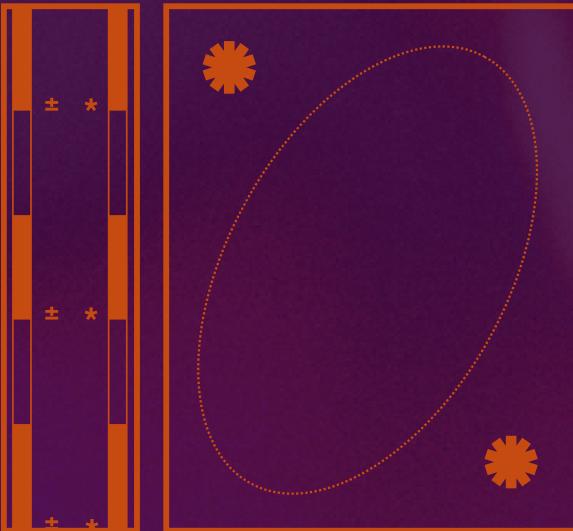


# APRENDIZADO DE MÁQUINA



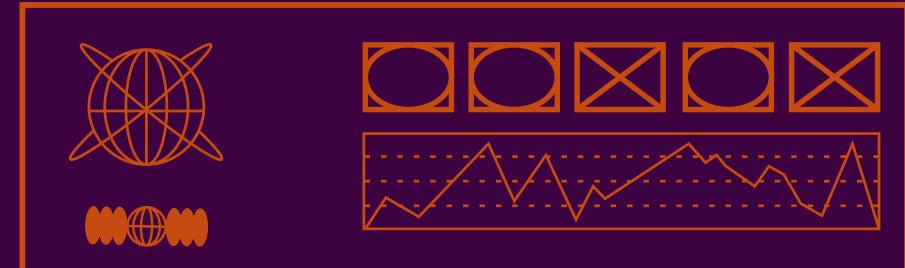
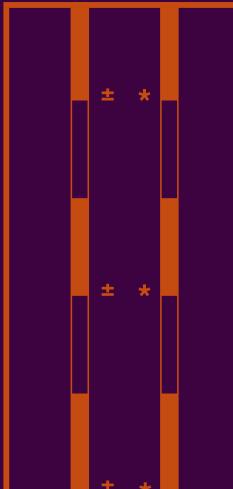
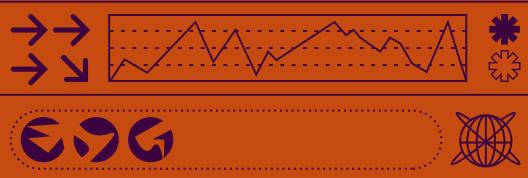
## DIA 2



# TÓPICOS



## Tópicos explorados na Aula:



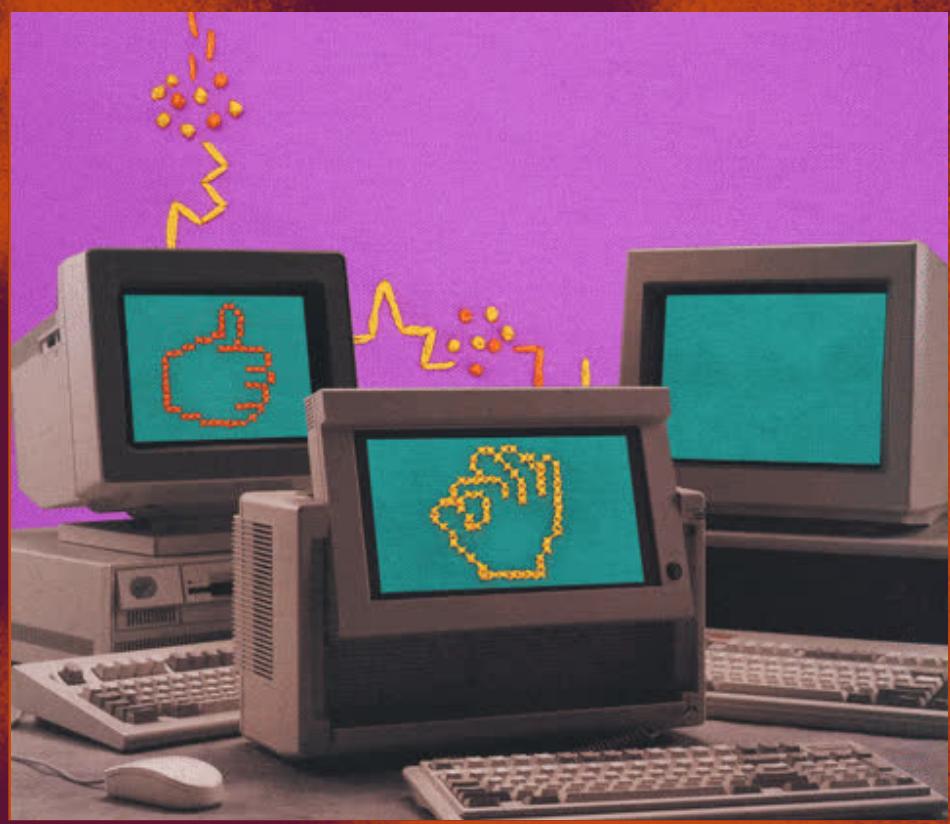
- » Classificação de Modelos
- » Modelos de Classificação
- » K-Nearest Neighbors
- » Naives Bayes
- » Regressão Linear
- » Regressão Logística
- » Regularização

AGENDA AGENDA AGENDA AGENDA AGENDA AGENDA



# Classificação de Modelos de ML

- Quais são as características que podemos classificar em um modelo?



# **MODELOS DE ML PODEM SER CLASSIFICADOS DE ACORDO COM:**

**TIPOS DE APRENDIZADO**

**SAÍDAS DO MODELO**

**ENTRADAS DO MODELO**

**TIPOS DE MODELO**

# TIPOS DE APRENDIZADO:



# SAÍDAS DO MODELO:

MACHINE LEARNING

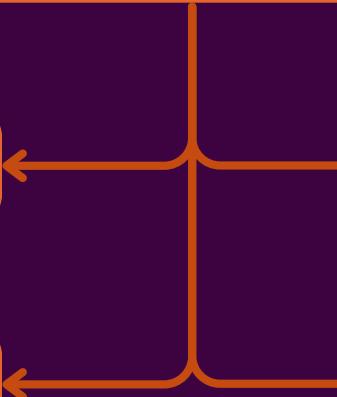
(APRENDIZADO DE MÁQUINA)

