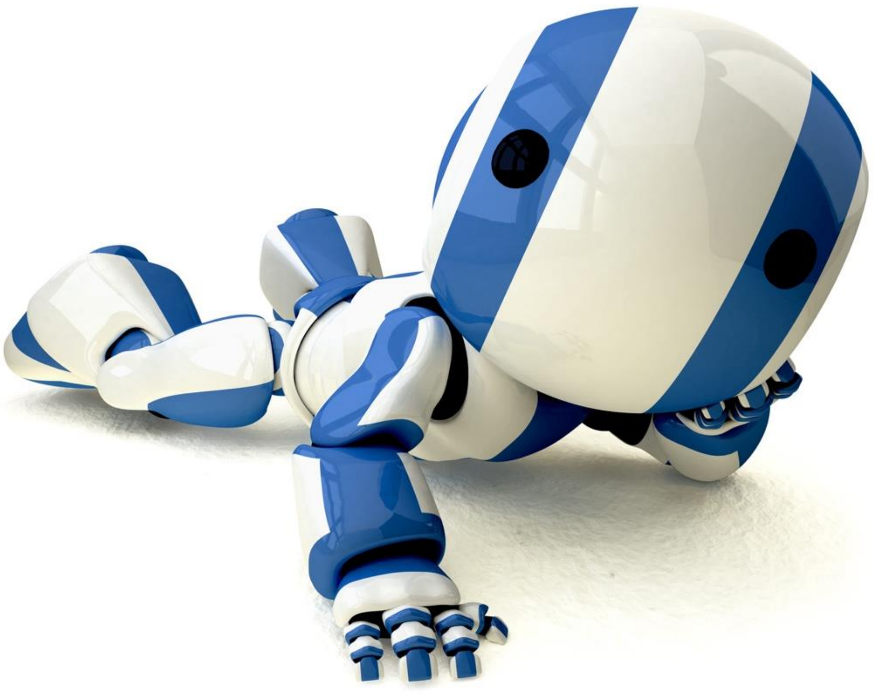
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
ESCOLA POLITÉCNICA

Aprendizado de Máquina

Aprendizado Supervisionado II
Aprendizado Probabilístico

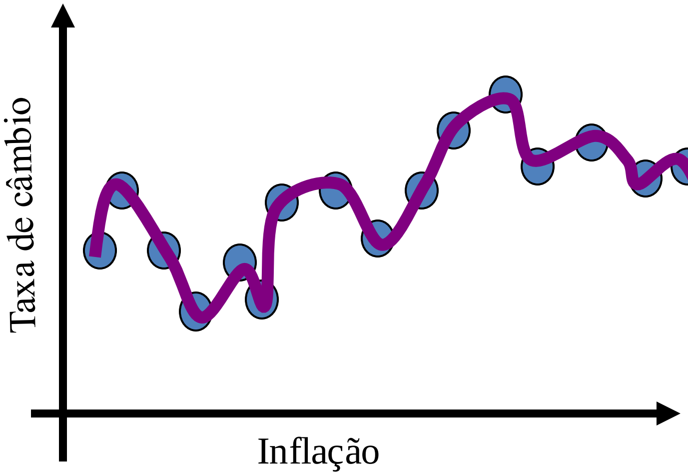
Prof. Me. Otávio Parraga

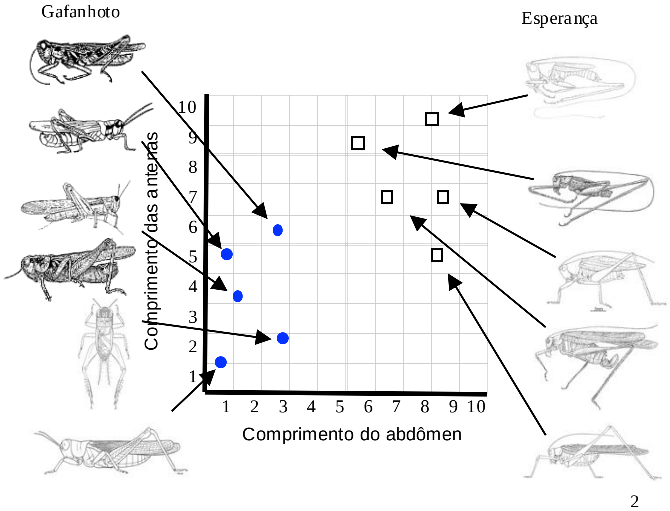


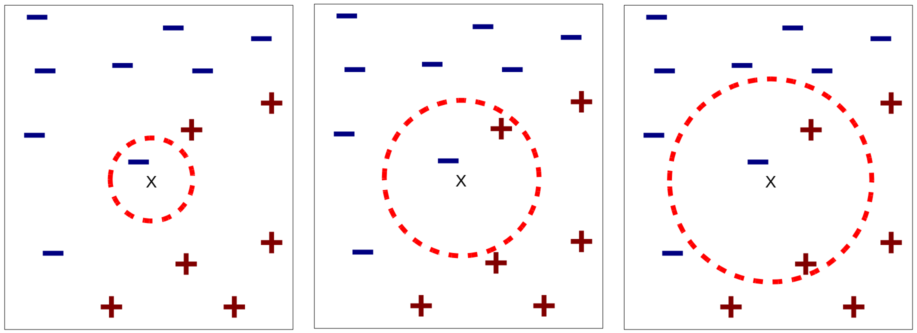
MALTA

Machine Learning Theory
and Applications Lab

Aula Passada







(a) 1-nearest neighbor (b) 2-nearest neighbor (c) 3-nearest neighbor

Aula de Hoje

- Revisão de Teoria das Probabilidades
- Introdução ao Aprendizado Bayesiano
- Classificador Naïve Bayes

Conceito de Probabilidade

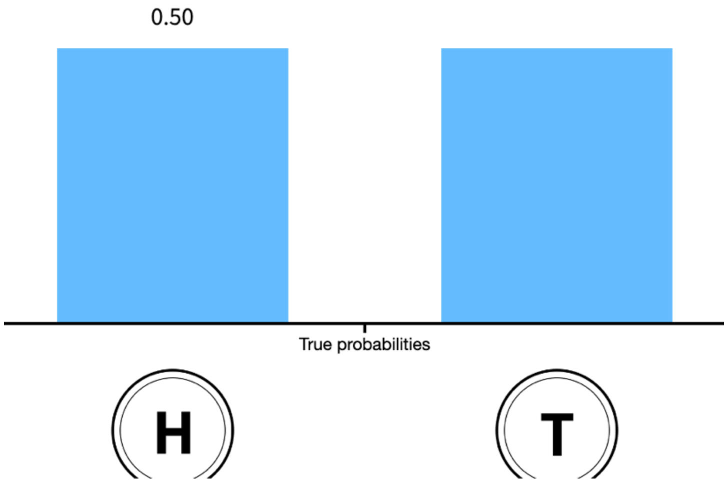
- Intuição:
  - Em algum fenômeno, vários resultados são possíveis. Quando o fenômeno se repete por um número grande de vezes, cada resultado ocorre com uma determinada frequência relativa, ou probabilidade. Se um resultado ocorre mais frequentemente que outro, dizemos que ele é mais provável

Conceito de Probabilidade

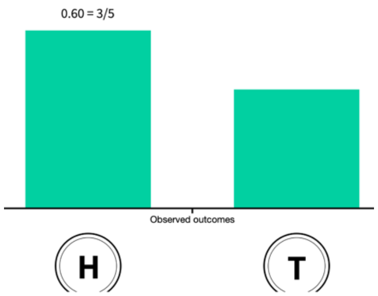
- Surge em dois contextos:
  - Ao se realmente repetir experimentos
    - Ex: anotar a cor de 1000 carros que passam na rua. Destes, 57 são verdes. Pode-se, então, estimar a probabilidade de um carro ser verde como $57/1000 = 0.057$
  - Em “conceitos idealizados” de um processo repetido
    - Ex: o comportamento de um dado não viciado de 6 faces. A probabilidade esperada de tirar um 5 é de $1/6 = 0.1667$
    - Ex2: escolher uma distribuição normal como modelo para representar as probabilidades esperadas da altura de um adulto do sexo masculino

Conceito de Probabilidade

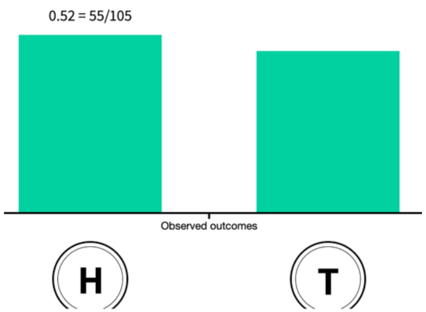
- Conceito idealizado:
  moeda justa (50% vs 50%)



- Resultados Experimentais:



5 lançamentos



105 lançamentos

Espaço de Probabilidades

- Um espaço de probabilidades é um processo (experimento, fenômeno) aleatório com 3 componentes:
  - Ω (espaço amostral): conjunto de todos possíveis resultados do fenômeno
    - Número de possíveis resultados $= |\Omega|$
  - F: conjunto de todos possíveis eventos
    - Um evento é um sub-conjunto de resultados (elementos do espaço amostral)
    - Denotado por letra maiúscula (ex: evento E)
    - Um evento possui de 0 a $|\Omega|$ resultados
  - P: função de distribuição de probabilidades
    - Função que mapeia cada resultado e evento a um número real entre 0 e 1 (probabilidade do resultado ou evento)
    - Probabilidade de um evento é a soma das probabilidades dos possíveis resultados deste evento

Espaço de Probabilidades

- Exemplificando: Jogar um dado de 6 lados
  - Ω (espaço amostral)= {1,2,3,4,5,6}
  - $F: \{\{\theta\}, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{1,6\}, \{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \dots \{1,2,3,4,5,6\}\}$
  - Evento: $E = \{2,4,6\}$
    - Obter um número par
  - $P: P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$

Teoria das Probabilidades

- $P(X)$
  - Probabilidade do evento X ocorrer
  - Ex: X = "Vai chover amanhã"
    - X = variável aleatória binária
    - $P(X) = 1$: com certeza vai chover amanhã
    - $P(X) = 0$: com certeza não vai chover amanhã
    - $0 \leq P(X) \leq 1$: níveis de incerteza!

