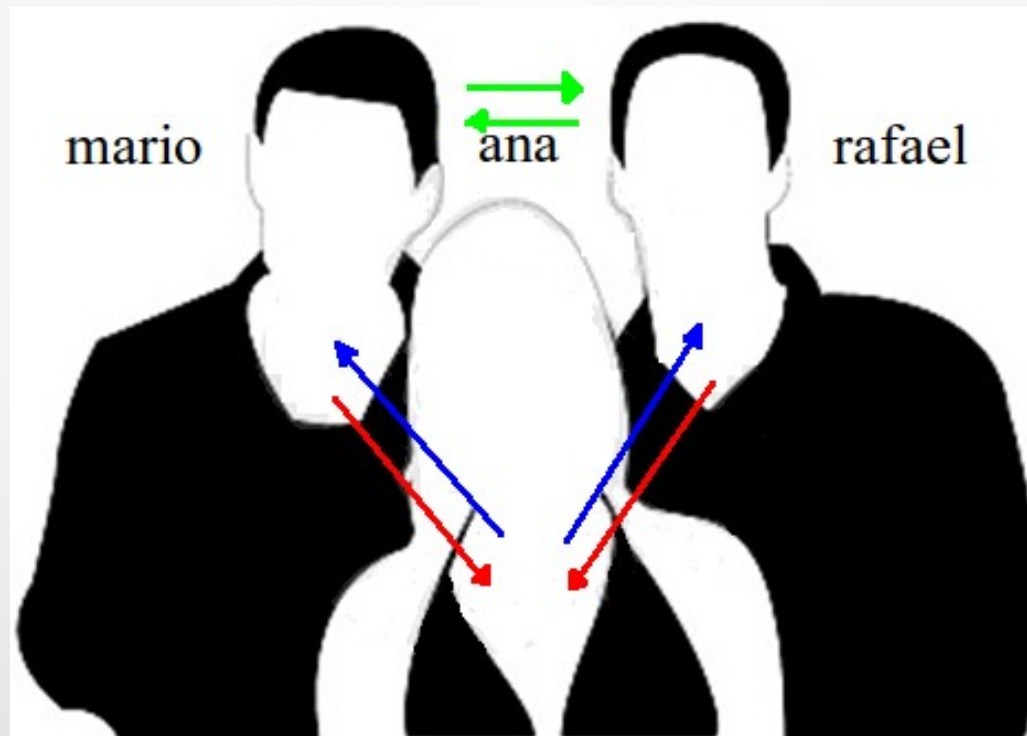
- Sem que saibam, a probabilidade de Mario ir influencia indiretamente na probabilidade de Rafael ir, e vice-versa, pois ambos influenciam na probabilidade de Ana ir.