CLASSIFICAÇÃO

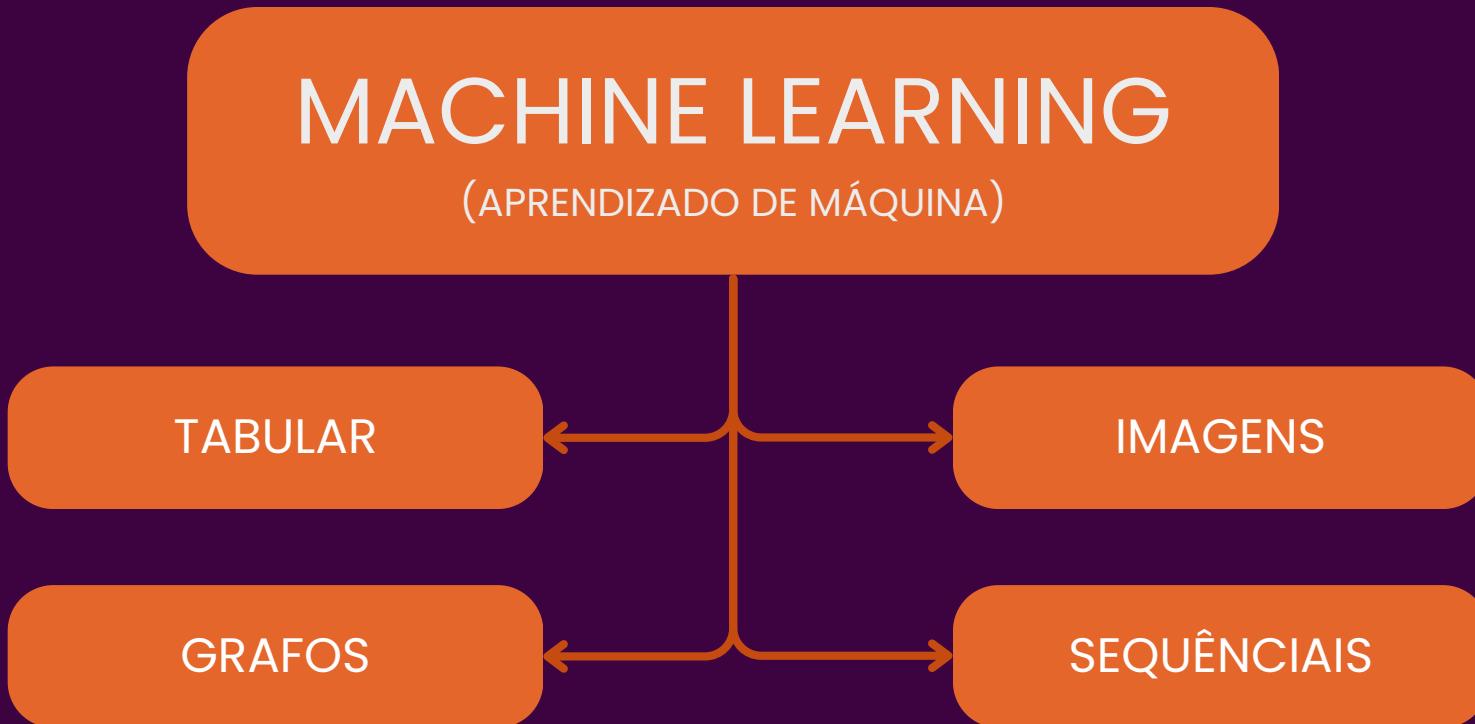
REGRESSÃO

GERAÇÃO

CLUSTERIZAÇÃO



# ENTRADAS DO MODELO:



# TIPOS DE MODELO:

## MACHINE LEARNING

(APRENDIZADO DE MÁQUINA)

LINEARES

BASEADOS EM VIZINHOS

ÁRVORE

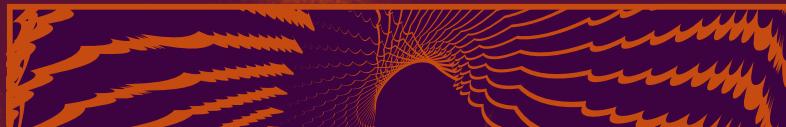
PROBABILÍSTICOS

REDES NEURAIS



# Modelos de Classificação

- Quais algoritmos possuímos para esse tipo de problema?



KNN

Árvore de  
Decisão

Random  
Forest

Naive Bayes

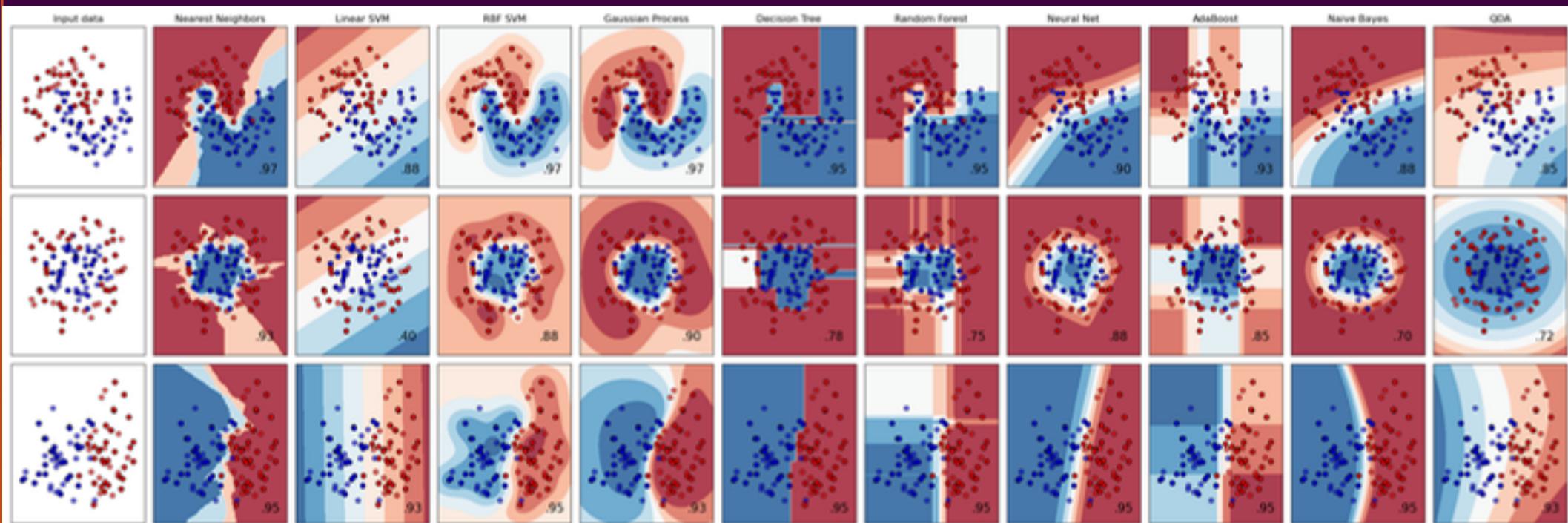
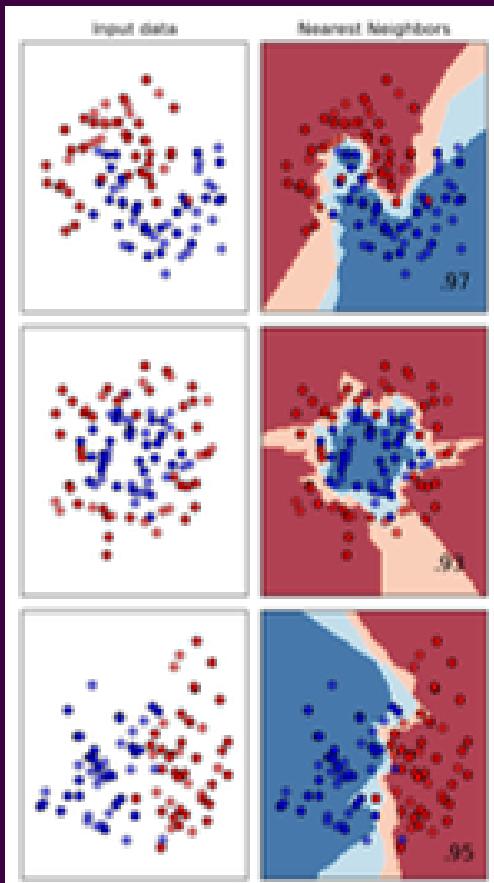
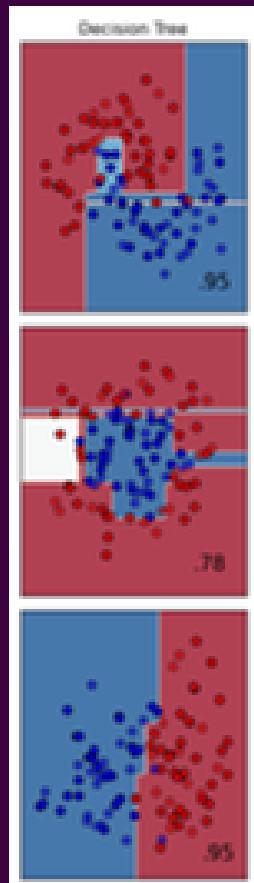


