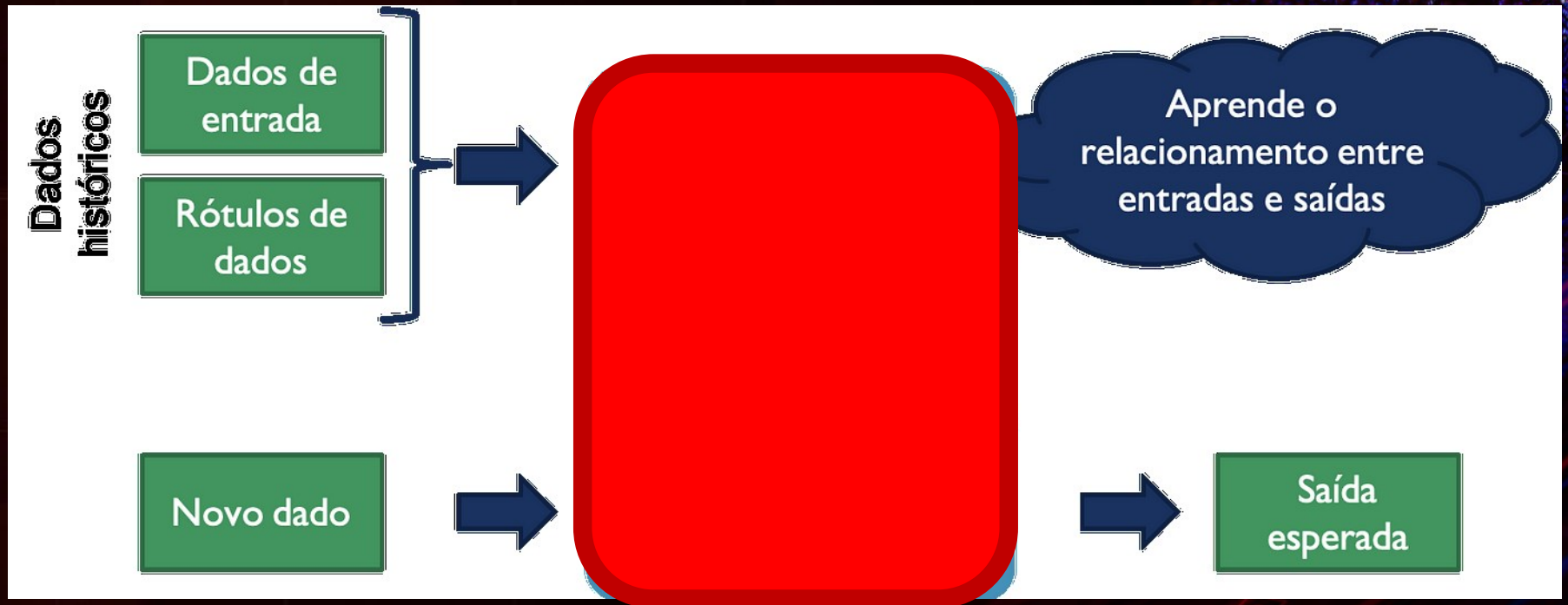
# PROCESSAMENTO DE LINGUAGEM NATURAL

## Aprendizado Bayesiano

# TÓPICOS

1. Aprendizado Bayesiano
2. Classificador Naïve Bayes
3. Análise

# CLASSIFICAÇÃO



# TEOREMA DE BAYES

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

**Mas há um problema!**

- Estimar a probabilidade condicional $p(x_1 = a, x_2 = b, \dots x_m = z|c_j)$ *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!**

$$p(x_1, x_2, \ldots, x_m) = p(x_1) \square p(x_2) \square \ldots \square p(x_m)$$ [independência]

$$p(x_1, x_2, \ldots, x_m \mid c_j) = p(x_1 \mid c_j) \square p(x_2 \mid c_j) \square \ldots \square p(x_m \mid c_j)$$ [independência condicional]

- **Reescrevendo o Teorema de Bayes com a hipótese de independência condicional:**

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

# TÓPICOS

1. Aprendizado Bayesiano
2. Classificador Naïve Bayes
3. Análise

# CLASSIFICADOR *NAÏVE BAYES*

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

Thomas Bayes
1702 - 1761

# NAÏVE BAYES

- Também conhecido por *Idiot Bayes* ou *Simple Bayes*
- *Naïve* = ingênuo
- Hipótese de independência entre atributos é quase sempre violada!
- Na prática, porém, *Naïve Bayes* se mostra bastante competitivo!