Axiomas das Probabilidades

- Não-negatividade:
  - Para qualquer evento $E \in F, P(E) \geq 0$

- Todos possíveis resultados:
  - $P(\Omega) = \sum_{\{x \in \Omega\}} P(X = x) = 1$

- Aditividade de eventos disjuntos:
  - Para quaisquer dois eventos $E, E' \in F$, se $E \cap E' = \emptyset$, então $P(E \cup E') = P(E) + P(E')$

Tipos de Espaços Amostrais

- Espaço amostral discreto
  - $|\Omega|$ é finito
  - Análise envolve somatórios $\sum$

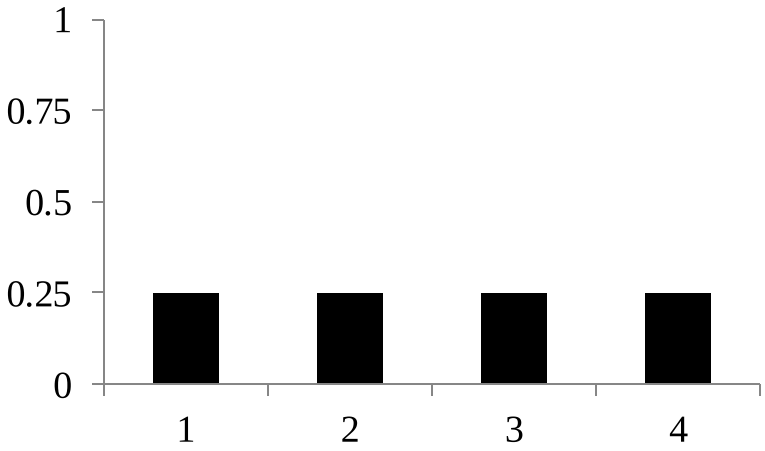
- Espaço amostral contínuo
  - $|\Omega|$ é infinito
  - Análise envolve integrais $\int$

Exemplo de um Espaço Amostral Discreto

- Jogar um dado (comum, 6 faces)
  - 6 possíveis resultados: $\Omega = \{1,2,3,4,5,6\}$
  - $2^6 = 64$ possíveis eventos
    - Ex: $E = \{1,3,5\}$, ou seja, resultado é ímpar
  - Se o dado não é viciado, resultados são equiprováveis:
    - $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$
  - Pelo axioma da aditividade, $P(E)$ (probabilidade do resultado ser ímpar) é dada por:
    - $P(1) + P(3) + P(5) = 3/6 = 0.5$
- Chamamos *P()* de *probability mass function* (PMF)

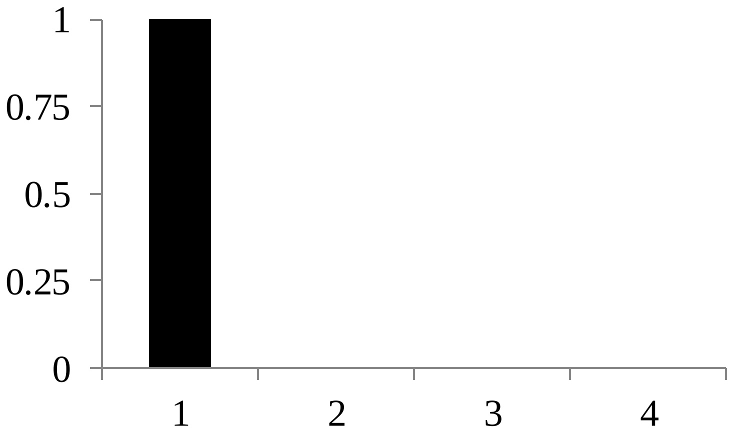
Teoria das Probabilidades

- Ex: $\Omega = \{1,2,3,4\}, x \in \Omega$



Distribuição Uniforme

$$P(X = x) = \frac{1}{|\Omega|}$$



Distribuição Degenerada

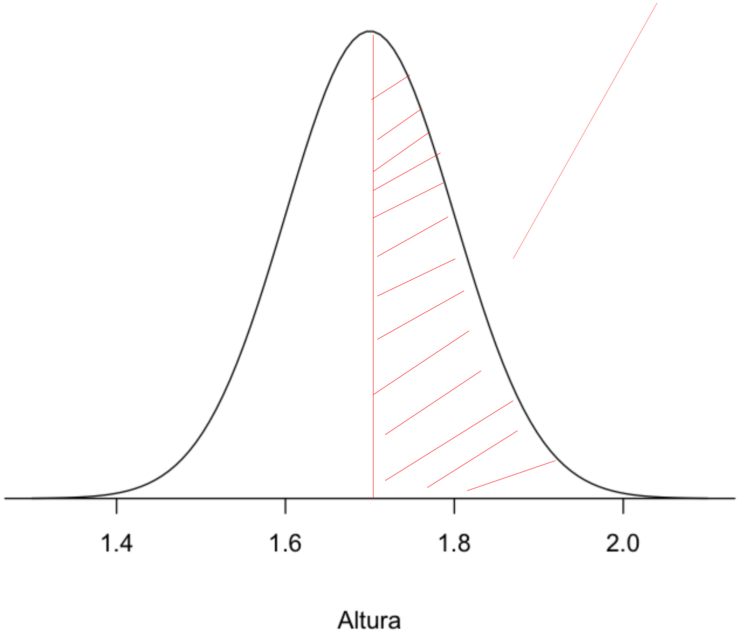
$$P(X = x) = \begin{cases} 1 \text{ se } x = 1 \\ 0 \text{ se } x \in \{2,3,4\} \end{cases}$$

Exemplo de um Espaço Amostral Contínuo

- Altura de um adulto do sexo masculino
  - Número infinito de resultados dentro de uma determinada faixa de valores (ex: 0.6m até 2.5m)
  - Número infinito de possíveis eventos
    - Ex: $\mathrm{E} = (R|R < 1.85m)$ = indivíduos com menos de 1.85m
  - Resultados não são equiprováveis, e são descritos por uma função contínua (*probability density function*)

Exemplo de um Espaço Amostral Contínuo

- Altura de um adulto do sexo masculino
  - $P(R)$ para um resultado em particular é **zero**
  - $\int P(R)$ com $R = [-\infty, +\infty]$ (área sob a curva) é 1
  - Ex: $\mathrm{P}(\mathrm{R} > 1.7\mathrm{m}) = \int \mathrm{P}(\mathrm{R})$ com $\mathrm{R} = [1.7, +\infty] \approx 0.5$



Prior Probability

(Probabilidade *a priori* ou incondicional)

- Probabilidade (ou grau de incerteza, ou grau de crença) de algum evento na ausência de qualquer outra informação

- Exemplo: $P(Moeda = coroa) = 0.5$

Probabilidade da União de Eventos

- Se eventos A e B forem mutuamente exclusivos (disjuntos):

$$P(A \cup B) = P(A) + P(B)$$

- Caso contrário, é necessário computar a probabilidade conjunta $P(A \cap B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

$P(Dado = 1)$?

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

$P(Dado = 1) = 1/6 \ (0.167)$

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

$P(Dado = 1) = 1/6\ (0.167)$
$P(Dado = 2)\ ?$

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

$$P(Dado = 1) = 1/6\ (0.167)$$
$$P(Dado = 2) = 1/6\ (0.167)$$
$$P(Dado = 3) = 1/6\ (0.167)$$
$$P(Dado = 4) = 1/6\ (0.167)$$
$$P(Dado = 5) = 1/6\ (0.167)$$
$$P(Dado = 6) = 1/6\ (0.167)$$

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

$$P(DadoA = 1 \cap DadoB = 1) \ = \ ?$$

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

$$P(DadoA = 1 \cap DadoB = 1) = 1/6 \times 1/6 = 1/36$$

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

$P(DadoA = 1 \cap DadoB = 1) = 1/6 \times 1/6 = 1/36$
$P(DadoA = 1 \cap DadoB = 2) = ?$

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

$P(DadoA = 1 \cap DadoB = 1) = 1/6 \times 1/6 = 1/36$
$P(DadoA = 1 \cap DadoB = 2) = 1/6 \times 1/6 = 1/36$
$P(DadoA = 1 \cap DadoB = 3) = 1/6 \times 1/6 = 1/36$

..

..

$P(DadoA = 6 \cap DadoB = 6) = 1/6 \times 1/6 = 1/36$

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

O que podemos concluir da probabilidade conjunta
$P(DadoA = x \cap DadoB = y)$?

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

O que podemos concluir da probabilidade conjunta
$P(DadoA = x \cap DadoB = y)$?

$$= P(DadoA = x) \times P(DadoB = y)$$

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

$P(A \cap B)$ é sempre $= P(A) \times P(B)$???

Probabilidade Conjunta

- É a probabilidade de dois eventos A e B ocorrerem
  - $P(A)$: probabilidade do evento A ocorrer
  - $P(B)$: probabilidade do evento B ocorrer
  - $P(A \cap B)$: probabilidade de A e B ocorrerem

$P(A \cap B)$ é sempre $= P(A) \times P(B)$???

NÃO!!!!!!

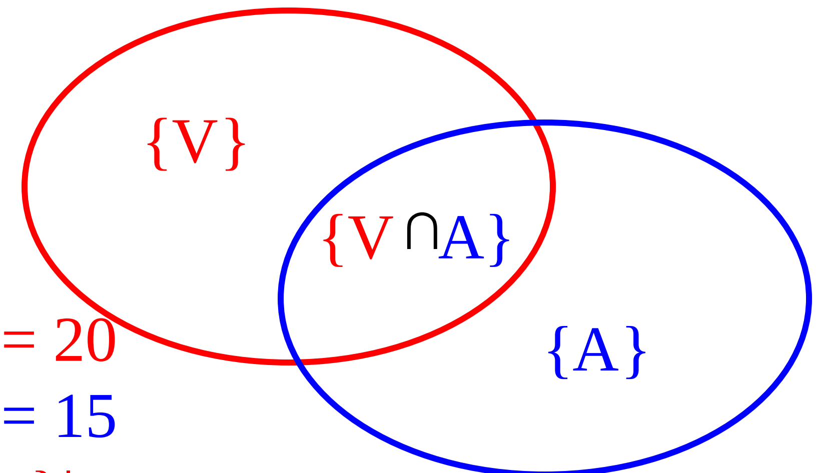
Apenas quando A e B forem eventos independentes

Probabilidade Condicional

- Se A e B **não forem eventos independentes**, tem-se:
  - $P(A \cap B) = P(B|A) \times P(A)$ Regra do Produto
    - Onde $P(B|A)$ é a probabilidade que B ocorra dado que A ocorreu (probabilidade condicional de B dado A)

Probabilidade Condicional

- Se A e B **<u>não forem eventos independentes</u>**, tem-se:
  - $P(A \cap B) = P(B|A) \times P(A)$ Regra do Produto
    - Onde $P(B|A)$ é a probabilidade que B ocorra dado que A ocorreu (probabilidade condicional de B dado A)