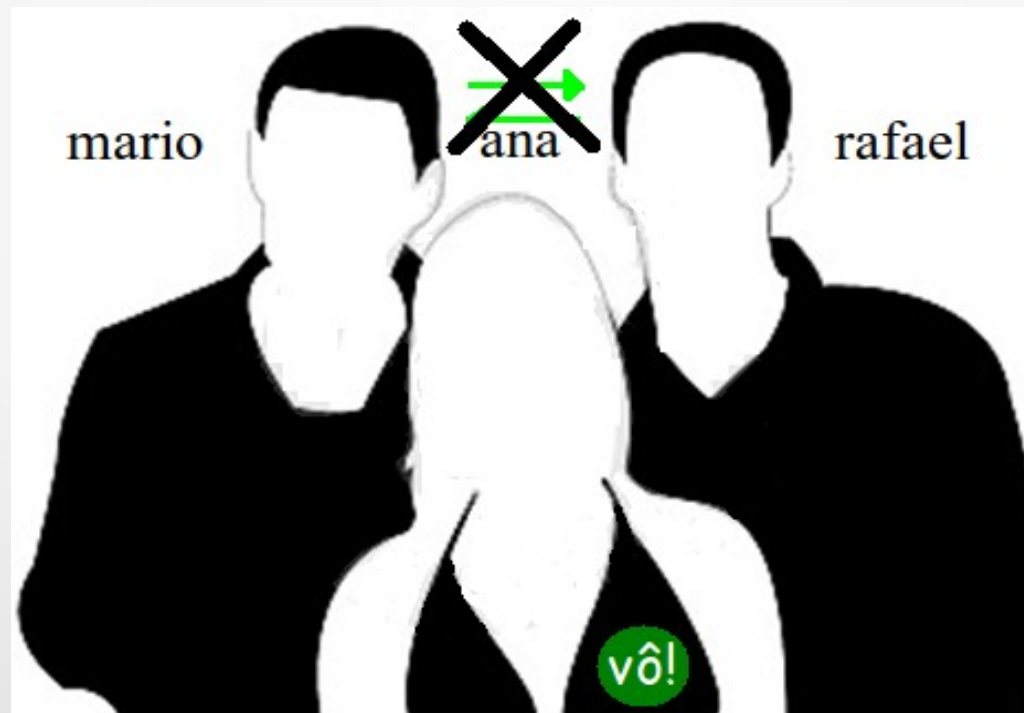
# Independência condicional

- Ou seja, existe dependência entre  $P(\text{Mario})$  e  $P(\text{Rafael})$  enquanto houver incerteza em relação a  $P(\text{Ana})$ .



# Independência condicional

- Entretanto, se Ana decide ir (ou não) no TUSCA, as probabilidades das variáveis *Mario* e *Rafael* deixam de se influenciar
- Isso é independência condicional!



# Independência condicional

- Duas variáveis são independentes condicionalmente se, dada uma terceira variável  $Z$ , temos:

$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

- Sendo assim, no exemplo dado temos:

$$P(Mario, Rafael|Ana) = P(Mario|Ana)P(Rafael|Ana)$$

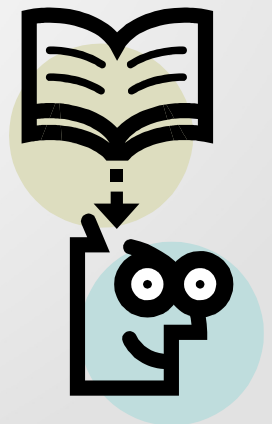
# Fatoração

- Asserções de independência condicional também permitem a decomposição da distribuição conjunta total em itens menores
- Exemplo:

$$\begin{aligned} P(\textit{Mario}, \textit{Rafael}, \textit{Ana}) &= \\ P(\textit{Mario}, \textit{Rafael} | \textit{Ana}) P(\textit{Ana}) &= \\ P(\textit{Mario} | \textit{Ana}) P(\textit{Rafael} | \textit{Ana}) P(\textit{Ana}) \end{aligned}$$

# Redução de complexidade

- A decomposição em problemas menores permite a redução significativa da complexidade de sistemas probabilísticos.
- Se existirem  $n$  variáveis independentes condicionalmente dado um evento, a complexidade do tamanho do sistema é reduzido de  $O(2^n)$  para  $O(n)$ !



# Causa X Múltiplos Efeitos

- Podemos modelar o problema em que uma variável aleatória influencia de maneira direta um conjunto de variáveis aleatórias condicionalmente independentes.
- Neste modelo, chamamos a variável que influencia de causa e as influenciadas de efeitos.

# Regra da cadeia ingênua

- Assumindo independência condicional entre os efeitos,  $P(\text{Efeito}_i | \text{Efeito}_{i+1}, \dots, \text{Efeito}_n, \text{Causa}) = P(\text{Efeito}_i | \text{Causa})$ , a distribuição conjunta pode ser escrita como:

$$P(Causa, \text{Efeito}_1, \dots, \text{Efeito}_n) = P(Causa) \prod_{i=1}^n P(\text{Efeito}_i | Causa)$$

## Regra da cadeia ingênua

Ingênuo!!!