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

# EXEMPLO

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

| 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 |

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

# EXEMPLO

| 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 | | | | | | | | |

| 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 |

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

# EXEMPLO

| 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 |

| 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 | | | | | | | | |

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

P(Yes|Sunny, Cool, High, True) = (2/9x3/9x3/9x3/9x9/14) / P(Sunny, Cool, High, True)
P(No|Sunny, Cool, High, True) = (3/5x1/5x4/5x3/5x5/14) / P(Sunny, Cool, High, True)

# EXEMPLO

| 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 | | | | | | | | |

| Outlook | Temp. | Humidity | Windy | Play |
|---|---|---|---|---|
| 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

| 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 |

# PROBLEMA: FREQUÊNCIA ZERO

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

# PROBLEMA: FREQUÊNCIA 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):

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

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

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

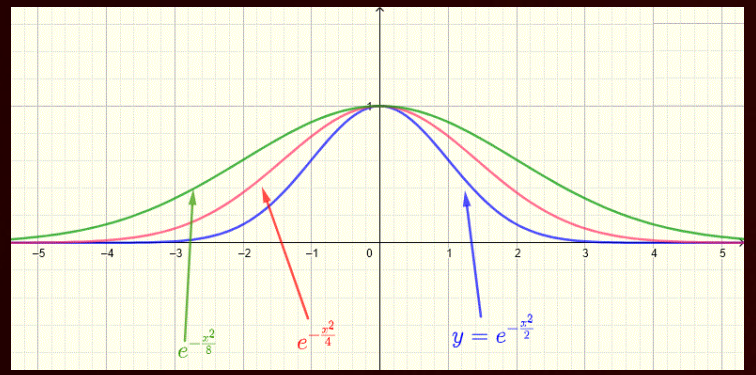
# ATRIBUTOS NUMÉRICOS

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

$$\mu_j = \frac{1}{N}\sum_{i=1}^{N} x_j^{(i)}$$

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

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



# ESTATÍSTICAS

| 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 | 64, 68, 69, 70, 72, … | 65, 71, 72, 80, 85, … | 65, 70, 70, 75, 80, … | 70, 85, 90, 91, 95, … | False | 6 | 2 | 9 | 5 |
| Overcast | 4 | 0 | | | | | True | 3 | 3 | | |
| Rainy | 3 | 2 | | | | | | | | | |
| 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 | | | | | | | | | |

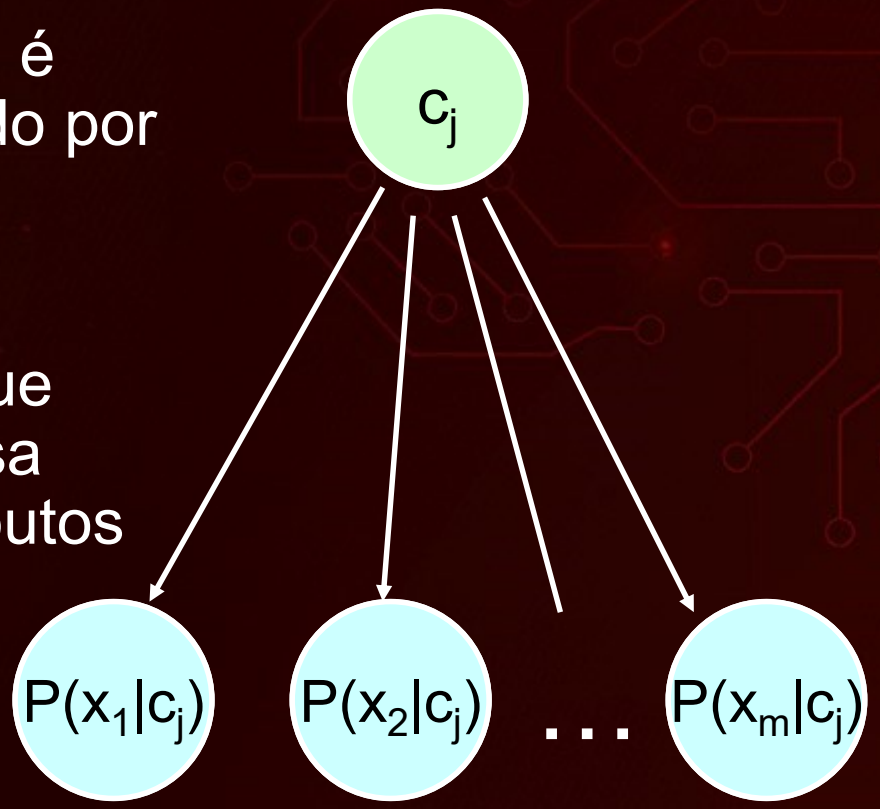
- Valor de densidade:

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

# REPRESENTAÇÃO

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

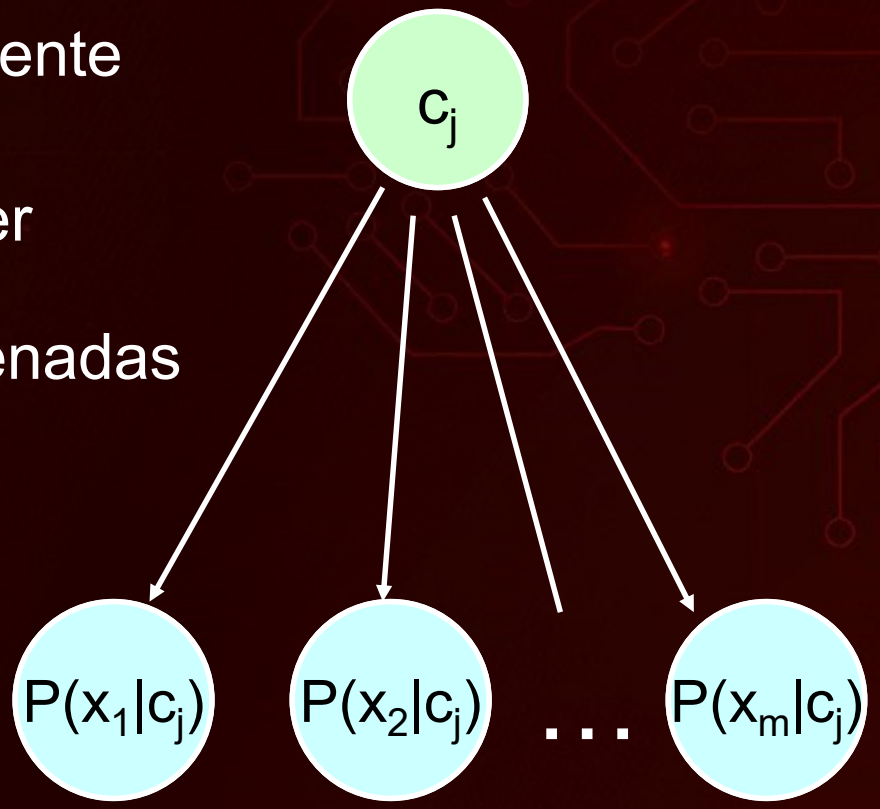
- Note a direção das setas, que dizem que cada classe causa certas combinações de atributos com uma determinada probabilidade



# REPRESENTAÇÃO

- 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 |



# TÓPICOS

1. Aprendizado Bayesiano
2. Classificador Naïve Bayes
3. Análise

# ATRIBUTOS IRRELEVANTES

- **Naïve Bayes NÃO É sensível a atributos irrelevantes...**
- **Suponha que estejamos tentando rotular o gênero de uma pessoa baseado em vários atributos, dentre eles a cor dos olhos. (É claro que a cor dos olhos é irrelevante na previsão do gênero de uma pessoa)**