$|\{V\}| = 20$

$|\{A\}| = 15$

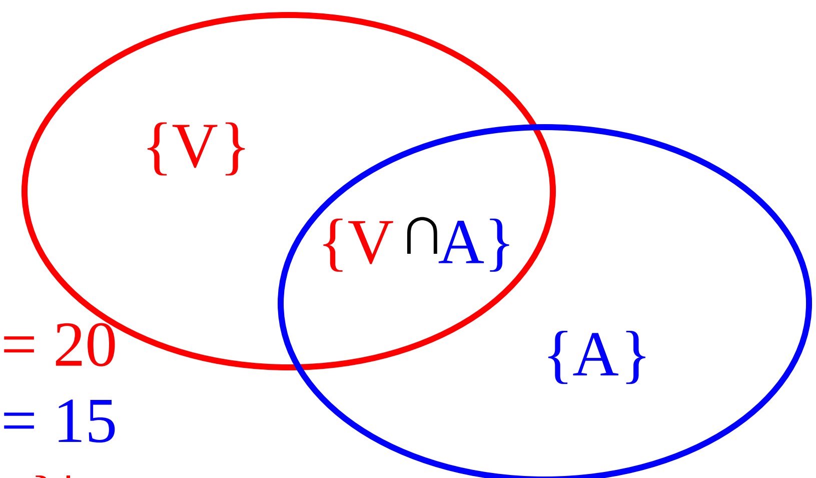
$|\{A \cap V\}| = 5$

$|X| = 30$

Probabilidade Condicional

- Se A e B **<u>não forem eventos independentes</u>**, tem-se:
  - $P(A \cap B) = P(B|A) \times P(A)$ Regra do Produto
    - Onde $P(B|A)$ é a probabilidade que B ocorra dado que A ocorreu (probabilidade condicional de B dado A)

$$P(A) = \frac{15}{30} = \frac{1}{2}$$



$|\{V\}| = 20$

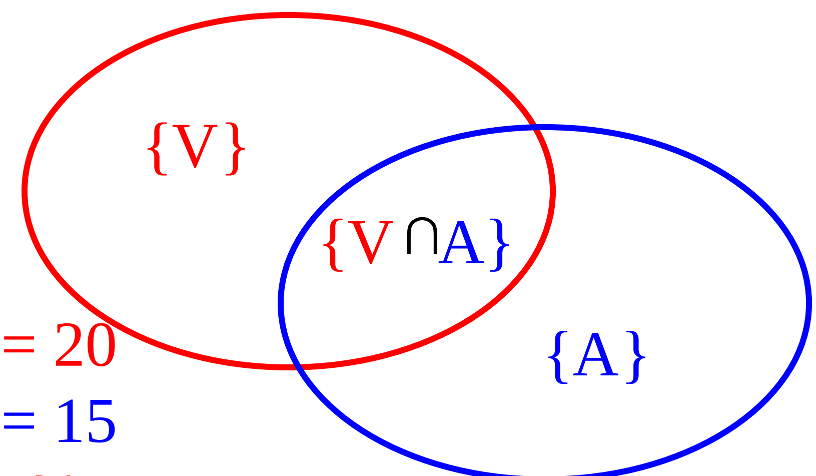
$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

Probabilidade Condicional

- Se A e B **<u>não forem eventos independentes</u>**, tem-se:
  - $P(A \cap B) = P(B|A) \times P(A)$ Regra do Produto
    - Onde $P(B|A)$ é a probabilidade que B ocorra dado que A ocorreu (probabilidade condicional de B dado A)



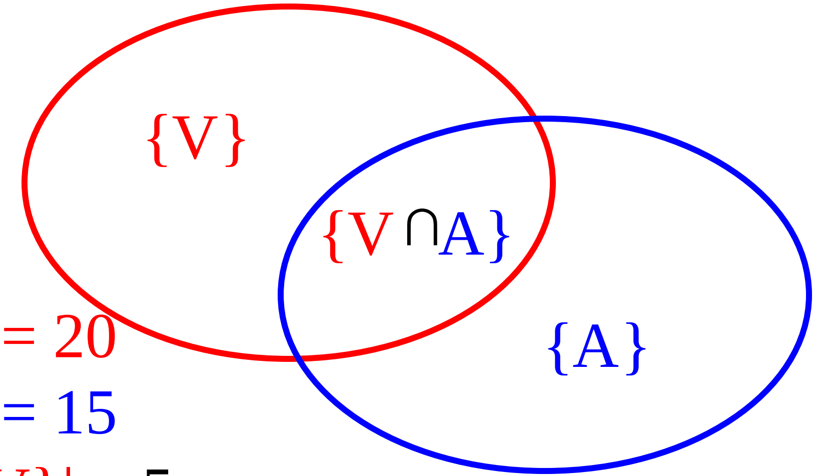
$|\{V\}| = 20$
$|\{A\}| = 15$
$|\{A \cap V\}| = 5$
$|X| = 30$

$$P(A) = \frac{15}{30} = \frac{1}{2}$$

$$P(V) = \frac{20}{30} = \frac{2}{3}$$

Probabilidade Condicional

- Se A e B **<u>não forem eventos independentes</u>**, tem-se:
  - $P(A \cap B) = P(B|A) \times P(A)$ Regra do Produto
    - Onde $P(B|A)$ é a probabilidade que B ocorra dado que A ocorreu (probabilidade condicional de B dado A)



$$P(A) = \frac{15}{30} = \frac{1}{2}$$

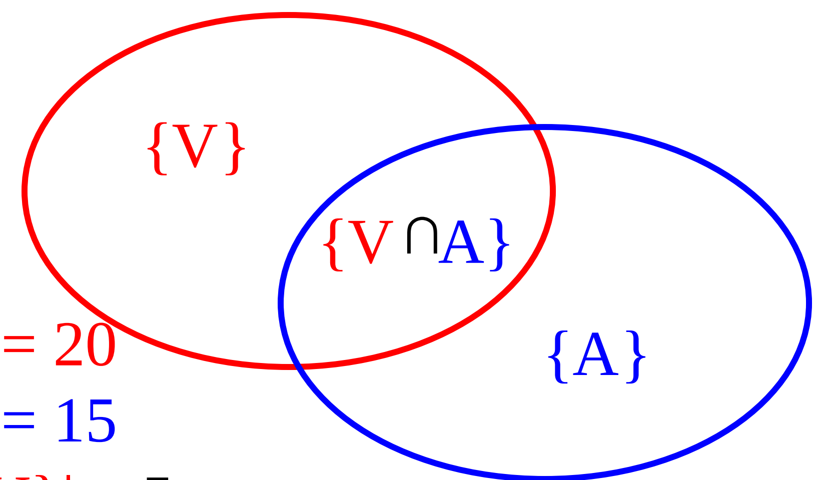
$$P(V) = \frac{20}{30} = \frac{2}{3}$$

$$P(A \cap V) = ?$$

$|\{V\}| = 20$
$|\{A\}| = 15$
$|\{A \cap V\}| = 5$
$|X| = 30$

Probabilidade Condicional

- Se A e B **<u>não forem eventos independentes</u>**, tem-se:
  - $P(A \cap B) = P(B|A) \times P(A)$ Regra do Produto
    - Onde $P(B|A)$ é a probabilidade que B ocorra dado que A ocorreu (probabilidade condicional de B dado A)



|{V}| = 20

|{A}| = 15

|{A ∩V}| = 5

|X| = 30

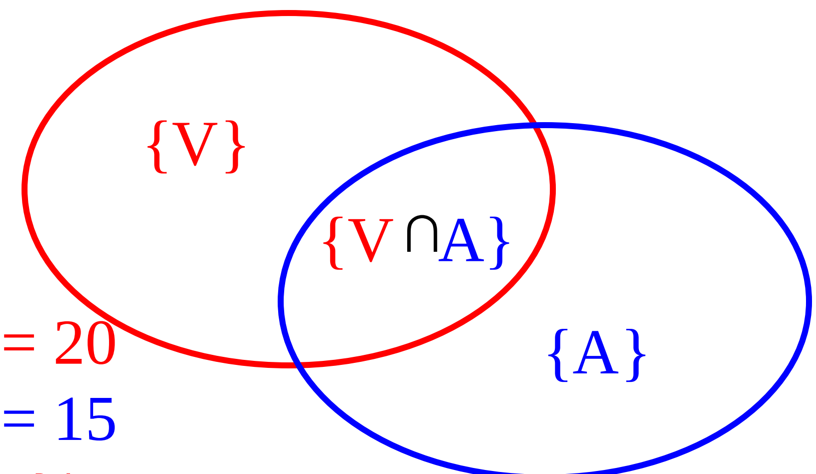
$$P(A) = \frac{15}{30} = \frac{1}{2}$$

$$P(V) = \frac{20}{30} = \frac{2}{3}$$

$$P(A \cap V) = P(V|A) \times P(A)$$

Probabilidade Condicional

- Se A e B **não forem eventos independentes**, tem-se:
  - $P(A \cap B) = P(B|A) \times P(A)$ Regra do Produto
    - Onde $P(B|A)$ é a probabilidade que B ocorra dado que A ocorreu (probabilidade condicional de B dado A)



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

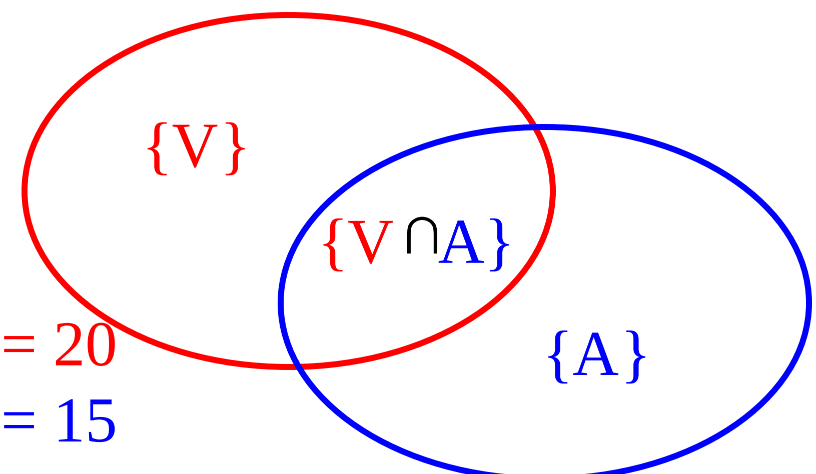
$$P(A) = \frac{15}{30} = \frac{1}{2}$$

$$P(V) = \frac{20}{30} = \frac{2}{3}$$

$$P(A \cap V) = P(V|A) \times \frac{1}{2}$$

Probabilidade Condicional

- Se A e B **não forem eventos independentes**, tem-se:
  - $P(A \cap B) = P(B|A) \times P(A)$ Regra do Produto
    - Onde $P(B|A)$ é a probabilidade que B ocorra dado que A ocorreu (probabilidade condicional de B dado A)



