

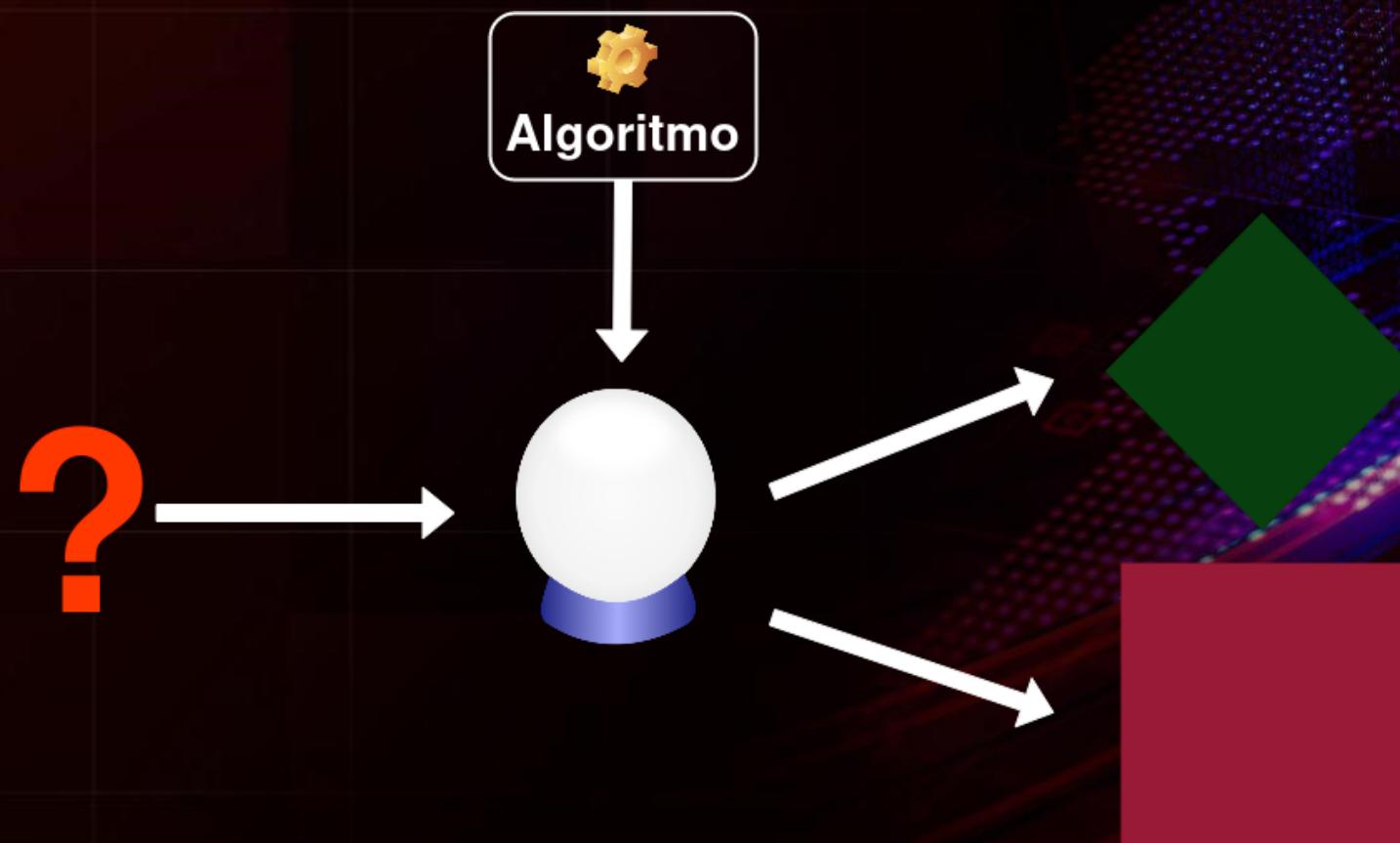
MINERAÇÃO DE DADOS

Algoritmos de classificação



ALGORITMOS DE CLASSIFICAÇÃO

Geram o modelo de classificação (classificador)



