

# Naive Bayes e Redes Bayesianas

Breno Sanchi Cardoso do Amaral  
Nathália Lobo da Silva

27/01/2025

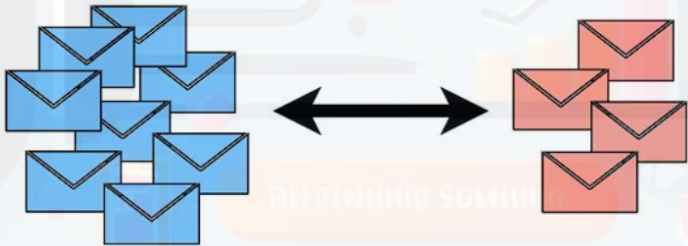


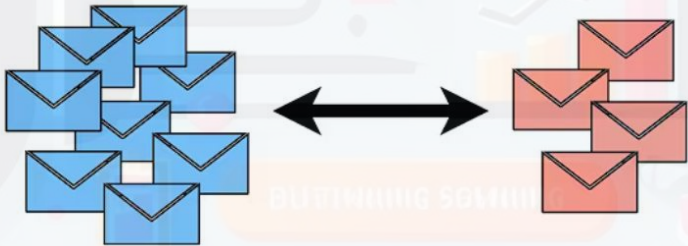
# Naive Bayes

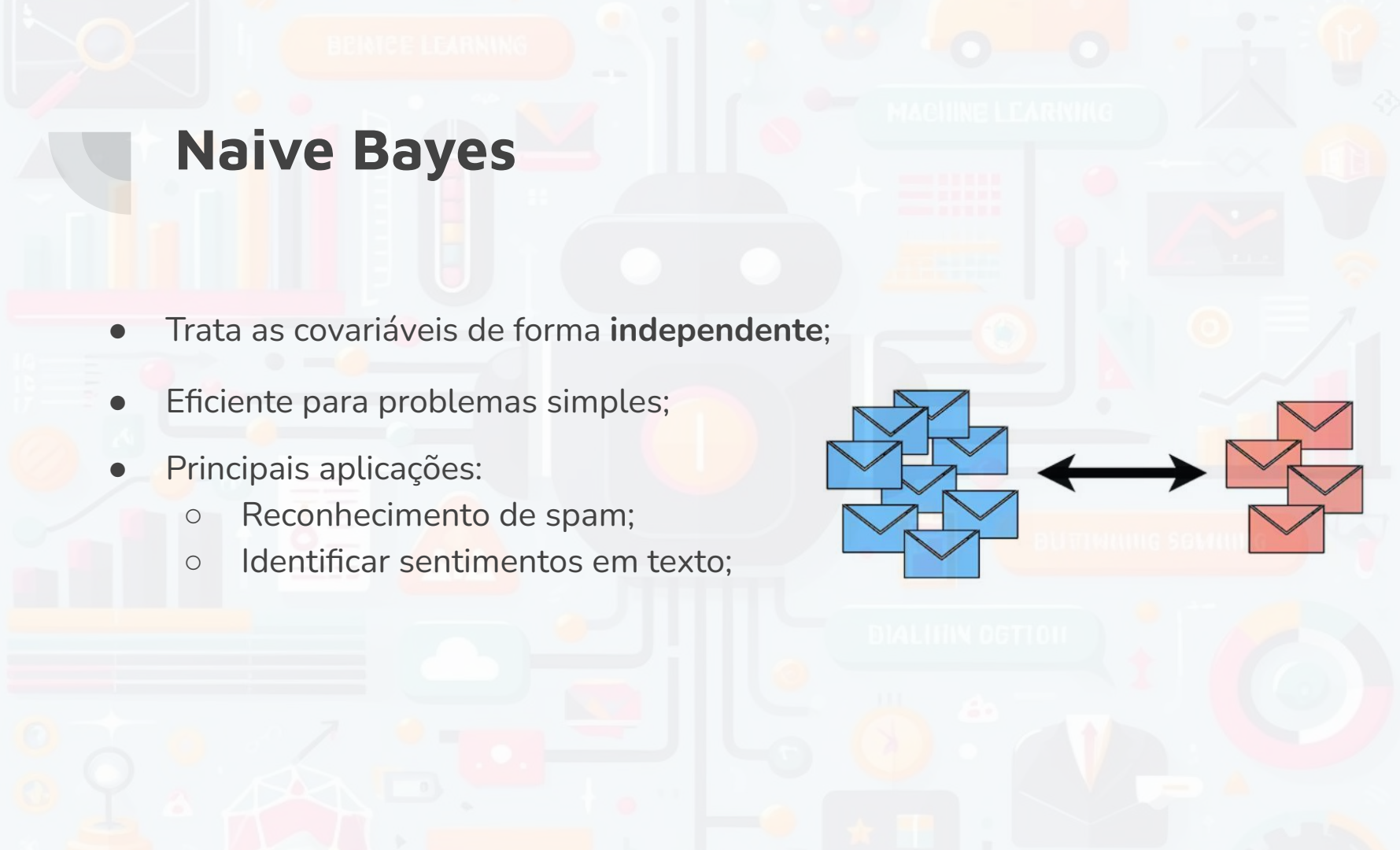
- É um algoritmo de **classificação**;
- Seu aprendizado é:
  - supervisionado;
  - baseado em modelo;
  - em batch (offline).
- Suas covariáveis podem ser numéricas ou categóricas, mas funciona melhor com **categóricas**.