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

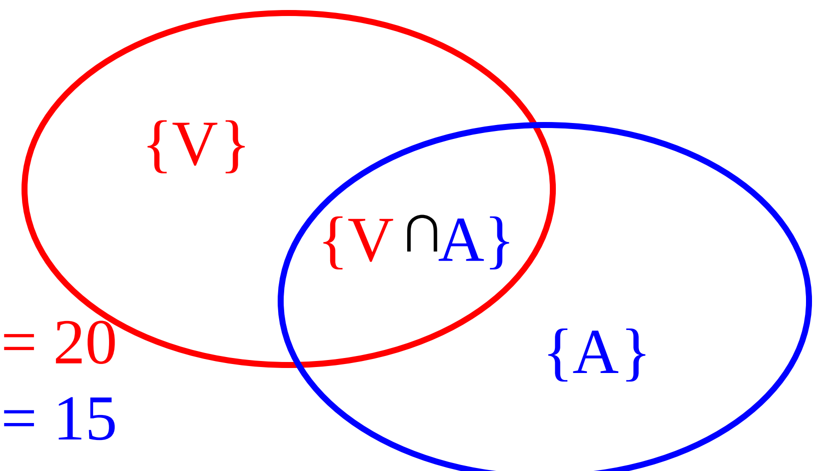
$$P(A) = \frac{15}{30} = \frac{1}{2}$$

$$P(V) = \frac{20}{30} = \frac{2}{3}$$

$$P(A \cap V) = \frac{5}{15} \times \frac{1}{2} = \frac{1}{6}$$

Probabilidade Condicional

- **Atenção!** $P(A \cap V) = P(V \cap A)$



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

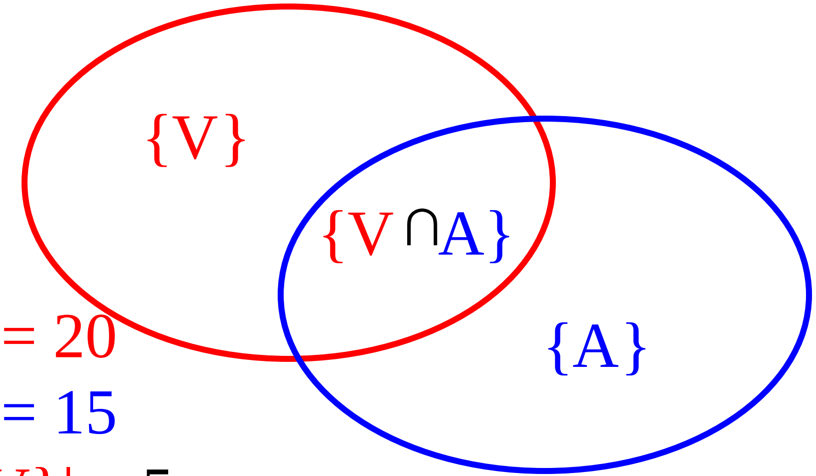
Probabilidade Condicional

- **Atenção!** $P(A \cap V) = P(V \cap A)$

$$P(A) = \frac{15}{30} = \frac{1}{2}$$

$$P(V) = \frac{20}{30} = \frac{2}{3}$$

$$P(A \cap V) = \frac{5}{15} \times \frac{1}{2} = \frac{1}{6}$$



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

Probabilidade Condicional

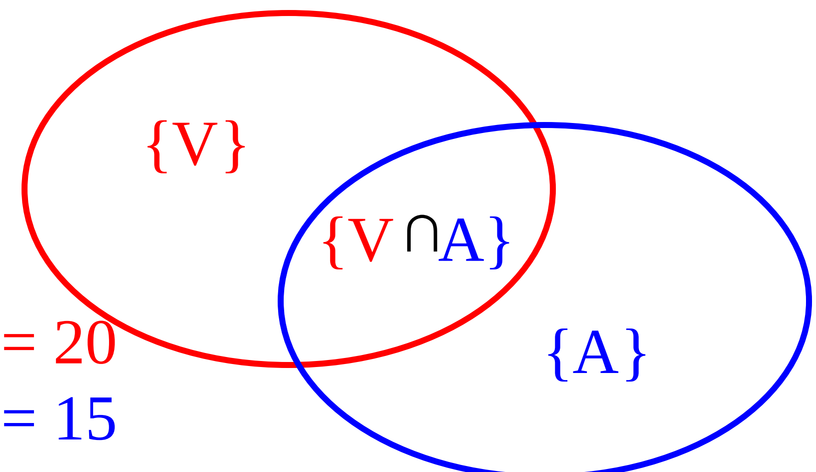
- **Atenção!** $P(A \cap V) = P(V \cap A)$

$$P(A) = \frac{15}{30} = \frac{1}{2}$$

$$P(V) = \frac{20}{30} = \frac{2}{3}$$

$$P(A \cap V) = \frac{5}{15} \times \frac{1}{2} = \frac{1}{6}$$

$$P(A \cap V) = P(A|V) \times P(V)$$



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

Probabilidade Condicional

- **Atenção!** $P(A \cap V) = P(V \cap A)$

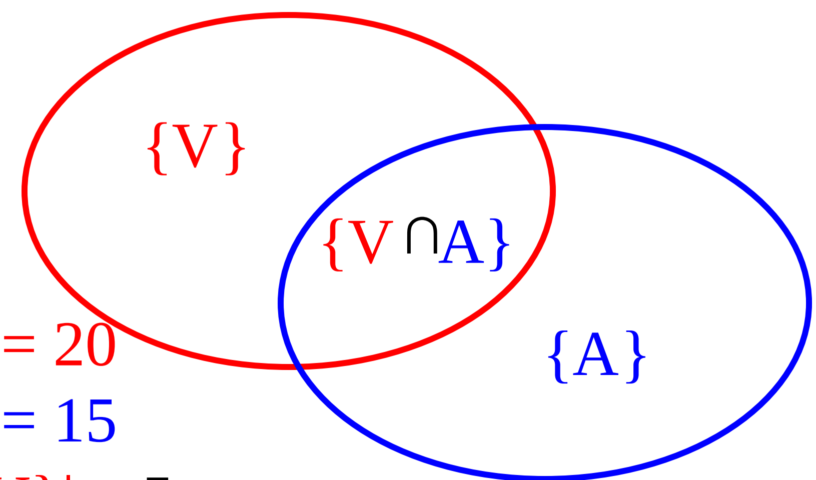
$$P(A) = \frac{15}{30} = \frac{1}{2}$$

$$P(V) = \frac{20}{30} = \frac{2}{3}$$

$$P(A \cap V) = \frac{5}{15} \times \frac{1}{2} = \frac{1}{6}$$

$$P(A \cap V) = P(A|V) \times P(V)$$

$$P(A \cap V) = P(A|V) \times \frac{2}{3}$$



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

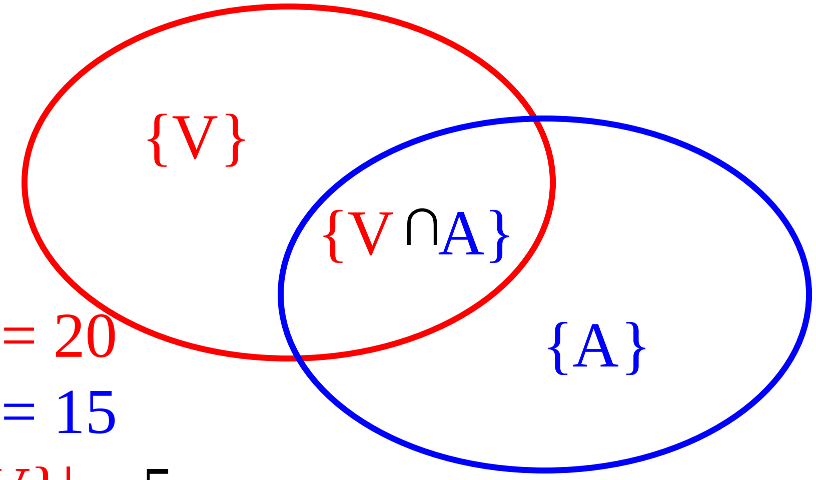
Probabilidade Condicional

- **Atenção!** $P(A \cap V) = P(V \cap A)$

$$P(A) = \frac{15}{30} = \frac{1}{2}$$

$$P(V) = \frac{20}{30} = \frac{2}{3}$$

$$P(A \cap V) = \frac{5}{15} \times \frac{1}{2} = \frac{1}{6}$$



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

$$P(A \cap V) = P(A|V) \times P(V)$$

$$P(A \cap V) = P(A|V) \times \frac{2}{3}$$

$$P(A \cap V) = \frac{5}{20} \times \frac{2}{3} = \frac{1}{6}$$

Probabilidade Condicional

$$P(A \cap V) = P(V|A) \times P(A)$$

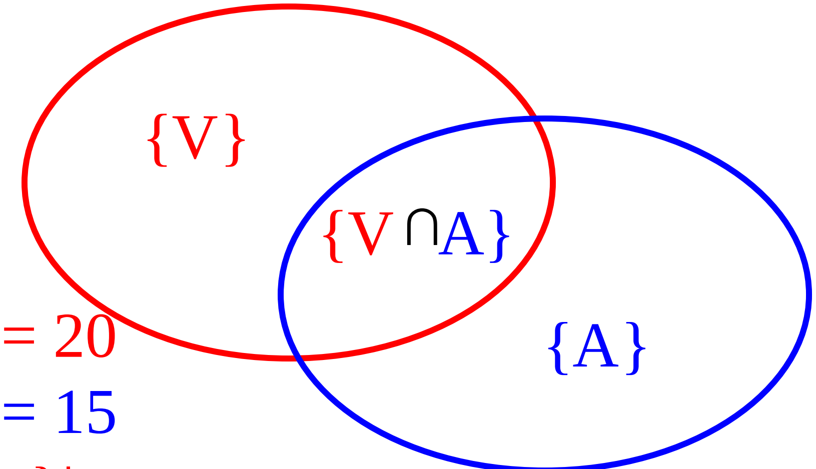
$$P(A \cap V) = P(V|A) \times \frac{1}{2}$$