Imagen: [Scikit-Learn](#)

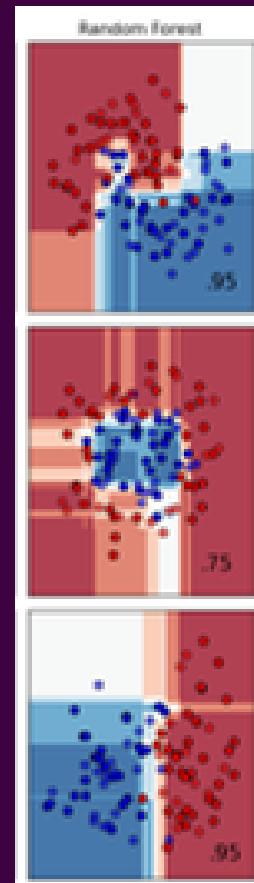
KNN



Árvore de  
Decisão



Random  
Forest



Naive Bayes

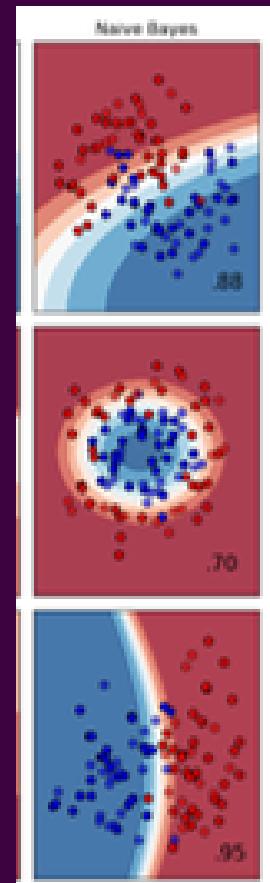


Imagen: [Scikit-Learn](#)



EDG



# K-Nearest Neighbors (KNN)

- Como funciona o algoritmo baseado nos vizinhos mais próximos?



# KNN: PASSO A PASSO

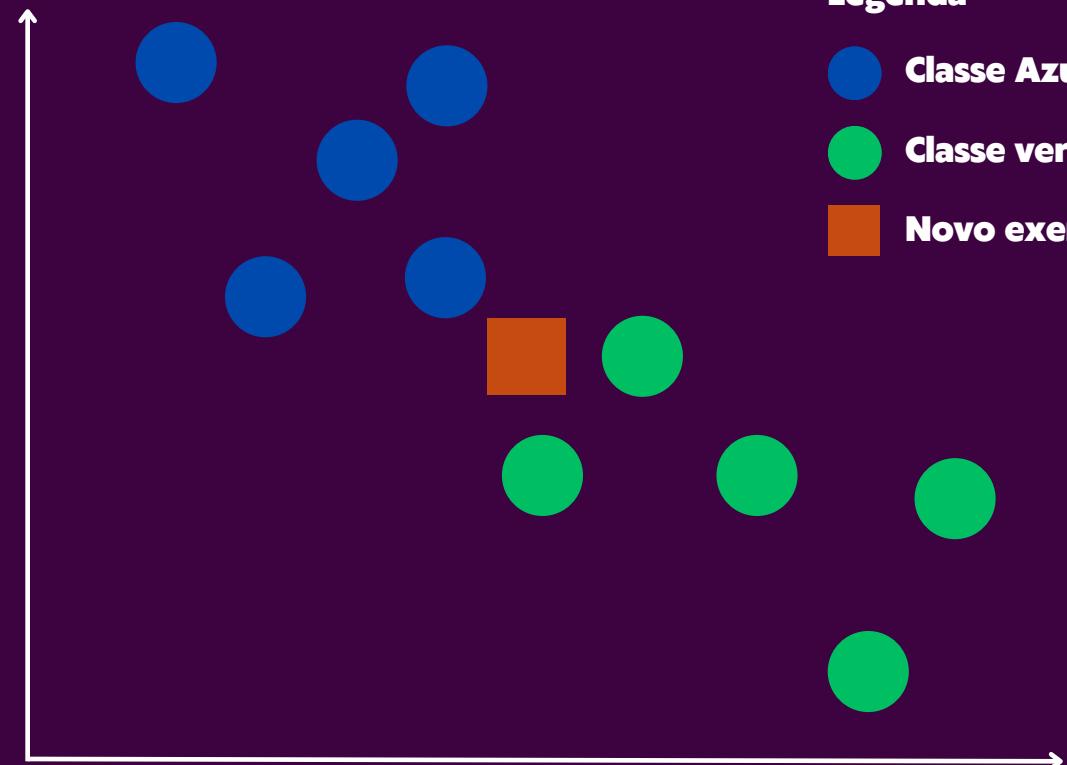
**1. Calcular a distância do novo ponto para todos os pontos de treino**