# Naive Bayes

- Trata as covariáveis de forma **independente**;
- Eficiente para problemas simples;
- Principais aplicações:
  - Reconhecimento de spam;
  - Identificar sentimentos em texto;



- # Naive Bayes
- Trata as covariáveis de forma **independente**;
  - Eficiente para problemas simples;
  - Principais aplicações:
    - Reconhecimento de spam;
    - Identificar sentimentos em texto;
- 
- The diagram illustrates the Naive Bayes model for sentiment analysis. It shows a cluster of blue envelopes on the left and a cluster of red envelopes on the right, connected by a double-headed arrow. The background is a collage of various icons related to machine learning, data analysis, and technology.



# Como funciona?

- Ajusta um modelo que estima a probabilidade de uma nova observação pertencer a cada classe.

- Utiliza o Teorema de Bayes:

- X: Covariáveis;
- C: Classe (label).

$$\overbrace{P(C|X)}^{\text{Posteriori}} = \frac{\overbrace{P(X|C)}^{\text{Verossimilhança}} \overbrace{P(C)}^{\text{Priori}}}{\underbrace{P(X)}_{\text{Marginalização}}}$$

$$P(C|X_1, X_2, \dots, X_n) = \frac{P(X_1, X_2, \dots, X_n|C)P(C)}{P(X_1, X_2, \dots, X_n)}$$

# Como funciona?

- Com base nas covariáveis, compara a probabilidade de cada classe e escolhe a mais provável.

$$P(C = c_1|X) \quad vs \quad P(C = c_2|X)$$

$$\frac{P(X|C = c_1)P(C = c_1)}{P(X)} \quad vs \quad \frac{P(X|C = c_2)P(C = c_2)}{P(X)}$$

$$P(X|C = c_1)P(C = c_1) \quad vs \quad P(X|C = c_2)P(C = c_2)$$

# Como funciona?

- Qual a condição para que a equação abaixo seja verdadeira?

$$P(X_1, X_2, \dots, X_n|C) = P(X_1|C)P(X_2|C)\dots P(X_n|C)$$



# Como funciona?

- Qual a condição para que a equação abaixo seja verdadeira?

$$P(X_1, X_2, \dots, X_n|C) = P(X_1|C)P(X_2|C)...P(X_n|C)$$

Quando as covariáveis X's são **condicionalmente independentes!**

- Logo:

$$P(X_1|C = c_1)...P(X_n|C = c_1) P(C = c_1) \quad vs \quad P(X_1|C = c_2)...P(X_n|C = c_2) P(C = c_2)$$

# Redes Bayesianas

Fornecem uma estrutura intuitiva e abrangente para modelar as dependências entre as variáveis em dados estáticos

- Teoria dos grafos
- Teorema de Bayes



# Teoria dos grafos

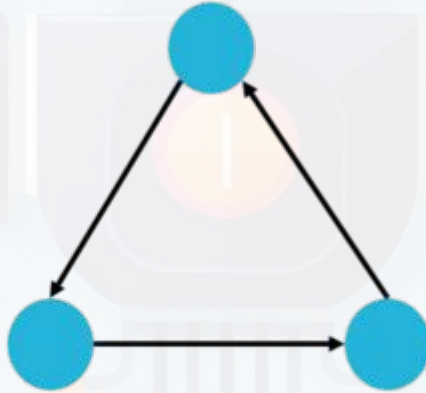
- **Grafos:** um conjunto de variáveis aleatórias que estão relacionados de certa forma
- **Nós/Vértices:** covariáveis
- **Arcos/Links:** as relações entre as covariáveis