$$P(A \cap V) = \frac{5}{15} \times \frac{1}{2} = \frac{1}{6}$$

$$P(A \cap V) = P(A|V) \times P(V)$$

$$P(A \cap V) = P(A|V) \times \frac{2}{3}$$

$$P(A \cap V) = \frac{5}{20} \times \frac{2}{3} = \frac{1}{6}$$



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

Probabilidade Condicional

$$P(A \cap V) = P(V|A) \times P(A)$$

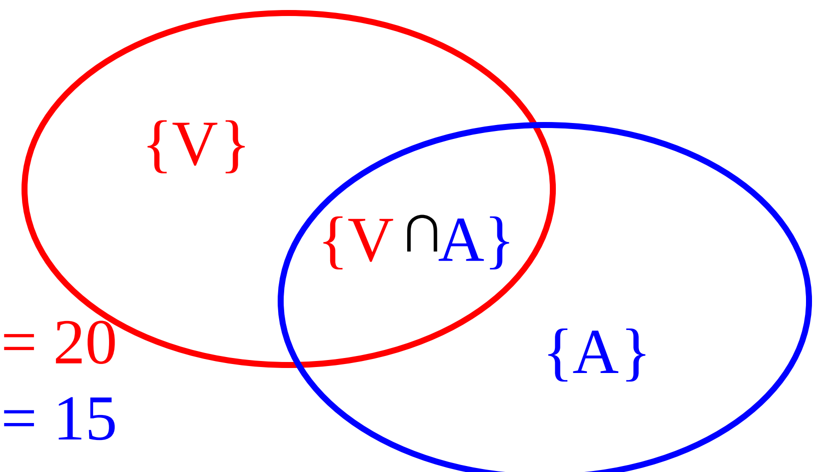
$$P(A \cap V) = P(A|V) \times P(V)$$

$$P(A \cap V) = P(V|A) \times \frac{1}{2}$$

$$P(A \cap V) = P(A|V) \times \frac{2}{3}$$

$$P(A \cap V) = \frac{5}{15} \times \frac{1}{2} = \frac{1}{6}$$

$$P(A \cap V) = \frac{5}{20} \times \frac{2}{3} = \frac{1}{6}$$



$$P(V|A) \times P(A)$$
$$= P(A|V) \times P(V)$$

$|\{V\}| = 20$

$|\{A\}| = 15$

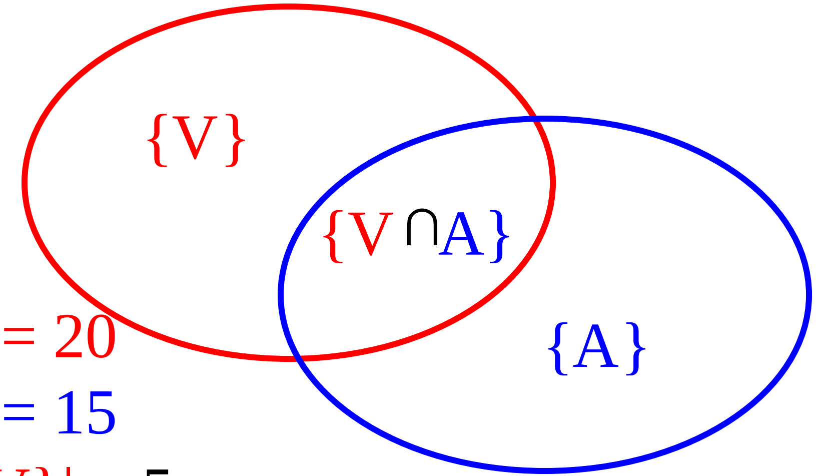
$|\{A \cap V\}| = 5$

$|X| = 30$

Probabilidade Condicional

$$P(V|A) \times P(A)$$
$$= P(A|V) \times P(V)$$

$$P(V|A) = \frac{P(A|V) \times P(V)}{P(A)}$$



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

Teorema de Bayes

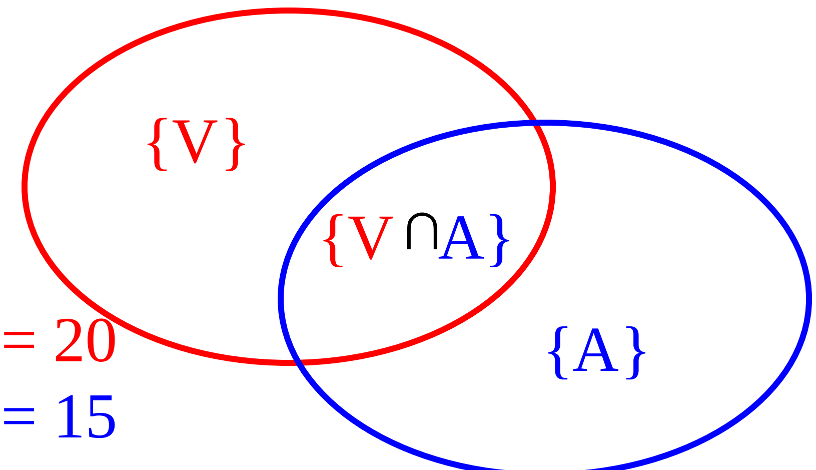
$$P(V|A) = \frac{P(A|V) \times P(V)}{P(A)}$$

Prior

Likelihood

Evidence

Posterior



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

Teorema de Bayes

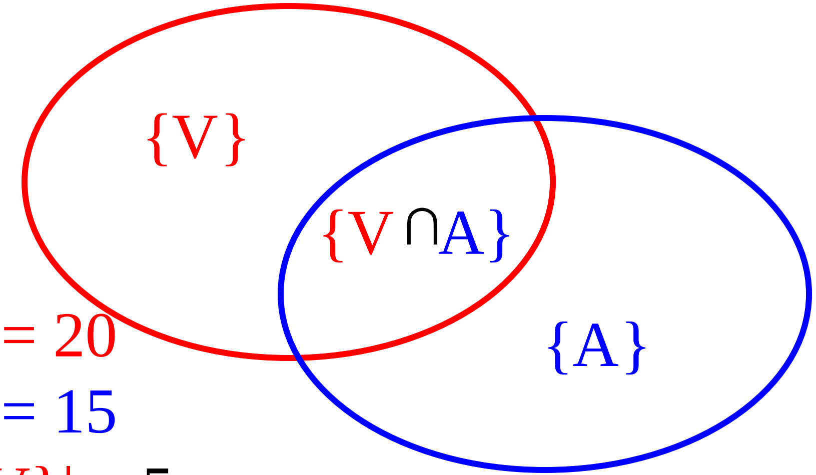
Prior

Likelihood

$$\frac{P(V|A) \times P(A)}{P(V)} = P(A|V)$$

Evidence

Posterior



$|\{V\}| = 20$

$|\{A\}| = 15$

$|\{A \cap V\}| = 5$

$|X| = 30$

Exemplo de Aplicação do Teorema de Bayes

- Dados:
  - Meningite causa rigidez no pescoço 50% das vezes
  - Probabilidade a priori de se ter meningite é de $1/50{,}000$
  - Probabilidade a priori de se ter rigidez no pescoço é de $1/20$
- **Problema:** se um paciente está com rigidez no pescoço (evidência), qual a probabilidade a posteriori que o paciente esteja com meningite?

Exemplo de Aplicação do Teorema de Bayes

- Dados:
  - Meningite causa rigidez no pescoço 50% das vezes
  - Probabilidade a priori de se ter meningite é de $1/50{,}000$
  - Probabilidade a priori de se ter rigidez no pescoço é de $1/20$
- **Problema:** se um paciente está com rigidez no pescoço (evidência), qual a probabilidade a posteriori que o paciente esteja com meningite?

$$P(R|M) = 0.5$$

Exemplo de Aplicação do Teorema de Bayes

- Dados:
  - Meningite causa rigidez no pescoço 50% das vezes
  - Probabilidade a priori de se ter meningite é de $1/50,000$
  - Probabilidade a priori de se ter rigidez no pescoço é de $1/20$
- **Problema:** se um paciente está com rigidez no pescoço (evidência), qual a probabilidade a posteriori que o paciente esteja com meningite?

$$P(R|M) = 0.5$$

$$P(M) = 1/50,000$$

Exemplo de Aplicação do Teorema de Bayes

- Dados:
  - Meningite causa rigidez no pescoço 50% das vezes
  - Probabilidade a priori de se ter meningite é de $1/50{,}000$
  - Probabilidade a priori de se ter rigidez no pescoço é de $1/20$
- **Problema:** se um paciente está com rigidez no pescoço (evidência), qual a probabilidade a posteriori que o paciente esteja com meningite?

$$P(R|M) = 0.5$$

$$P(M) = 1/50{,}000$$

$$P(R) = 1/20$$

Exemplo de Aplicação do Teorema de Bayes

- Dados:
  - Meningite causa rigidez no pescoço 50% das vezes
  - Probabilidade a priori de se ter meningite é de $1/50,000$
  - Probabilidade a priori de se ter rigidez no pescoço é de $1/20$
- **Problema:** se um paciente está com rigidez no pescoço (evidência), qual a probabilidade a posteriori que o paciente esteja com meningite?

$P(R|M) = 0.5$ $\qquad\qquad$ $P(M|R) = ?$

$P(M) = 1/50,000$

$P(R) = 1/20$

Exemplo de Aplicação do Teorema de Bayes

- Dados:
  - Meningite causa rigidez no pescoço 50% das vezes
  - Probabilidade a priori de se ter meningite é de $1/50,000$
  - Probabilidade a priori de se ter rigidez no pescoço é de $1/20$
- **Problema:** se um paciente está com rigidez no pescoço (evidência), qual a probabilidade a posteriori que o paciente esteja com meningite?

$$P(R|M) = 0.5$$

$$P(M) = 1/50,000$$

$$P(R) = 1/20$$

$$P(M|R) = \frac{P(R|M) \times P(M)}{P(R)}$$

Exemplo de Aplicação do Teorema de Bayes

- Dados:
  - Meningite causa rigidez no pescoço 50% das vezes
  - Probabilidade a priori de se ter meningite é de $1/50{,}000$
  - Probabilidade a priori de se ter rigidez no pescoço é de $1/20$
- **Problema:** se um paciente está com rigidez no pescoço (evidência), qual a probabilidade a posteriori que o paciente esteja com meningite?

$$P(R|M) = 0.5$$

$$P(M) = 1/50{,}000$$

$$P(R) = 1/20$$

$$P(M|R) = \frac{P(R|M) \times P(M)}{P(R)} = \frac{0.5 \times 0.00002}{0.05}$$

Exemplo de Aplicação do Teorema de Bayes

- Dados:
  - Meningite causa rigidez no pescoço 50% das vezes
  - Probabilidade a priori de se ter meningite é de $1/50{,}000$
  - Probabilidade a priori de se ter rigidez no pescoço é de $1/20$
- **Problema:** se um paciente está com rigidez no pescoço (evidência), qual a probabilidade a posteriori que o paciente esteja com meningite?

$$P(R|M) = 0.5$$

$$P(M) = 1/50{,}000$$

$$P(R) = 1/20$$

$$P(M|R) = \frac{P(R|M) \times P(M)}{P(R)} = \frac{0.5 \times 0.00002}{0.05}$$

$$= 0.0002\ (0.02\%)$$

Aprendizado Bayesiano

$$p(c_j|\mathbf{x}) = \frac{p(\mathbf{x}|c_j) \times p(c_j)}{p(\mathbf{x})}$$

- $p(c_j|\mathbf{x})$ = probabilidade da instância $\mathbf{x}$ pertencer a classe c_j

  Isto é o que queremos calcular!!
- $p(\mathbf{x}|c_j)$ = probabilidade da classe c_j gerar a instância $\mathbf{x}$

  Podemos imaginar que pertencer à classe $\mathbf{c}_\mathbf{j}$ faz com que se tenha a instância $\mathbf{x}$ com alguma probabilidade
- $p(c_j)$ = probabilidade de ocorrência da classe c_j

  Frequência da classe $\mathbf{c}_j$ na base de dados
- $p(\mathbf{x})$ = probabilidade de ocorrência da instância $\mathbf{x}$

  Podemos ignorar este valor, pois é o mesmo para todas as classes

Aprendizado Bayesiano

$$p(c_j|\mathbf{x}) = \frac{p(\mathbf{x}|c_j) \times p(c_j)}{p(\mathbf{x})}$$

- Para classificar a nova instância $\mathbf{x}$:

$$\hat{f}(\mathbf{x}) = \underset{k}{\operatorname{argmax}}[p(c_k|\mathbf{x})]$$

Nesse caso, $\hat{f}(\mathbf{x})$ é chamada de estimativa MAP (*Maximum A Posteriori*)

Assuma que temos duas classes

c_1 = gato, c_2 = cachorro

Um amigo comentou que possui um animal de estimação chamado Bob.