**2. Ordenar essas distâncias**

**3. Selecionar os k vizinhos mais próximos**

**4. Votação (classificação) ou média (regressão)**

**5. Atribuir o resultado ao novo ponto**



**Legenda**

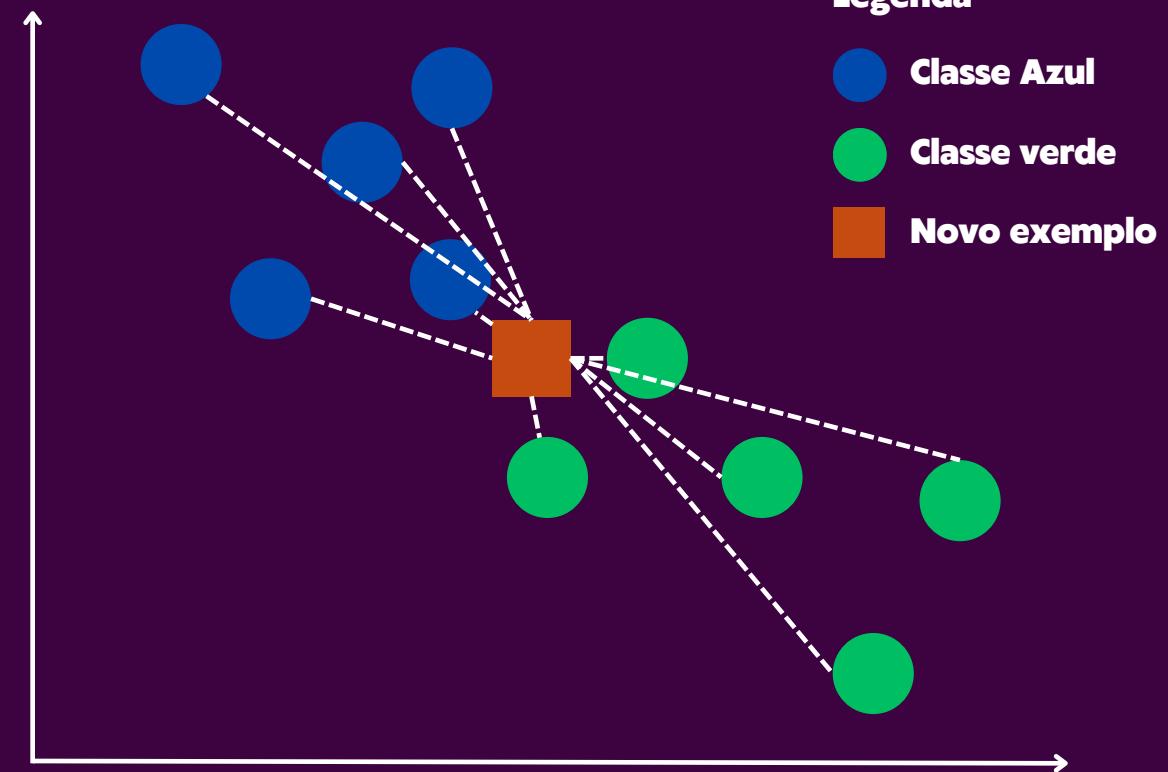
● Classe Azul

● Classe verde

■ Novo exemplo

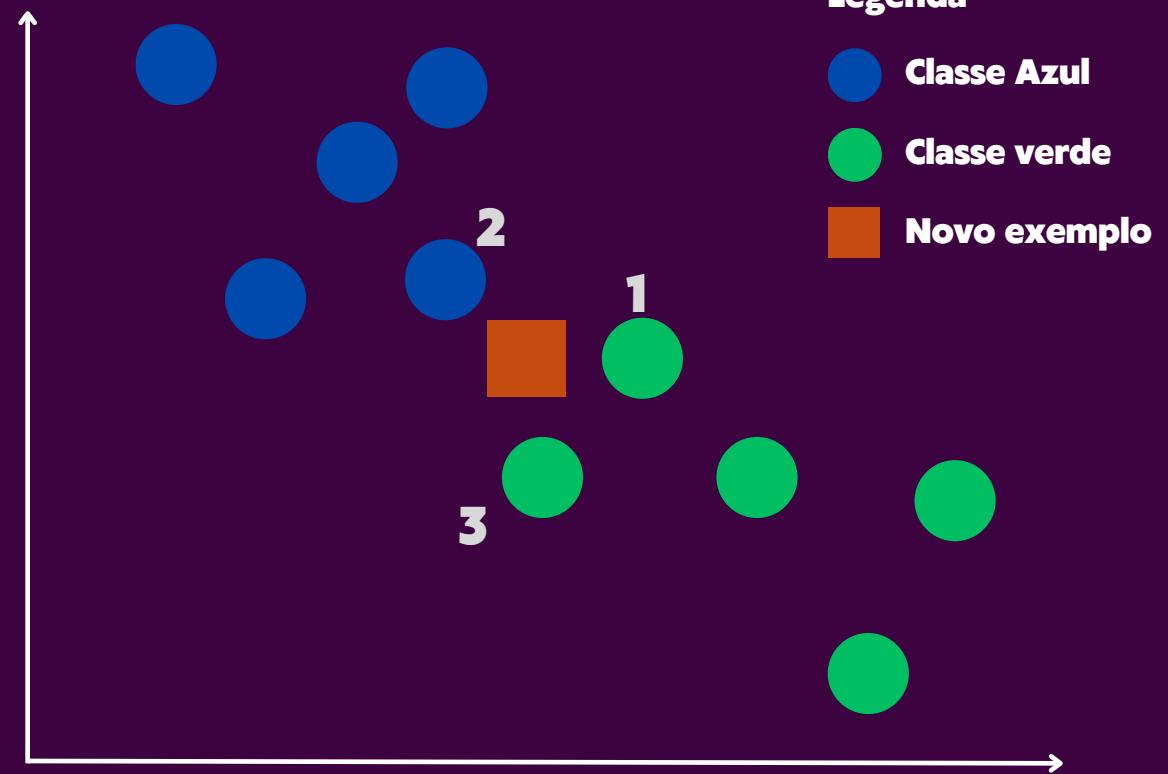
# KNN: PASSO A PASSO

**1. Calcular a distância do novo ponto para todos os pontos de treino**



# KNN: PASSO