Um grafo  $G = (V, A)$  consiste em um conjunto não vazio  $V$  de nós e um conjunto finito (mas possivelmente vazio)  $A$  de pares de vértices, ou seja, arcos

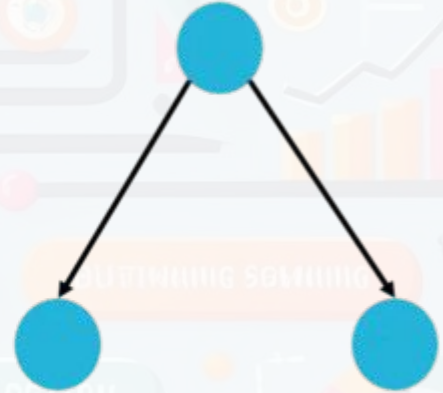
# DAG (Directed Acyclic Graph)



An undirected graph



A directed graph which isn't acyclic

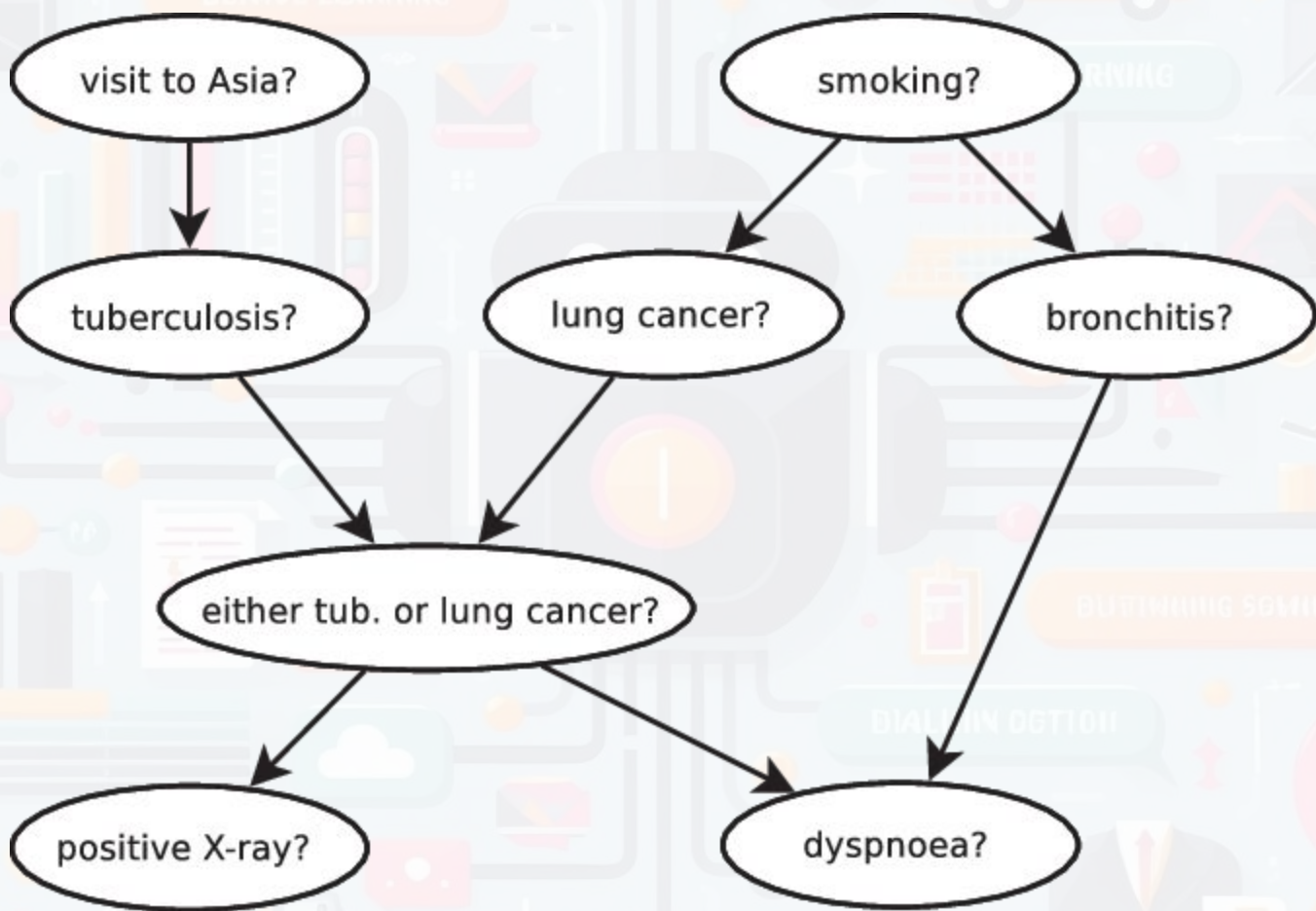


A directed acyclic graph

# Redes Bayesianas

É um modelo de Aprendizado de Máquina que captura dependências entre variáveis aleatórias  $X = \{X_1, \dots, X_p\}$  com um Gráfico Acíclico Direcionado (DAG), em que cada nó  $v_i \in V$  corresponde a uma variável aleatória  $X_i$ .