Classificar Bob como gato ou cachorro é equivalente a perguntar se é mais provável que Bob seja um gato ou um cachorro, ou seja, verificar qual probabilidade é maior:

$p(gato|bob)$ ou $p(cachorro|bob)$

$$p(\text{gato} \mid bob) = \frac{p(bob \mid \text{gato})\, p(\text{gato})}{p(bob)}$$

Assuma que temos duas classes

c_1 = gato, c_2 = cachorro

Um amigo comentou que possui um animal de estimação chamado Bob.

Classificar Bob como gato ou cachorro é equivalente a perguntar se é mais provável que Bob seja um gato ou um cachorro, ou seja, verificar qual probabilidade é maior:

$$p(gato|bob) \text{ ou } p(cachorro|bob)$$

Qual a probabilidade de um animal se chamar "bob" considerando que seja do um gato?

$$p(\text{gato} \mid bob) = \frac{p(bob \mid \text{gato}) \, p(\text{gato})}{p(bob)}$$

Assuma que temos duas classes

c_1 = gato, c_2 = cachorro

Um amigo comentou que possui um animal de estimação chamado Bob.

Classificar Bob como gato ou cachorro é equivalente a perguntar se é mais provável que Bob seja um gato ou um cachorro, ou seja, verificar qual probabilidade é maior:

$p(gato|bob)$ ou $p(cachorro|bob)$

Qual a probabilidade de um animal se chamar "bob" considerando que seja do um gato?

Qual a probabilidade de algum animal ser um gato?

$$p(\text{gato} \mid bob) = \frac{p(bob \mid \text{gato}\,)\, p(\text{gato})}{p(bob)}$$

Assuma que temos duas classes

c_1 = gato, c_2 = cachorro

Um amigo comentou que possui um animal de estimação chamado Bob.

Classificar Bob como gato ou cachorro é equivalente a perguntar se é mais provável que Bob seja um gato ou um cachorro, ou seja, verificar qual probabilidade é maior:

$p(gato|bob)$ ou $p(cachorro|bob)$

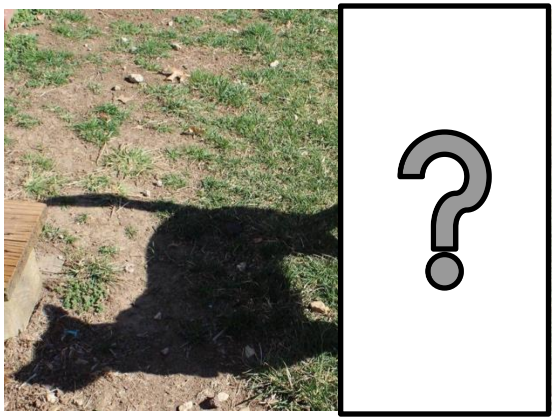
Qual a probabilidade de um animal se chamar "bob" considerando que seja do um gato?

Qual a probabilidade de algum animal ser um gato?

$$p(\text{gato} \mid bob) = \frac{p(bob \mid \text{gato})\, p(\text{gato})}{p(bob)}$$

Qual a probabilidade de algum animal se chamar "bob"? (irrelevante, pois é a mesma para todas as classes)

Este é Bob... seria Bob um Gato ou Cachorro?



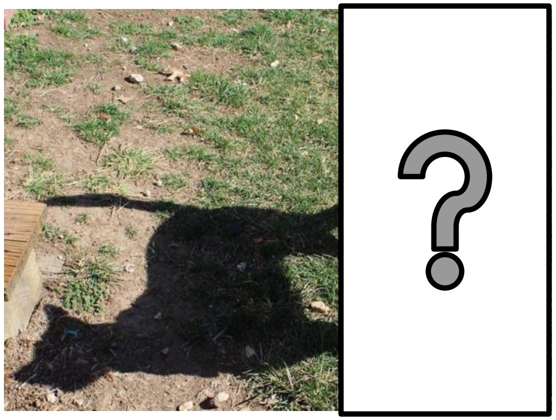
Animal Bob

Ainda bem que temos uma pequena base de dados com nomes e Espécies.

Vamos utilizá-la para aplicar o teorema de Bayes...

Nome	Espécie
Bob	Gato
Claudia	Cachorro
Bob	Cachorro
Bob	Cachorro
Alberto	Gato
Karin	Cachorro
Nina	Cachorro
Sergio	Gato

$$p(c_j \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid c_j)\, p(c_j)}{p(\mathbf{x})}$$



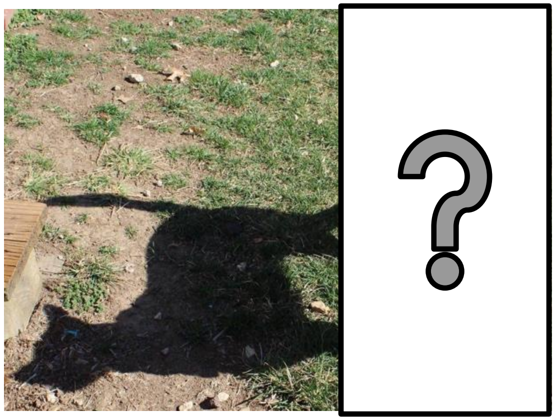
$$p(c_j \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid c_j)\, p(c_j)}{p(\mathbf{x})}$$

Animal Bob

$$p(\text{gato} \mid bob) = \frac{p(bob \mid \text{gato}) \quad * \; p(\text{gato})}{p(bob)}$$

$$p(\text{gato} \mid bob) = \frac{1/3 \times 3/8}{p(bob)} = \frac{0.125}{p(bob)}$$

Nome	Espécie
Bob	Gato
Claudia	Cachorro
Bob	Cachorro
Bob	Cachorro
Alberto	Gato
Karin	Cachorro
Nina	Cachorro
Sergio	Gato



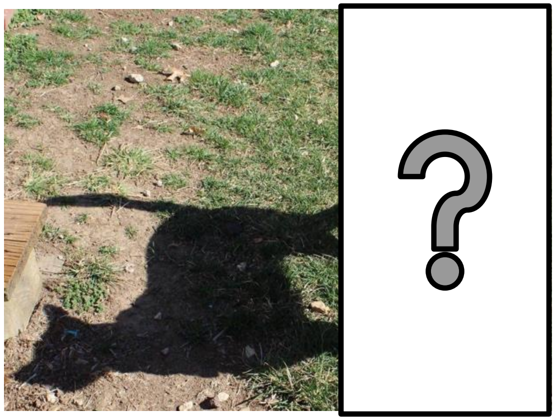
$$p(c_j \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid c_j)\, p(c_j)}{p(\mathbf{x})}$$

Animal Bob

$$p(dog \mid bob) = \frac{p(bob \mid dog) \; * \, p(dog)}{p(bob)}$$

$$p(dog \mid bob) = \frac{2/5 \times 5/8}{p(bob)} = \frac{0.250}{p(bob)}$$

Nome	Espécie
Bob	Gato
Claudia	Cachorro
Bob	Cachorro
Bob	Cachorro
Alberto	Gato
Karin	Cachorro
Nina	Cachorro
Sergio	Gato



$$p(c_j \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid c_j)\, p(c_j)}{p(\mathbf{x})}$$

Animal Bob

Nome	Espécie
Bob	Gato
Claudia	Cachorro
Bob	Cachorro
Bob	Cachorro
Alberto	Gato
Karin	Cachorro
Nina	Cachorro
Sergio	Gato

$$p(\text{gato} \mid bob) = \frac{1/3 \times 3/8}{p(bob)} = \frac{0.125}{p(bob)}$$

$$p(\text{dog} \mid bob) = \frac{2/5 \times 5/8}{p(bob)} = \frac{0.250}{p(bob)}$$

É mais provável que Bob seja um cachorro

Entendendo o Cálculo da Evidência

- Até então ignoramos o cálculo da evidência $p(\mathbf{x})$
  - Por se tratar apenas de um termo normalizador
  - Por ser constante para todas as classes
    - Ex: $p(bob)$

- Mas como calcular $p(\mathbf{x})$?

Entendendo o Cálculo da Evidência

- Voltando ao exemplo anterior:

$$p(\text{gato} \mid bob) = \frac{1/3 * 3/8}{p(bob)} = \frac{0.125}{p(bob)} =$$

$$p(\text{dog} \mid bob) = \frac{2/5 * 5/8}{p(bob)} = \frac{0.250}{p(bob)} =$$

Nome	Espécie
Bob	Gato
Claudia	Cachorro
Bob	Cachorro
Bob	Cachorro
Alberto	Gato
Karin	Cachorro
Nina	Cachorro
Sergio	Gato

Interpretação natural:

- *p*(*bob*) é a probabilidade *a priori* de algum animal se chamar Bob

Entendendo o Cálculo da Evidência

- Voltando ao exemplo anterior:

$$p(\text{gato} \mid bob) = \frac{1/3 * 3/8}{3/8} = \frac{0.125}{3/8} =$$

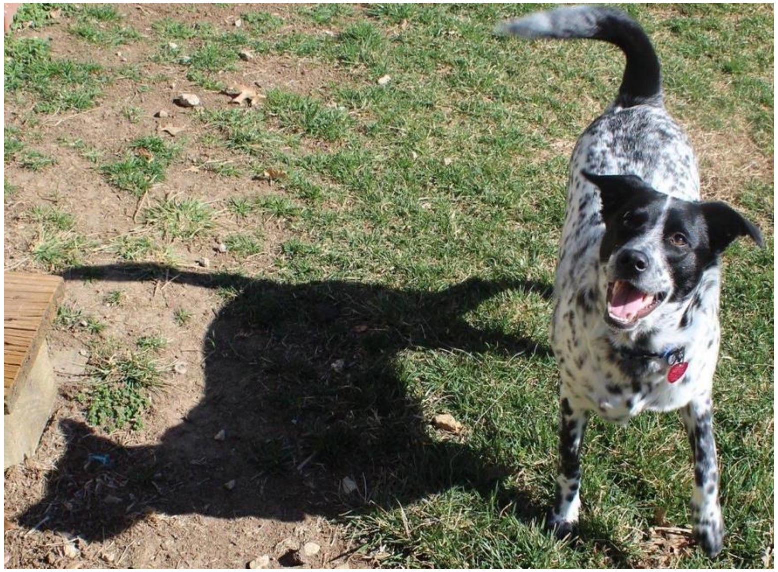
$$p(\text{dog} \mid bob) = \frac{2/5 * 5/8}{3/8} = \frac{0.250}{3/8} =$$

Nome	**Espécie**
Bob	Gato
Claudia	Cachorro
Bob	Cachorro
Bob	Cachorro
Alberto	Gato
Karin	Cachorro
Nina	Cachorro
Sergio	Gato

Interpretação natural:

- *p*(*bob*) é a probabilidade *a priori* de algum animal se chamar Bob

Bob é um cachorro!