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

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

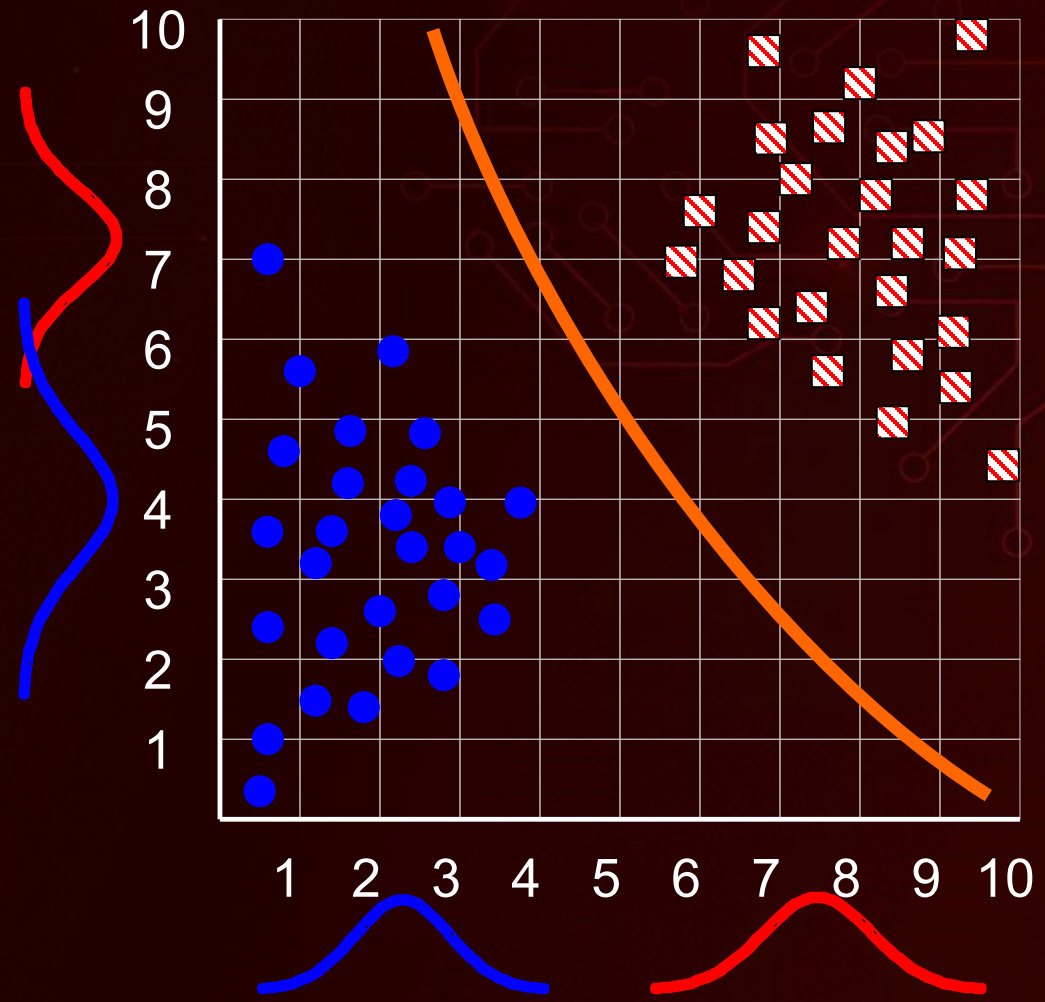
$p(\text{Jessica} \mid \text{Masc}) = 9{,}001/10{,}000 * 2/10{,}000 * \ldots.$

Quase o mesmo valor!

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

# FRONTEIRA DE DECISÃO

- O classificador Naïve Bayes gera uma fronteira de decisão quadrática



# CONCLUSÕES

- 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)

# O QUE VIMOS?

- Aprendizado Bayesiano
- Classificador Naïve Bayes
- Análise

# PRÓXIMA VIDEOAULA

- **Prática: Aprendizado de Máquina**