TIPOS DE ALGORITMOS DE CLASSIFICAÇÃO

Baseados em
conhecimento

Baseados em
árvores

Conexionistas

Baseados em
distância

Baseados em
função

Probabilísticos



TIPOS DE ALGORITMOS DE CLASSIFICAÇÃO

Baseados em
conhecimento

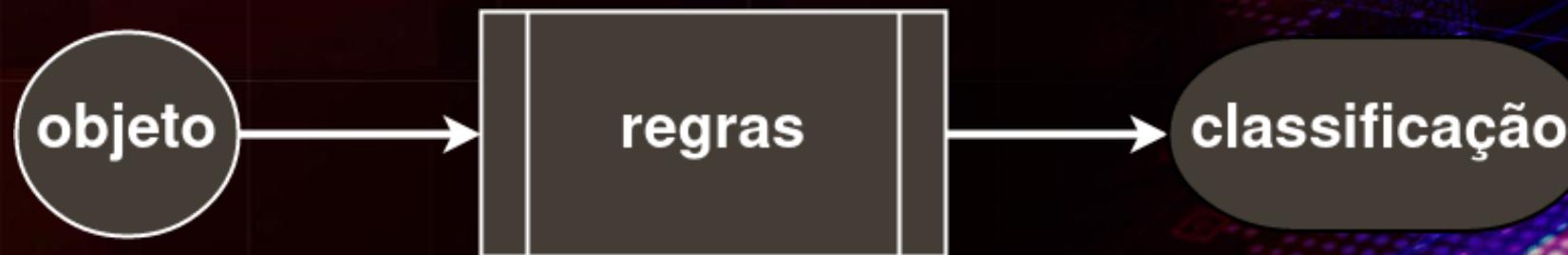
Baseados em
árvores

Conexionistas

Baseados em
distância

Baseados em
função

Probabilísticos



if..then..else, switch..case

TIPOS DE ALGORITMOS DE CLASSIFICAÇÃO

Baseados em