Animal Bob

$$p(\textcolor{blue}{\text{gato}} \mid bob) = \frac{1/3 * 3/8}{3/8} = \frac{0.125}{3/8} \textcolor{blue}{= 0.33\ (33\%)}$$

$$p(\textcolor{red}{\text{dog}} \mid bob) = \frac{2/5 * 5/8}{3/8} = \frac{0.250}{3/8} = \textcolor{red}{0.66\ (66\%)}$$

Estendendo o Teorema de Bayes para vários atributos

$$p\left(c_j \middle| x_1 = a, x_2 = b, \dots x_m = z\right) = \frac{p\left(x_1 = a, x_2 = b, \dots x_m = z \middle| c_j\right) \times p(c_j)}{p(x_1 = a, x_2 = b, \dots x_m = z)}$$

Mas há um problema!

Estendendo o Teorema de Bayes para vários atributos

$$p\left(c_j \middle| x_1 = a, x_2 = b, \dots x_m = z\right) = \frac{p\left(x_1 = a, x_2 = b, \dots x_m = z \middle| c_j\right) \times p(c_j)}{p(x_1 = a, x_2 = b, \dots x_m = z)}$$

Mas há um problema!

Estimar a probabilidade condicional $p\left(x_1 = a, x_2 = b, \dots x_m = z \middle| c_j\right)$ *e a evidência* $p(x_1 = a, x_2 = b, \dots x_m = z)$ demandaria uma quantidade mínima de exemplos de cada combinação possível de valores dos atributos $x_1, x_2, \dots, x_m$

IMPRATICÁVEL, ESPECIALMENTE PARA QUANTIDADES ELEVADAS DE ATRIBUTOS!!

Possível Solução?

- Assumir independência entre atributos!

[independência]

$$p(x_1, x_2, ..., x_m) = p(x_1) \times p(x_2) \times ... \times p(x_m)$$

[independência condicional]

$$p(x_1, x_2, ..., x_m \mid c_j) = p(x_1 \mid c_j) \times p(x_2 \mid c_j) \times ... \times p(x_m \mid c_j)$$

Possível Solução?

Re-escrevendo o Teorema de Bayes com a hipótese de independência condicional:

$$p(c_j \mid x_1, x_2, \ldots, x_m) = \frac{p(c_j) \times \prod_{i=1}^{m} p(x_i \mid c_j)}{\prod_{i=1}^{m} p(x_i)}$$

Classificador Naïve Bayes

- Mais simples e bem difundido classificador baseado no Teorema de Bayes

Thomas Bayes
1702 - 1761

Naïve Bayes

- Naïve = ingênuo
  - Hipótese de independência entre atributos é quase sempre violada!
  - Na prática, Naïve Bayes se mostra bastante competitivo!

$$p(c_j \mid x_1, x_2, \ldots, x_m) = \frac{p(c_j) \times \prod_{i=1}^{m} p(x_i \mid c_j)}{\prod_{i=1}^{m} p(x_i)}$$

Exemplo:

(Witten & Frank, 2005)

Outlook (A_1)			Temperature (A_2)			Humidity (A_3)			Windy (A_4)			Play (B)	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny			Hot			High			False				
Overcast			Mild			Normal			True				
Rainy			Cool										
Sunny			Hot			High			False				
Overcast			Mild			Normal			True				
Rainy			Cool										

Estimar a probabilidade das classes (jogar tênis ou não) com base nos atributos preditivos (referentes ao clima do dia)

$$p(c_j \mid x_1, x_2, \ldots, x_m) = \frac{p(c_j) \times \prod_{i=1}^{m} p(x_i \mid c_j)}{\prod_{i=1}^{m} p(x_i)}$$

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**

Exemplo:

(Witten & Frank, 2005)

Outlook (A_1)			**Temperature (A_2)**			**Humidity (A_3)**			**Windy (A_4)**			**Play (B)**	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	Hot			High			False				
Overcast			Mild			Normal			True				
Rainy			Cool										
Sunny			Hot			High			False				
Overcast			Mild			Normal			True				
Rainy			Cool										

Estimar a probabilidade das classes (jogar tênis ou não) com base nos atributos preditivos (referentes ao clima do dia)

$$p(c_j \mid x_1, x_2, \ldots, x_m) = \frac{p(c_j) \times \prod_{i=1}^{m} p(x_i \mid c_j)}{\prod_{i=1}^{m} p(x_i)}$$

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**

Exemplo:

(Witten & Frank, 2005)

Outlook (A_1)			Temperature (A_2)			Humidity (A_3)			Windy (A_4)			Play (B)	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	Hot			High			False				
Overcast	4	0	Mild			Normal			True				
Rainy			Cool										
Sunny			Hot			High			False				
Overcast			Mild			Normal			True				
Rainy			Cool										

Estimar a probabilidade das classes (jogar tênis ou não) com base nos atributos preditivos (referentes ao clima do dia)

$$p(c_j \mid x_1, x_2, \ldots, x_m) = \frac{p(c_j) \times \prod_{i=1}^{m} p(x_i \mid c_j)}{\prod_{i=1}^{m} p(x_i)}$$

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**

Exemplo:

(Witten & Frank, 2005)

Outlook (A_1)			Temperature (A_2)			Humidity (A_3)			Windy (A_4)			Play (B)	
	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			Hot			High			False				
Overcast			Mild			Normal			True				
Rainy			Cool										

Estimar a probabilidade das classes (jogar tênis ou não) com base nos atributos preditivos (referentes ao clima do dia)

$$p(c_j \mid x_1, x_2, \ldots, x_m) = \frac{p(c_j) \times \prod_{i=1}^{m} p(x_i \mid c_j)}{\prod_{i=1}^{m} p(x_i)}$$

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**

Exemplo:

(Witten & Frank, 2005)

Outlook (A_1)			Temperature (A_2)			Humidity (A_3)			Windy (A_4)			Play (B)	
	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			High			False				
Overcast			Mild			Normal			True				
Rainy			Cool										

Estimar a probabilidade das classes (jogar tênis ou não) com base nos atributos preditivos (referentes ao clima do dia)

$$p(c_j \mid x_1, x_2, \ldots, x_m) = \frac{p(c_j) \times \prod_{i=1}^{m} p(x_i \mid c_j)}{\prod_{i=1}^{m} p(x_i)}$$

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

Exemplo:

(Witten & Frank, 2005)

Outlook (A_1)			**Temperature (A_2)**			**Humidity (A_3)**			**Windy (A_4)**			**Play (B)**	
	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			High			False				
Overcast	4/9	0/5	Mild			Normal			True				
Rainy			Cool										

Estimar a probabilidade das classes (jogar tênis ou não) com base nos atributos preditivos (referentes ao clima do dia)

$$p(c_j \mid x_1, x_2, \ldots, x_m) = \frac{p(c_j) \times \prod_{i=1}^{m} p(x_i \mid c_j)}{\prod_{i=1}^{m} p(x_i)}$$

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**

Exemplo:

(Witten & Frank, 2005)

Outlook (A_1)			Temperature (A_2)			Humidity (A_3)			Windy (A_4)			Play (B)	
	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								

Estimar a probabilidade das classes (jogar tênis ou não) com base nos atributos preditivos (referentes ao clima do dia)

$$p(c_j \mid x_1, x_2, \ldots, x_m) = \frac{p(c_j) \times \prod_{i=1}^{m} p(x_i \mid c_j)}{\prod_{i=1}^{m} p(x_i)}$$

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

Continuando...

(Witten & Frank, 2005)

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 dia:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	**???**

↓

Continuando...

(Witten & Frank, 2005)

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 dia:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	???

↓

P(Yes|Sunny, Cool, High, True) = (2/9 × 3/9 × 3/9 × 3/9 × 9/14) / P(Sunny, Cool, High, True)

P(No|Sunny, Cool, High, True) = (3/5 × 1/5 × 4/5 × 3/5 × 5/14) / P(Sunny, Cool, High, True)

↓

P(Yes|Sunny, Cool, High, True) = **0.0053** / P(Sunny, Cool, High, True)

P(No|Sunny, Cool, High, True) = **0.0206** / P(Sunny, Cool, High, True)

➡ Play = No

Problema da Freqüência Zero

- O que acontece se um determinado valor de atributo não aparece na base de treinamento, mas aparece no exemplo de teste?
  - Por exemplo: "Outlook = Overcast" para classe "No"
  - Probabilidade correspondente será zero
    - P(Overcast | "No") = 0
  - *Probabilidade a posteriori* será também zero!
    - P("No" | Overcast, ...) = 0
    - Não importa as probabilidades referentes aos demais atributos !
    - Muito radical, especialmente considerando que a base de treinamento pode não ser totalmente representativa
      - Por exemplo, classes minoritárias com instâncias raras

Problema da Freqüência Zero

- Possível solução (**Estimador de Laplace**):
  - Adicionar 1 unidade fictícia para cada combinação de valor-classe
    - Como resultado, probabilidades nunca serão zero !
  - Exemplo (atributo Outlook – classe No):

$$\frac{3+1}{5+3} \qquad \frac{0+1}{5+3} \qquad \frac{2+1}{5+3}$$

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

  - Nota: Deve ser feito para todas as classes, para não inserir viés nas probabilidades de apenas uma classe

Atributos Numéricos

Atributos Numéricos

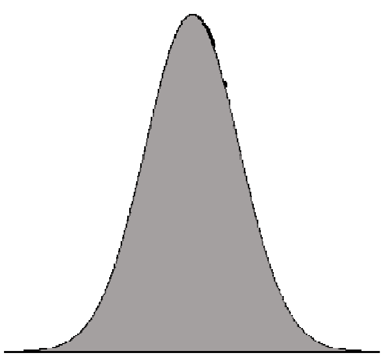
- **Alternativa 1:** Discretização

Atributos Numéricos

- **Alternativa 1:** Discretização
- **Alternativa 2:** Assumir ou estimar alguma função de densidade de probabilidade para estimar as probabilidades
  - Usualmente distribuição Gaussiana (Normal)

$$\mu_j = \frac{1}{N}\sum_{i=1}^{N} x_j^{(i)} \qquad \sigma_j^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_j^{(i)} - \mu_j\right)^2$$

$$f(x_j^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{\left(x_j^{(i)}-\mu_j\right)^2}{2\sigma_j^2}}$$