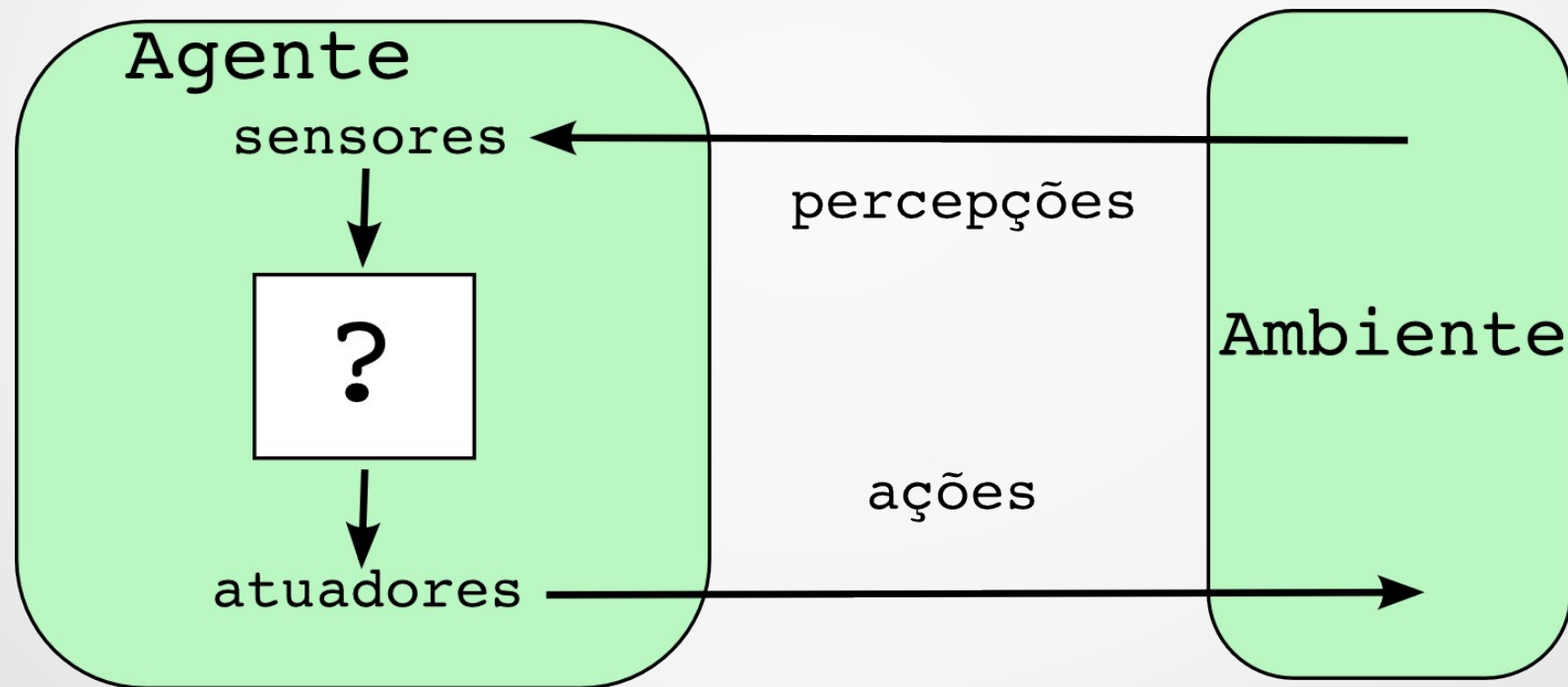
- Assumindo independência condicional entre os efeitos,  $P(\text{Efeito}_i | \text{Efeito}_{i+1}, \dots, \text{Efeito}_n, \text{Causa}) = P(\text{Efeito}_i | \text{Causa})$ , a distribuição conjunta pode ser escrita como:

$$P(\text{Causa}, \text{Efeito}_1, \dots, \text{Efeito}_n) = P(\text{Causa}) \prod_{i=1}^n P(\text{Efeito}_i | \text{Causa})$$



# Agentes Inteligentes Bayesianos

- O agente bayesiano utiliza como motor de inferência a lógica bayesiana para inferir sobre o ambiente e tomar a decisão da ação a ser aplicada



# Relação com classificação

- Existe uma relação entre a regra da cadeia para causa X efeito e classificação
  - No problema Causa X Efeito, o agente tem uma percepção do cenário (vetor  $x$  de entradas, que são os efeitos) e deve estimar qual é a causa para tomar decisão
  - No problema de classificação, o agente quer determinar a qual classe pertence o vetor  $x$ 
    - Ou seja, o funcionamento é igual!