- Previsão de falhas em máquinas
- Medicina e saúde
- Biologia (genética)



# Propriedades

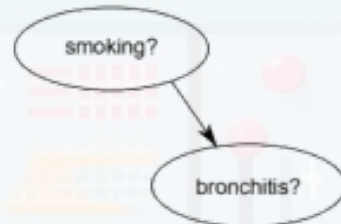
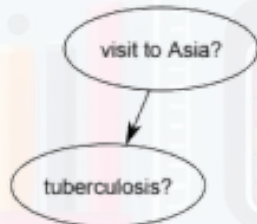
- **Propriedade de Markov das redes bayesianas:** cada variável é condicionalmente independente de suas não descendentes, dadas suas variáveis-mãe.
- **Probabilidade Condicional**
- **Fatoração da distribuição conjunta**

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{pais}(X_i))$$



visit to Asia?	yes	no
	0.01	0.99

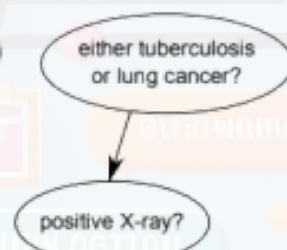
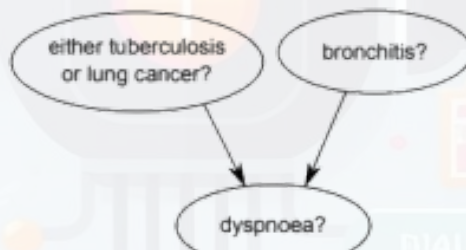
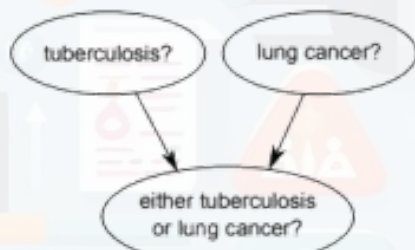
smoking?	yes	no
	0.50	0.50



	yes	no
yes	0.05	0.95
no	0.01	0.99

	yes	no
yes	0.10	0.90
no	0.01	0.99

	yes	no
yes	0.60	0.40
no	0.30	0.70



	yes	no
yes:yes	1	0
yes:no	1	0
no:yes	1	0
no:no	0	1

	yes	no
yes:yes	0.90	0.10
yes:no	0.70	0.30
no:yes	0.80	0.20
no:no	0.10	0.90

	yes	no
yes	0.98	0.02
no	0.05	0.95

# Modelagem/Aprendizado

O processo de aprendizagem é dividido em duas partes:

1. aprendizagem da estrutura (e relações entre as variáveis).
2. aprendizagem dos parâmetros numéricos (distribuição de probabilidade).

# 1. Aprendizagem da estrutura

- Consiste em identificar a estrutura do gráfico da rede bayesiana
- Existe vários algoritmos foram propostos na literatura
- Apesar dessa variedade de fundamentos teóricos e terminologias, eles podem se enquadrar em três categorias amplas:
  1. Algoritmos baseados em restrições (Constraint-based)
  2. Algoritmos baseados em pontuação (Score-based)
  3. Algoritmos híbridos

# 1.1 Algoritmos baseados em restrições

**Princípio:** realizar testes de independência condicional entre as variáveis para decidir a estrutura da rede.

Algoritmo IC (causalidade indutiva):

1. Identifica quais pares de variáveis são conectados por um arco, independentemente de sua direção, partindo de um grafo completo.
2. Identificação das v-estruturas entre todos os pares de nós não adjacentes A e B com um vizinho comum C. Essa condição pode ser verificada realizando um teste de independência condicional para A e B.
3. Identifica arcos compelidos e os orienta recursivamente para obter o DAG parcialmente completo (CPDAG).

## 1.2 Algoritmos baseados em pontuação

**Princípio:** otimizar uma pontuação que avalia o quão bem uma estrutura se ajusta aos dados.

Os exemplos mais simples são AIC e BIC, mas também há pontuações específicas para redes bayesianas como a log-verossimilhança de uma rede  $G$  penalizada por um termo que expressa a complexidade da rede.