Karl Gauss
1777-1855

(Witten & Frank, 2005)

Estatísticas para "weather"

(Witten & Frank, 2005)

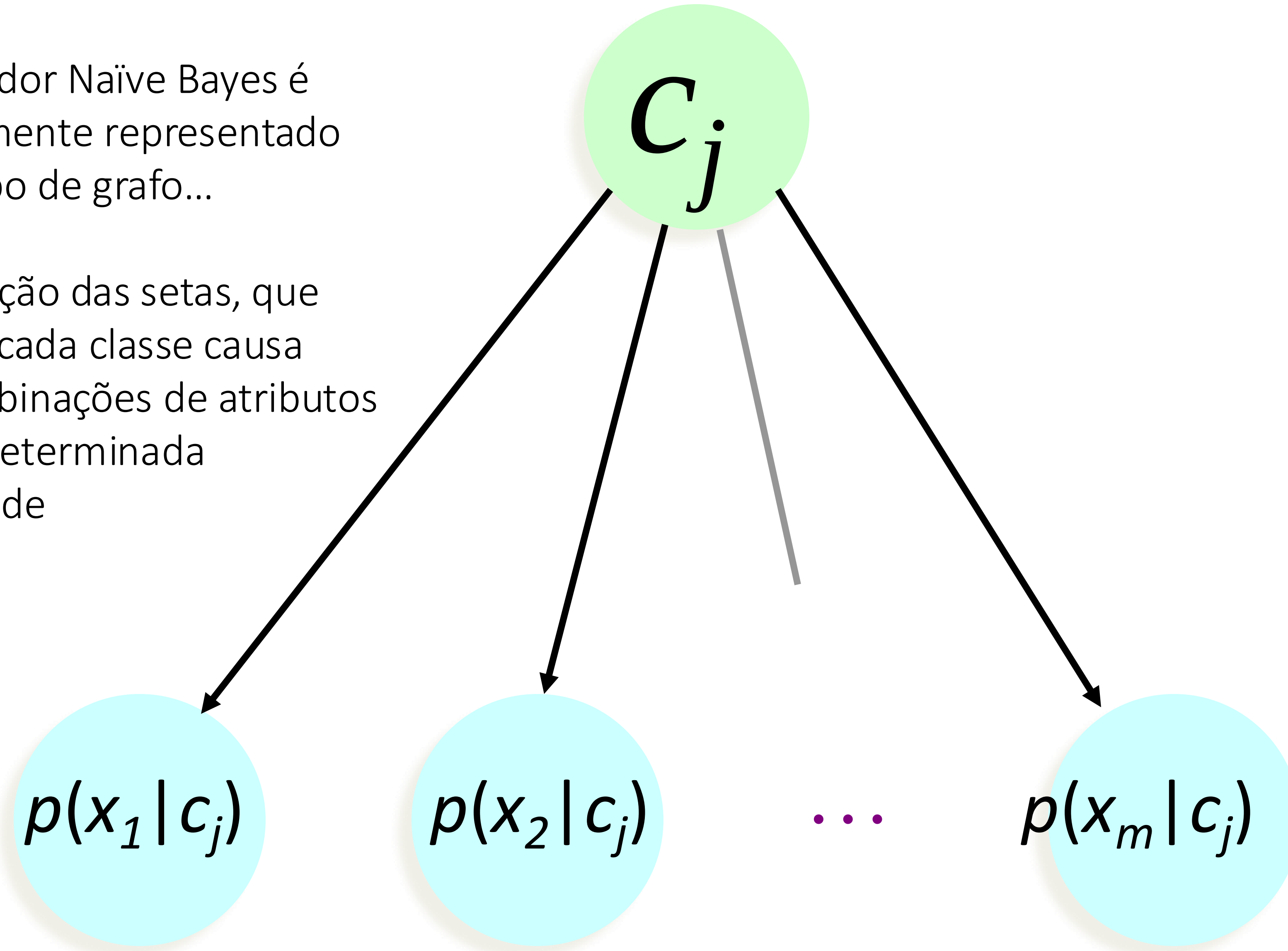
Outlook			Temperature		Humidity		Windy			Play	
	Yes	*No*	*Yes*	*No*	*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	64, 68,	65, 71,	65, 70,	70, 85,	False	6	2	9	5
Overcast	4	0	69, 70,	72, 80,	70, 75,	90, 91,	True	3	3		
Rainy	3	2	72, …	85, …	80, …	95, …					
Sunny	2/9	3/5	μ =73	μ =75	μ =79	μ =86	False	6/9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

- Valor de **densidade**:

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

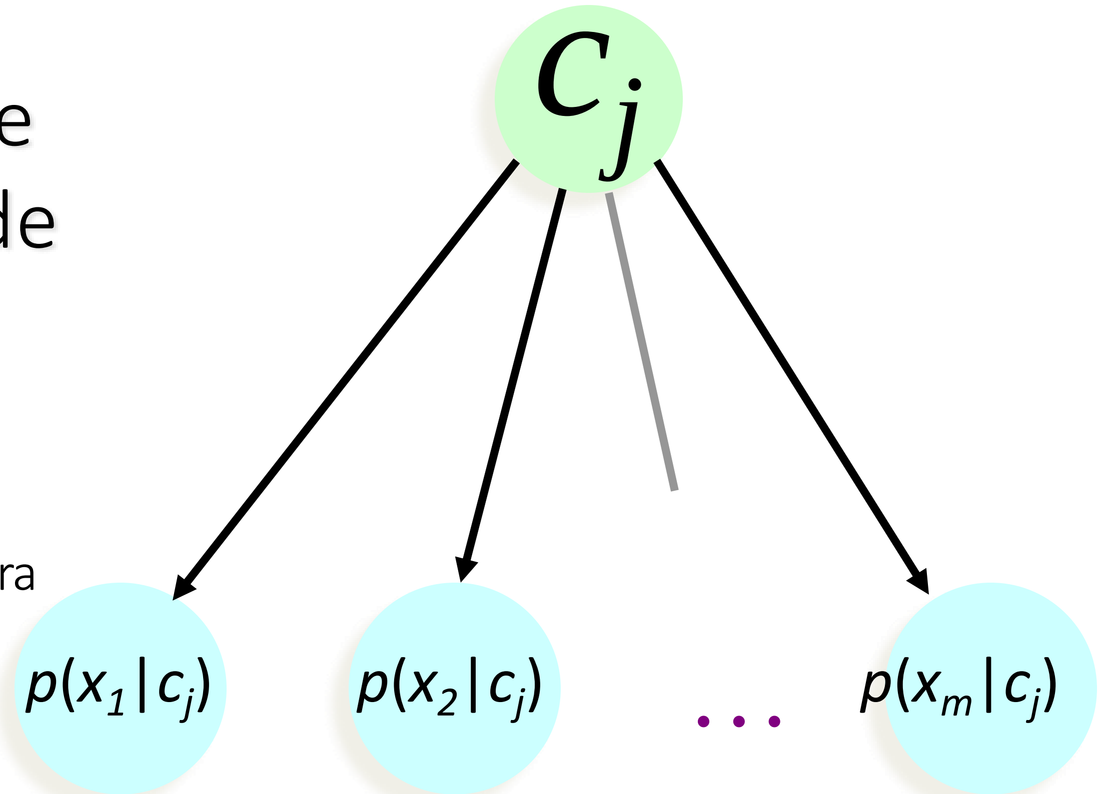
O classificador Naïve Bayes é frequentemente representado por este tipo de grafo…

Nota a direção das setas, que dizem que cada classe causa certas combinações de atributos com uma determinada probabilidade



Naïve Bayes é rápido e eficiente em termos de memória

As probabilidades podem ser calculadas com uma única varredura da base e armazenadas em uma (pequena) tabela…



Sexo	**> 190$_{cm}$**	
Masc	Sim	0.15
	Não	0.85
Fem	Sim	0.01
	Não	0.99

Sexo	**Cabelo longo**	
Masc	Sim	0.05
	Não	0.95
Fem	Sim	0.70
	Não	0.30

Sexo
Masc
Fem

Naïve Bayes NÃO É sensível a atributos irrelevantes...

Naïve Bayes NÃO É sensível a atributos irrelevantes...

Suponha que estejamos tentando rotular a faixa etária de uma pessoa baseado em vários atributos, dentre eles a cor dos olhos. (É claro que a cor dos olhos é irrelevante na previsão da faixa etária de uma pessoa)

Naïve Bayes NÃO É sensível a atributos irrelevantes...

Suponha que estejamos tentando rotular a faixa etária de uma pessoa baseado em vários atributos, dentre eles a cor dos olhos. (É claro que a cor dos olhos é irrelevante na previsão da faixa etária de uma pessoa)

$$p(\text{Jessica} \mid c_j) = p(\text{olho= castanho} \mid c_j) * p(\text{ cabelo\_longo= sim} \mid c_j) * \ldots.$$

$$p(\text{Jessica} \mid \text{Adulto}) = \textit{9,000/10,000} \;\; * \;\; \textit{9,975/10,000} \;\; * \ldots.$$

$$p(\text{Jessica} \mid \text{Idoso}) \;\; = \textit{9,001/10,000} \;\; * \;\; \textit{2/10,000} \;\; * \ldots.$$

Quase o mesmo valor!

No entanto, estamos assumindo que temos estimativas boas o suficiente: quanto mais dados, melhor!

Vantagens/Desvantagens do Naïve Bayes

- Vantagens:
  - Rápido para treinar (varredura única)
  - Rápido para classificar
  - Insensível a atributos irrelevantes
  - Lida com dados discretos e contínuos
  - Lida bem com fluxos de dados (data streams)
- Desvantagem:
  - Assume independência dos atributos
    - Caso haja alta redundância entre atributos, seleção de atributos resolve o problema!
    - Caso contrário, utilizar abordagem mais robusta (ex: Redes Bayesianas)

Sugestão de Leituras

- Seção 5.3 (Tan et al., 2006)
- Capítulo 5 (Faceli et al., 2011)

Créditos e Referências

Slides adaptados dos originais gentilmente cedidos por:

- Rodrigo Coelho Barros (PUCRS)
- André Carvalho, Eduardo Hruschka, Ricardo Campello (ICMC-USP)
- Pang-Ning Tan (Michigan State University)
- Eamon Keogh (University of California at Riverside)
  - http://www.cs.ucr.edu/~eamonn/
  - eamonn@cs.ucr.edu
- Jeff Howbert (University of Washington)

- Tan, P. N., Steinbach, M., Kumar, V. **Introduction to Data Mining.** Addison-Wesley, 2005. 769 p.

- Faceli et al. **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. LTC, 2011. 378 p.

- Murphy, K. P. **Machine Learning – A Probabilistic Approach**. MIT Press, 2012. 1071 p.