# Classificação usando *Bayes*

- Para classificação usando Regra de Bayes, ou seja, determinar qual é o efeito, utilizamos um modelo probabilístico condicional.
  - Consideramos um vetor de percepções  $\mathbf{x}=(x_1,\dots,x_n)$  a ser classificado, composto por  $n$  características (*variáveis independentes*)
  - É possível calcular a probabilidade deste objeto pertencer a classe  $C_k$  ela como:

$$P(C_k|\mathbf{x}) = P(C_k|x_1,\dots,x_n)$$

## Regra de Bayes para Classificação

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



$$\textit{Posteriori} = \frac{\textit{Verossimilhança} \times \textit{Priori}}{\textit{Evidência}}$$

# Regra de Bayes para Classificação

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

Denominador é constante e não depende das classes  $C_k$ !

$$\text{Posteriori} = \frac{\text{Verossimilhança} \times \text{Priori}}{\text{Evidência}}$$

# Naïve Bayes para Classificação

Denotando proporcionalidade por  $\propto$

$$P(C_k | \mathbf{x}) \propto P(C_k) P(\mathbf{x} | C_k)$$



Numerador é a probabilidade conjunta!

$$P(C_k | x_1, \dots, x_n) \propto P(C_k, x_1, \dots, x_n)$$



Assumindo independência condicional entre os atributos do objeto, temos:

$$= P(C_k) \prod_{i=1}^n P(x_i | C_k)$$

# Naïve Bayes para Classificação

- Portanto, o modelo probabilístico para a classe  $C_k$  é dado por:

$$\begin{aligned} P(C_k | x_1, \dots, x_n) &= \\ &= \frac{1}{Z} P(C_k) \prod_{i=1}^n P(x_i | C_k) \end{aligned}$$

onde  $Z$  é  $P(\mathbf{x})$ , constante e dependente de  $\mathbf{x}$ .

# Construindo Classificador

- O classificador Bayesiano utiliza o modelo probabilístico combinado com alguma regra de decisão que defina qual classe é preferida dentre as hipóteses testadas
  - Assim, a classe de um objeto  $x$  é dada por:

$$\hat{y} = \operatorname{argmax}_{k \in \{1, \dots, K\}} P(C_k) \prod_{i=1}^n P(x_i | C_k)$$



## O truque do log

- Ao multiplicar muitos valores ( $n$ ) abaixo de 0 faz com que haja instabilidade computacional (*underflow*).
  - Para evitar esse problema, pode-se aplicar o truque do log

$$\log \left( P \left( C_k \right) \right) \prod_{i=1}^n \log \left( P \left( x_i | C_k \right) \right)$$

# Exemplo: Agente jogador tênis com NB

Dia	Clima	Temperatura	Umidade	Vento	Jogar Tênis
1	ensolarado	quente	alta	fraco	não
2	ensolarado	quente	alta	forte	não
3	nublado	quente	alta	fraco	sim
4	chuva	suave	alta	fraco	sim
5	chuva	fria	normal	fraca	sim
6	chuva	fria	normal	forte	não
7	nublado	fria	normal	forte	sim
8	ensolarado	suave	alta	fraco	não
9	ensolarado	fria	normal	fraco	sim
10	chuva	suave	normal	fraco	sim
11	ensolarado	suave	normal	forte	sim
12	nublado	suave	alta	forte	sim
13	nublado	quente	normal	fraco	sim
14	chuva	suave	alta	forte	não

# Treinando agente NB para tênis

- $P(\text{Jogar} = \text{sim}) = 9/14$     $P(\text{Jogar} = \text{não}) = 5/14$

<b>Clima</b>	<b>Jogar = sim</b>	<b>Jogar = não</b>
ensolarado	2/9	3/5
nublado	4/9	0/5
chuva	3/9	2/5