Um dos métodos de otimização mais simples é o hill-climbing, sendo ele de otimização local que começa com um grafo inicial e em cada etapa, se escolhe a melhor operação elementar.



## 1.3 Algoritmos híbridos

**Princípio:** Combinar os testes de independência e otimização uma pontuação.

A ideia principal pode ser geralmente decomposta em 2 etapas principais:

- Redução do espaço de busca para cada nó/variável  $X$ , usando testes de independência condicional para selecionar os pais e filhos (potenciais) de  $X$ .
- Maximizar a pontuação neste espaço de busca reduzido, permitindo apenas as operações elementares compatíveis com os pais/filhos (potenciais).

Um dos principais exemplos é o algoritmo Max-Min Hill Climbing (MMHC).

## 2. Aprendizagem dos parâmetros

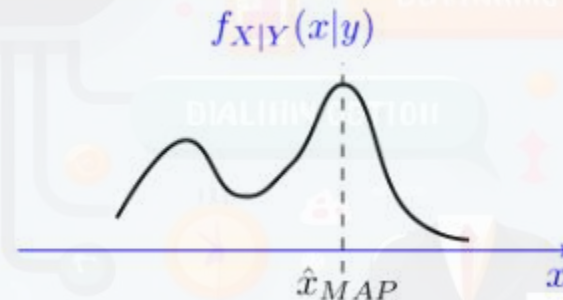
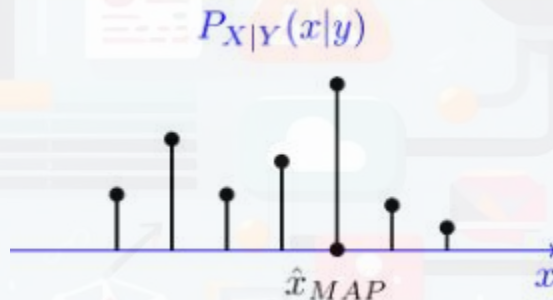
Uma vez que a estrutura da rede foi aprendida, a partir dos dados, a estimação dos parâmetros da distribuição global é bastante simplificada pela aplicação da propriedade de Markov.

Existem duas abordagens principais:

- baseada na estimativa de máxima verossimilhança
- baseada na estimativa bayesiana.

**Estimativa de Máxima Verossimilhança (EMV)** clássica, para nós discretos, é relativamente de cálculo simples: cada probabilidade pode ser estimada como o número de resultados favoráveis dividido pelo número total de resultados possíveis.

**Estimativa de Máxima a Posteriori (MAP)** da variável aleatória  $X$ , dado que observamos  $Y=y$ , é dada pelo valor de  $x$  que maximiza a distribuição a posteriori,  $f_{X|Y}(x|y)$  (ou  $P_{X|Y}(x|y)$ ), contém todo o conhecimento sobre a quantidade desconhecida  $X$



# Naive Bayes vs. Redes Bayesianas

## Naive Bayes

- Trata as covariáveis de forma **independente**;
- Eficiente para problemas simples.

## Redes Bayesianas

- Considera a **dependência** entre as covariáveis;
- Mais robusta para problemas complexos;



# Referências

- <https://medium.com/@mansi89mahi/simple-explanation-difference-between-naive-bayes-and-full-bayesian-network-model-505616545503>
- <https://medium.com/edureka/naive-bayes-in-r-37ca73f3e85c>
- <https://www.r-bloggers.com/2021/04/naive-bayes-classification-in-r/>
- <https://youtu.be/O2L2Uv9pdDA?si=MZkqTGWJESljYnmY>
- [https://youtu.be/3vY2L0ikq8w?si=LFL-f1MK3xwY3DD\\_](https://youtu.be/3vY2L0ikq8w?si=LFL-f1MK3xwY3DD_)
- <https://expleogroup.medium.com/introduction-to-bayesian-networks-and-predictive-maintenance-part-1-831d22cad158>
- <https://andreric.github.io/files/pdfs/bayesianas.pdf>
- <https://dipartimenti.unicatt.it/scienze-statistiche-23-25-1-17ScutariSlides.pdf>
- [https://www.probabilitycourse.com/chapter9/9\\_1\\_2\\_MAP\\_estimation.php](https://www.probabilitycourse.com/chapter9/9_1_2_MAP_estimation.php)
- <https://medium.com/@mansi89mahi/simple-explanation-difference-between-naive-bayes-and-full-bayesian-network-model-505616545503>



# Obrigado!