A PASSO

## 2. Ordenar essas distâncias

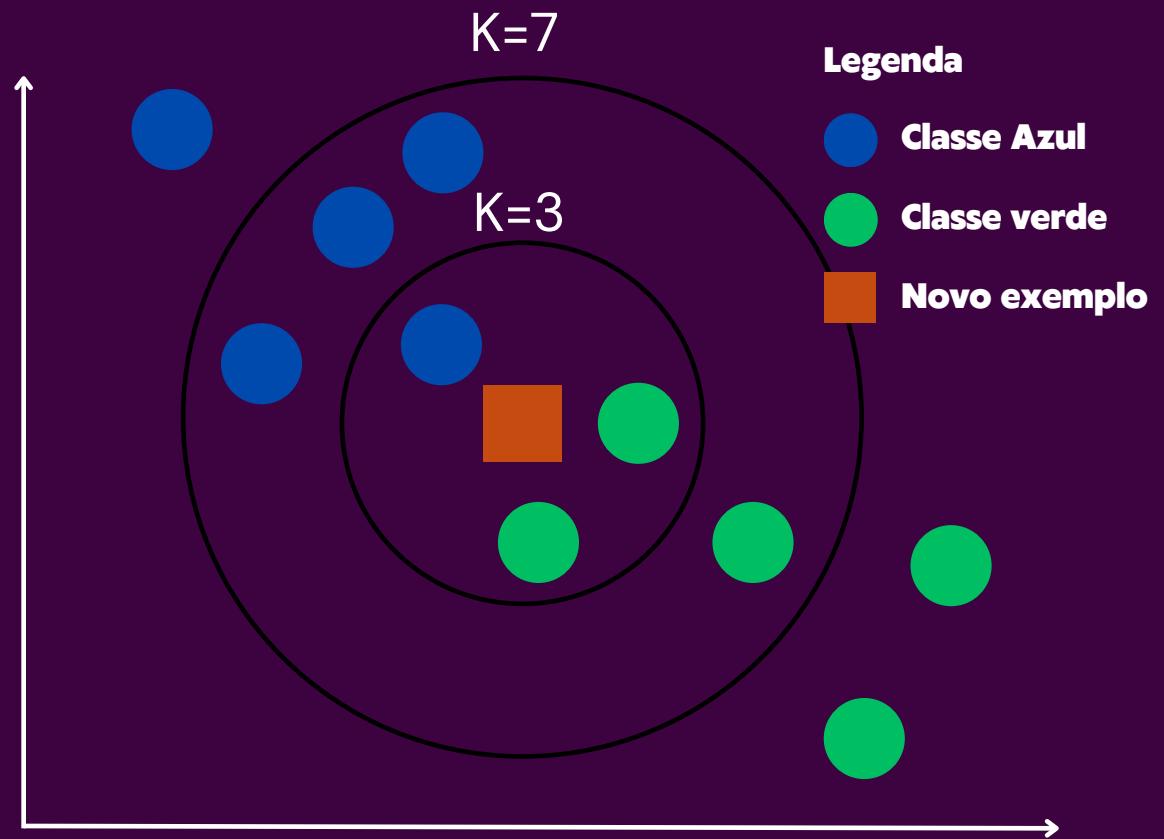


# KNN: PASSO A PASSO

**3. Selecionar os k vizinhos mais próximos**

**Exemplos:**

- **Selecionarmos  $k = 3$**
- **Selecionarmos  $k = 7$**

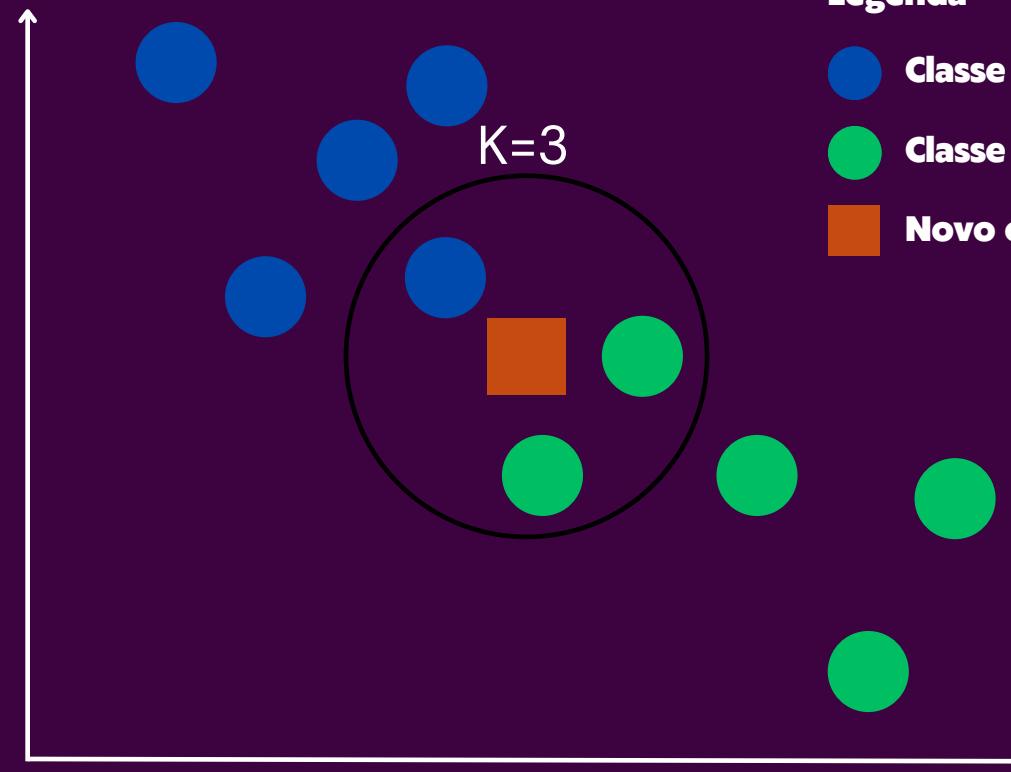


# KNN: PASSO A PASSO

## 4. Votação (classificação) ou média (regressão)

**Exemplo:**

- Selecionamos  $k = 3$ , então o novo exemplo será da classe verde!