conhecimento

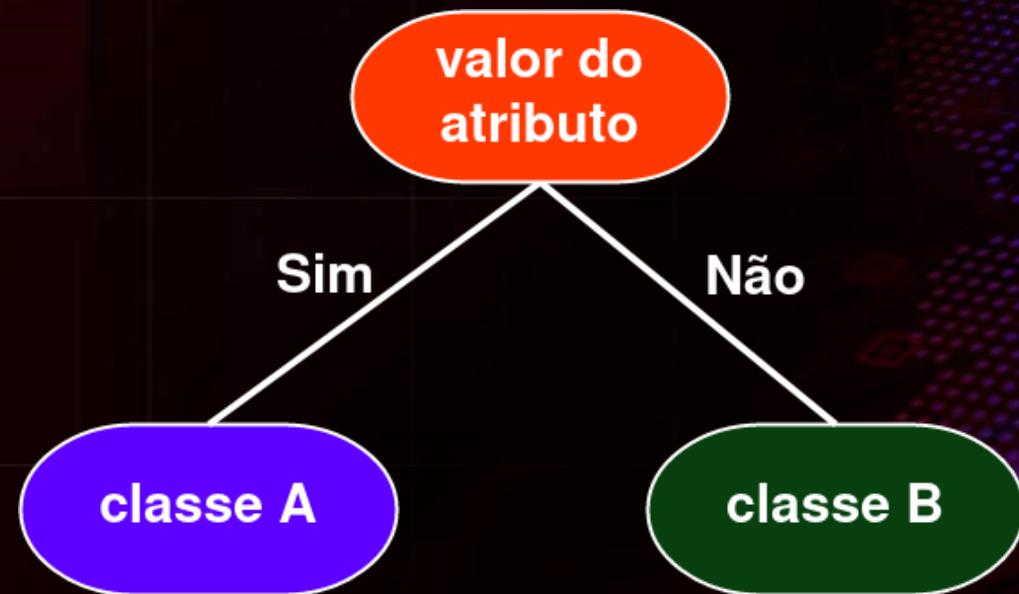
Baseados em
árvores

Conexionistas

Baseados em
distância

Baseados em
função

Probabilísticos



árvores de decisão

TIPOS DE ALGORITMOS DE CLASSIFICAÇÃO

Baseados em
conhecimento

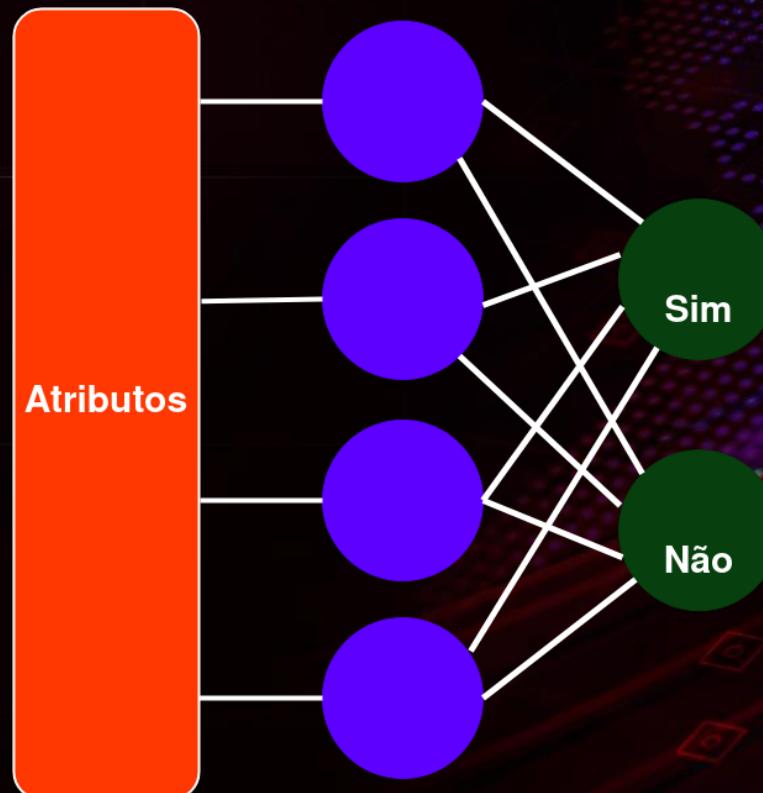
Baseados em
árvores

Conexionistas

Baseados em
distância

Baseados em
função

Probabilísticos



redes neurais
artificiais

TIPOS DE ALGORITMOS DE CLASSIFICAÇÃO