<b>Temperatura</b>	<b>Jogar = sim</b>	<b>Jogar = não</b>
quente	2/9	2/5
suave	4/9	2/5
fria	3/9	1/5

<b>Umidade</b>	<b>Jogar = sim</b>	<b>Jogar = não</b>
alta	3/9	4/5
normal	6/9	1/5

<b>Vento</b>	<b>Jogar = sim</b>	<b>Jogar = não</b>
forte	3/9	3/5
fraco	6/9	2/5

# Usando modelo treinado

- Considerando as probabilidades dos atributos temos

<b>Feature</b>	<b>Jogar=sim</b>	<b>Jogar = não</b>
Clima = ensolarado	2/9	3/5
Temperatura = fria	3/9	1/5
Umidade = alta	3/9	4/5
Vento = forte	3/9	3/5
<b>Soma</b>	<b>9/14</b>	<b>5/14</b>
<b>Probabilidade</b>	<b>0.0053</b>	<b>0.0206</b>

- $P(\text{Jogar=sim}|\text{ensolarado,fria,alta,forte}) = P(\text{ensolarado}|\text{sim}) * P(\text{fria}|\text{sim}) * P(\text{alta}|\text{sim}) * P(\text{forte}|\text{sim}) * P(\text{sim})$
- $P(\text{Jogar=não}|\text{ensolarado,fria,alta,forte}) = P(\text{ensolarado}|\text{não}) * P(\text{fria}|\text{não}) * P(\text{alta}|\text{não}) * P(\text{forte}|\text{não}) * P(\text{não})$

# Problema

- Categorias que não possuem entradas para todos os atributos resultam em 0 para probabilidades condicionais
  - $P(C|\mathbf{x}) = P(x_1|C) * P(x_2|C) * P(C)$

# Problema

- Categorias que não possuem entradas para todos os atributos resultam em 0 para probabilidades condicionais
  - $P(C|\mathbf{x}) = P(x_1|C) * P(x_2|C) * P(C) = 0$

# Suavização de Laplace

- Categorias que não possuem entradas para todos os atributos resultam em 0 para probabilidades condicionais
  - $P(C|\mathbf{x}) = P(x_1|C) * P(x_2|C) * P(C) = 0$
- Solução
  - Somar 1 no numerador e denominador das categorias vazias, onde  $n$  é o número de atributos no conjunto de treino

$$P(C|x) = \frac{\text{contagem}(x, C) + 1}{\text{contagem}(C) + n + 1}$$

# Tipos diferentes de Naïve Bayes

- Diferentes tipos de dados requerem diferentes tipos de distribuição de probabilidade para gerar o modelo

Modelo de NB	Tipo de dado
Bernoulli	Binário (V/F)
Multinomial	Discreto (e.g. contagens)
Gaussiano	Contínuo



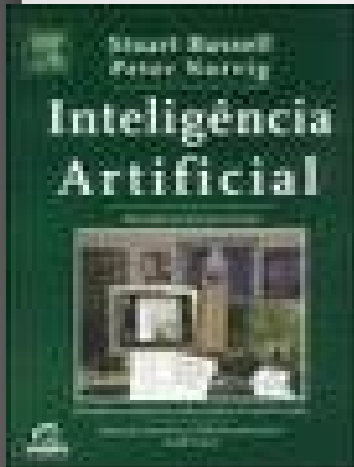
# Combinando tipos diferentes de atributos

- Problema
  - Os dados contêm diferente tipos de atributos
    - Ou seja, contínuos e categóricos/discretos

# Combinando tipos diferentes de atributos

- Problema
  - Os dados contêm diferentes tipos de atributos
    - Ou seja, contínuos e categóricos/discretos
- Possíveis soluções
  - Opção 1: dividir o atributo contínuo em intervalos e ajustar a um modelo multinomial
  - Opção 2: ajustar um modelo gaussiano aos atributos contínuos e categóricos e combinar

# Bibliografia



RUSSEL, S. and NORVIG, P. Inteligência Artificial: uma abordagem moderna. Editora Campus, 2003