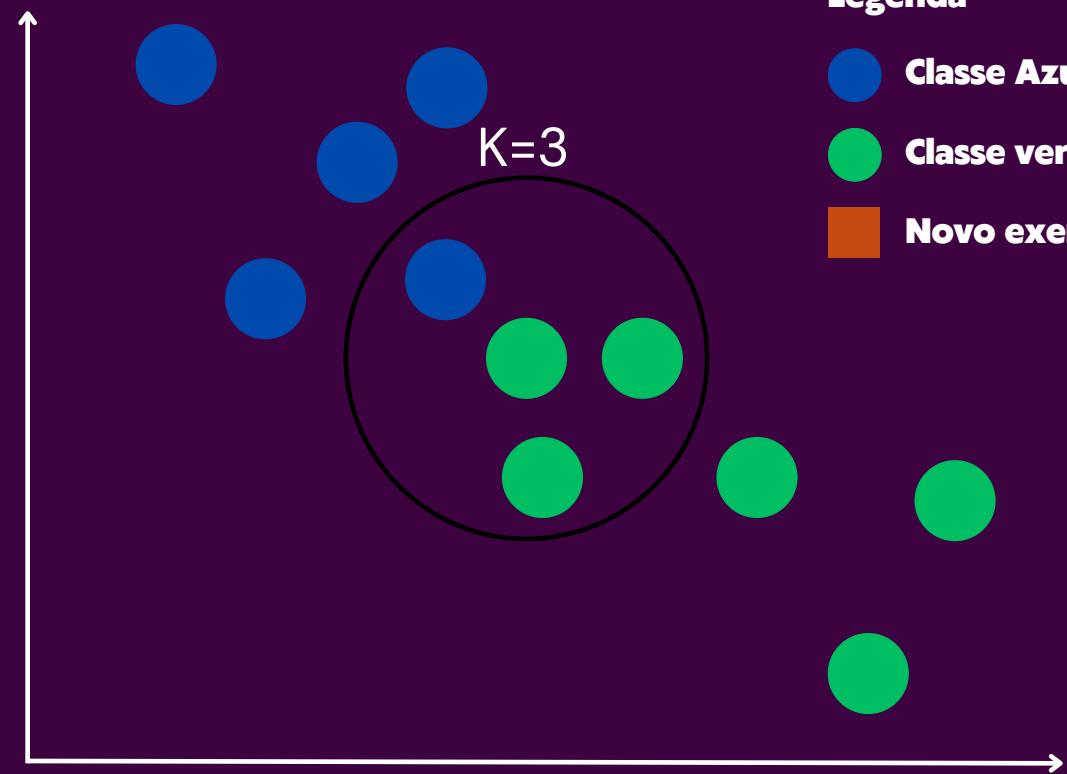


# KNN: PASSO A PASSO

## 5. Atribuir o resultado ao novo ponto

**Exemplo:**

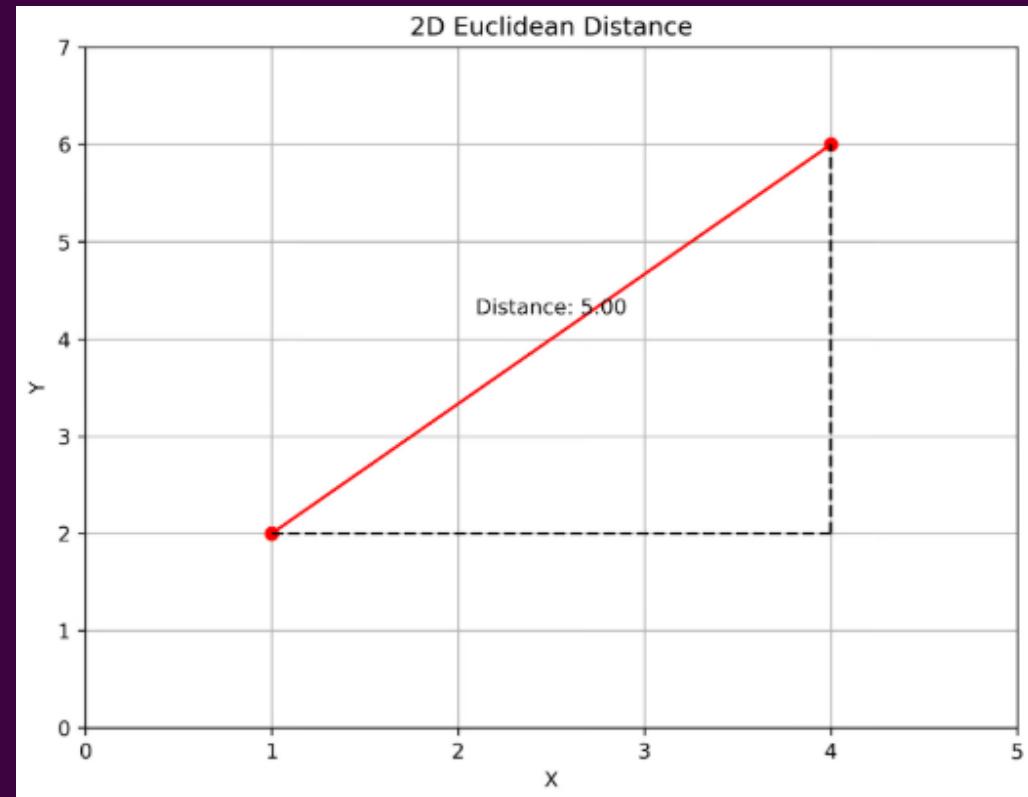
- Selecionamos  $k = 3$ , então o novo exemplo será da classe verde!



# KNN: DISTÂNCIAS MAIS COMUNS

Distância euclidiana 2D

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$



# KNN: DISTÂNCIAS MAIS COMUNS

Distância euclidiana 2D

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

CATETOS



# KNN: DISTÂNCIAS MAIS COMUNS

Distância euclidiana 2D

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

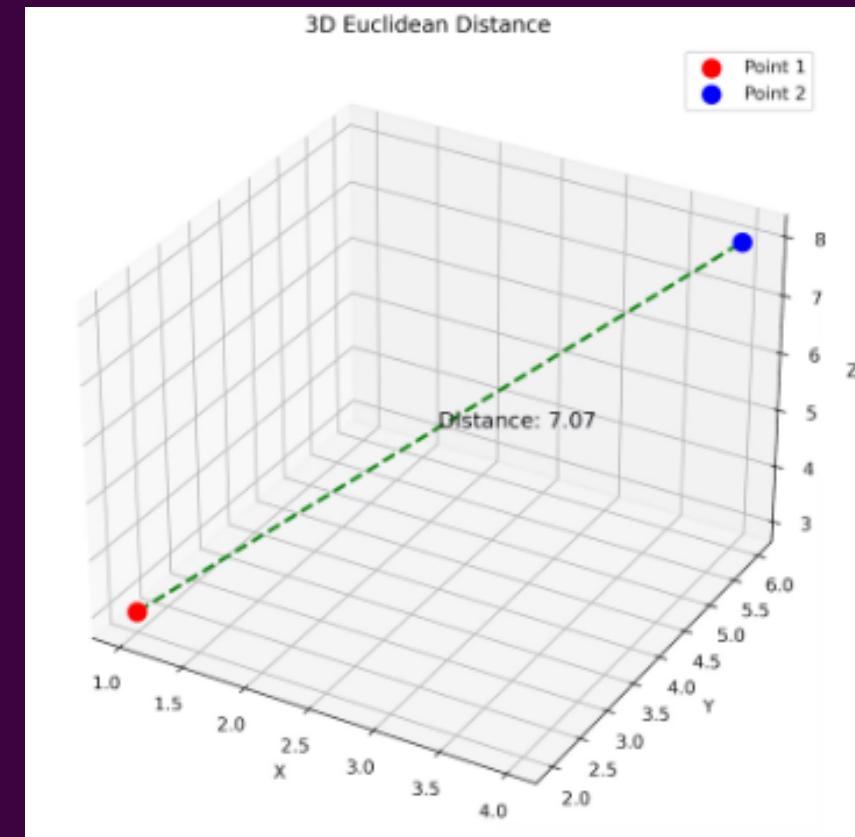
HIPOTENUSA



# KNN: DISTÂNCIAS MAIS COMUNS

Distância euclidiana 3D

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$



# KNN: DISTÂNCIAS MAIS COMUNS

Distância Euclidiana  
N-dimensional

$$d = \sqrt{\sum_{i=1}^n (b_i - a_i)^2}$$

# KNN: CARACTERÍSTICAS

Definição

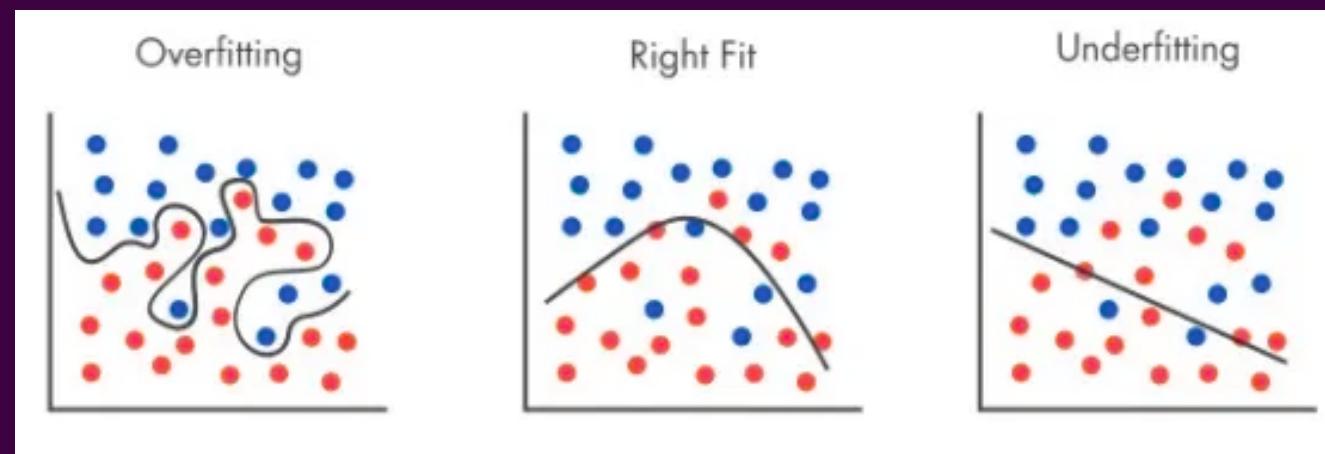
- **Algoritmo supervisionado, usado para classificação e regressão.**
- **Baseia-se na ideia de que "coisas parecidas estão próximas".**

Características

- **Não envolve fase de treinamento real: apenas armazena os dados.**
- **O "aprendizado" acontece na hora da previsão (lazy learner).**
  - **Toda vez que é chamado, aprende as distâncias**

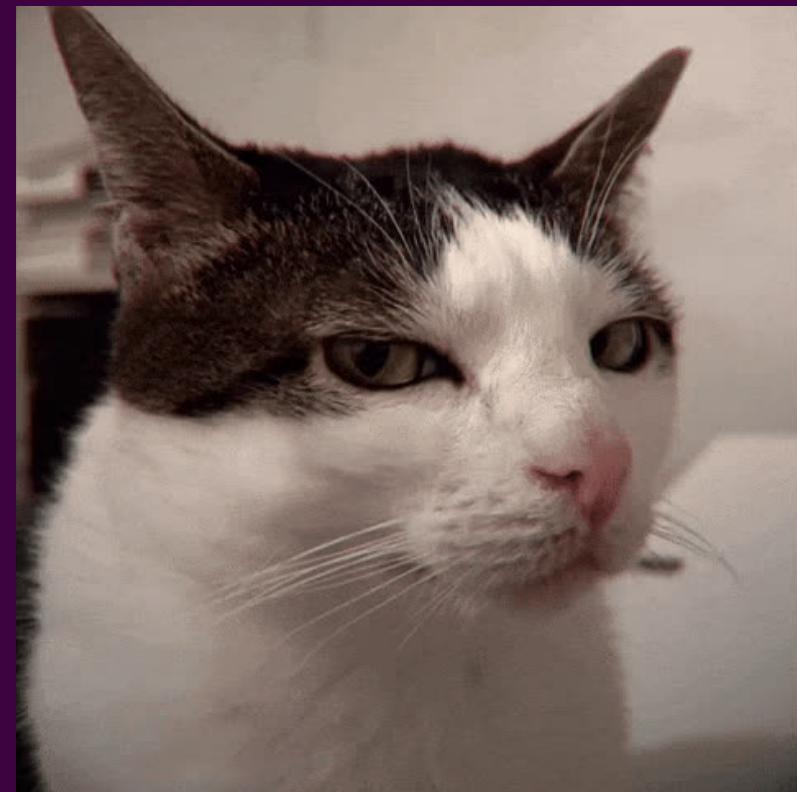
# KNN: COMO ESCOLHER O VALOR DE K?

- Se K é muito pequeno → sensível a ruído (overfitting).
- Se K é muito grande → suaviza demais (underfitting).



# KNN: COMO ESCOLHER O VALOR DE K?

VOCÊS:



# KNN: COMO ESCOLHER O VALOR DE K?

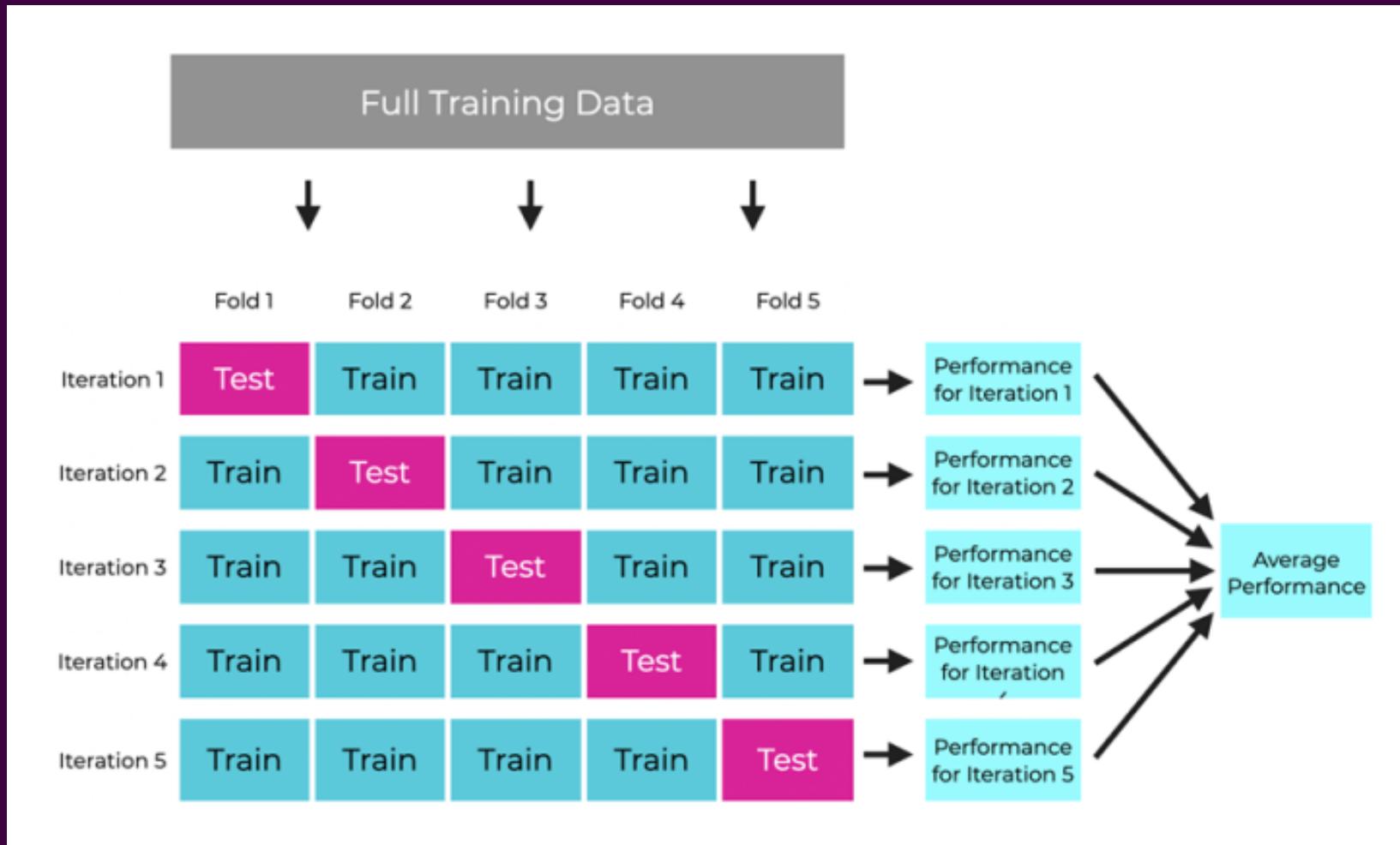
## VALIDAÇÃO CRUZADA!



# KNN: COMO ESCOLHER O VALOR DE K?

## Validação cruzada - k-fold

1. **Divida os dados em k partes (ou "folds") aproximadamente iguais.**
2. **Para cada uma das k rodadas:**
  - **Use k-1 partes para treinar o modelo.**
  - **Use a parte que sobrou para testar.**
3. **Repita isso k vezes, trocando a parte de teste em cada rodada.**
4. **No final, calcule a média das métricas (ex: acurácia, erro quadrático, etc.) obtidas em cada rodada.**



# DOCUMENTAÇÃO



[Documentação KNN](#)

# MÃO NA MASSA!



# VAMOS CODAR!

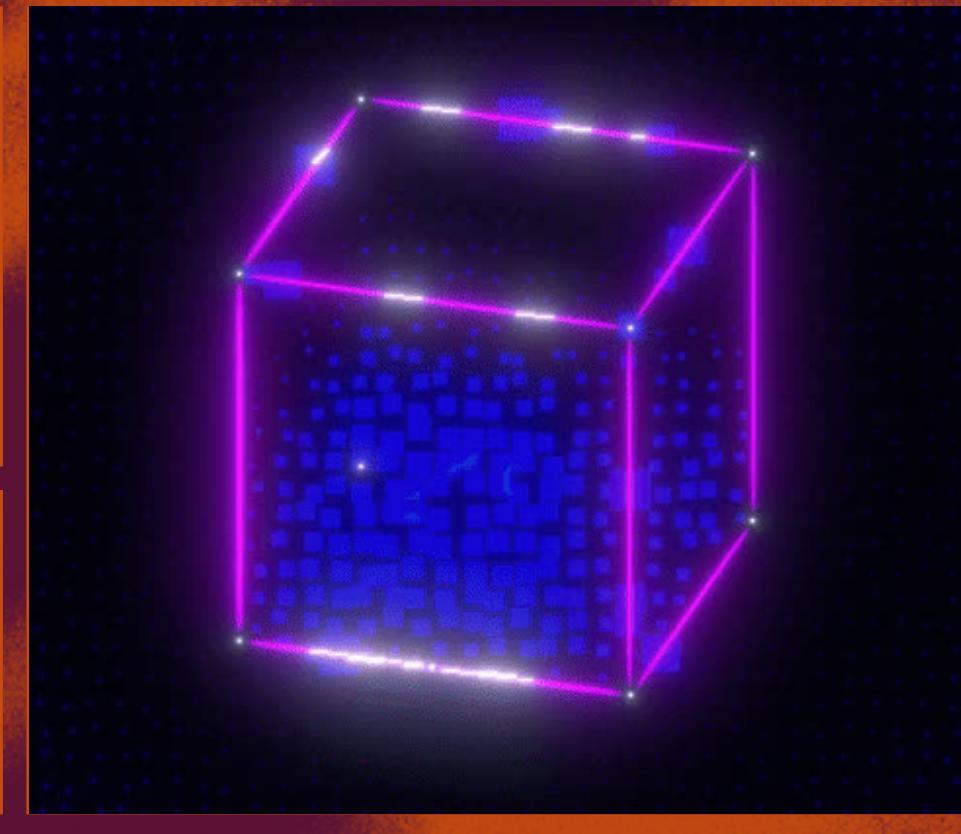
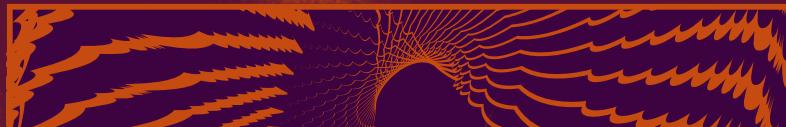


EDG

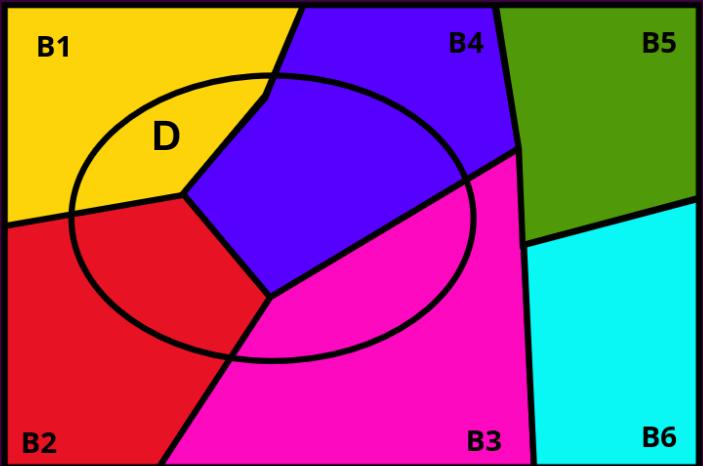


# Naive Bayes

- O que é o Teorema de Bayes e como ele se aplica em Machine Learning?



# TEOREMA DE BAYES

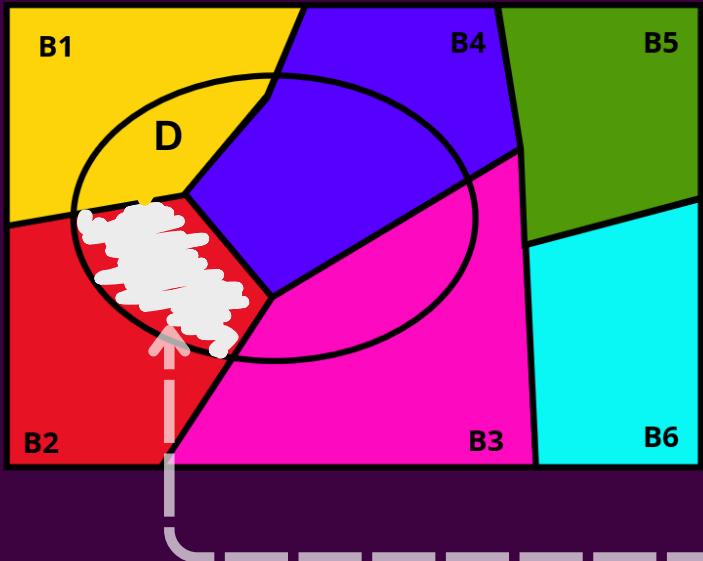


**Exemplo:**

- **D: Define se a pessoa está doente**
- **D': Define se a pessoa não está doente**
- **B1: Define Sintoma 1**
- **B2: Define Sintoma 2 ...**

**Qual a probabilidade de uma pessoa estar doente dado que ela está sentindo o sintoma 2?  $\rightarrow P(D|B2)$**

# TEOREMA DE BAYES



**Temos uma hipótese (pessoa estar doente) gerada a partir de uma evidência (pessoa possuir sintoma 2)**

**Do efeito para a causa → Diagnóstico**

**Qual a probabilidade de uma pessoa estar doente dado que ela está sentindo o sintoma 2? →  $P(D|B2)$**

# TEOREMA DE BAYES

## LIKELIHOOD

The probability of "B" being True, given "A" is True

## PRIOR

The probability "A" being True. This is the knowledge.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

## POSTERIOR

The probability of "A" being True, given "B" is True

## MARGINALIZATION

The probability "B" being True.

$$P(D|B_2) = \frac{P(B_2|D) * P(D)}{P(B_2)}$$

Imagens: [thebottomline](#) e [towardsdatascience](#)

# NAIVE BAYES

O que significa  
“Naive”?

**“Naive” em Inglês significa inocente**



**Logo, para esse modelo de Machine Learning de Classificação, ele considera que todos os atributos são independentes, o que é uma “suposição forte”.**

→ **Algumas doenças possuem sintomas muito correlacionados, isto é, que possuem algum tipo de ligação entre si quando mudam**

# NAIVE BAYES

## Detection of Cardiovascular Disease Risk's Level for Adults Using Naive Bayes Classifier

[Eka Miranda<sup>1</sup>](#), [Edy Irwansyah<sup>1,✉</sup>](#), [Alowisius Y Amelga<sup>1</sup>](#), [Marco M Maribondang<sup>1</sup>](#), [Mulyadi Salim<sup>1</sup>](#)

► Author information ► Article notes ► Copyright and License information

PMCID: PMC4981580 PMID: [27525161](#)

[LINK PESQUISA](#)

# NAIVE BAYES

O que significa  
“Naive”?

“Naive” em Inglês significa inocente

Logo, para esse modelo de Machine Learning de Classificação, ele considera que todos os atributos são independentes, o que é uma “suposição forte”.

$$P(D|B1, B2, B3 \dots) = \frac{P(B1, B2, B3 \dots | D) * P(D)}{P(B1, B2, B3 \dots)}$$

# NAIVE BAYES

O que significa  
“Naive”?

“Naive” em Inglês significa inocente

Logo, para esse modelo de Machine Learning de Classificação, ele considera que todos os atributos são independentes, o que é uma “suposição forte”.

$$\rightarrow P(D|B1, B2, B3 \dots) = P(D) * P(B1|D) * P(B2|D) * P(B3|D) \dots$$