Baseados em
conhecimento

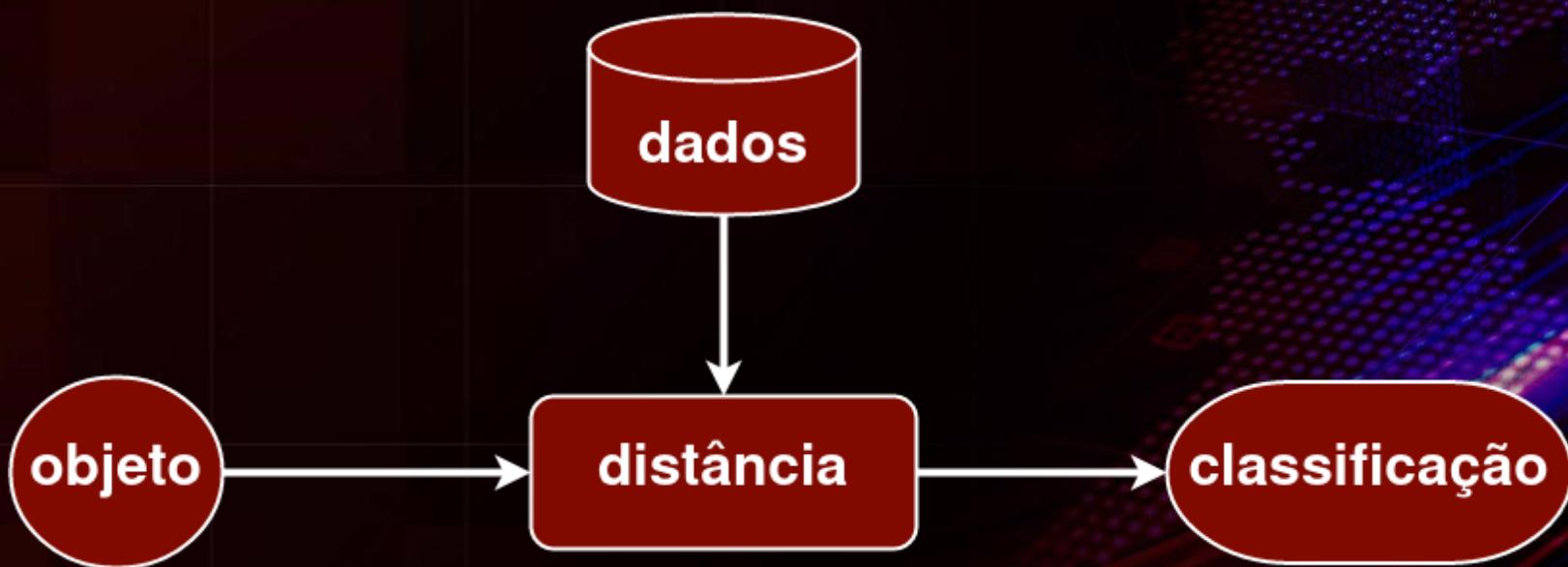
Baseados em
árvores

Conexionistas

Baseados em
distância

Baseados em
função

Probabilísticos



k-NN

TIPOS DE ALGORITMOS DE CLASSIFICAÇÃO

Baseados em
conhecimento

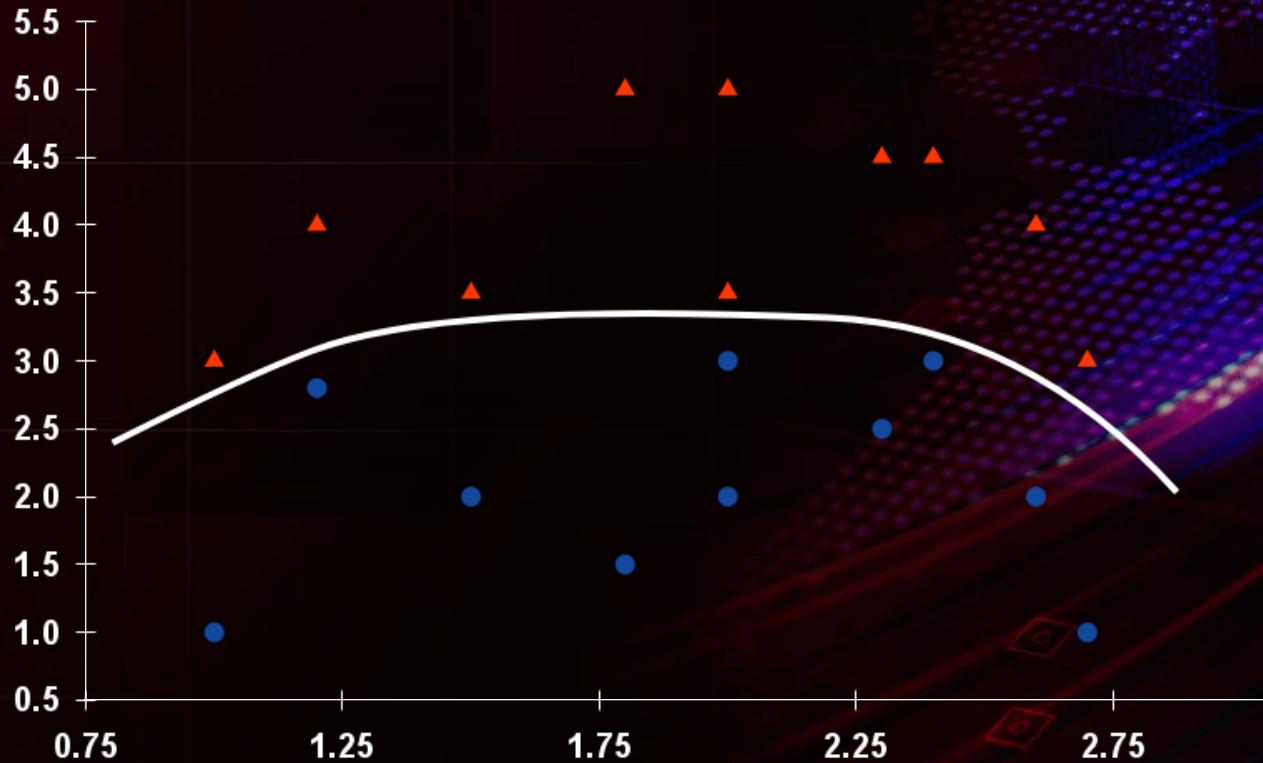
Baseados em
árvores

Conexionistas

Baseados em
distância

Baseados em
função

Probabilísticos



SVM

TIPOS DE ALGORITMOS DE CLASSIFICAÇÃO

Baseados em
conhecimento

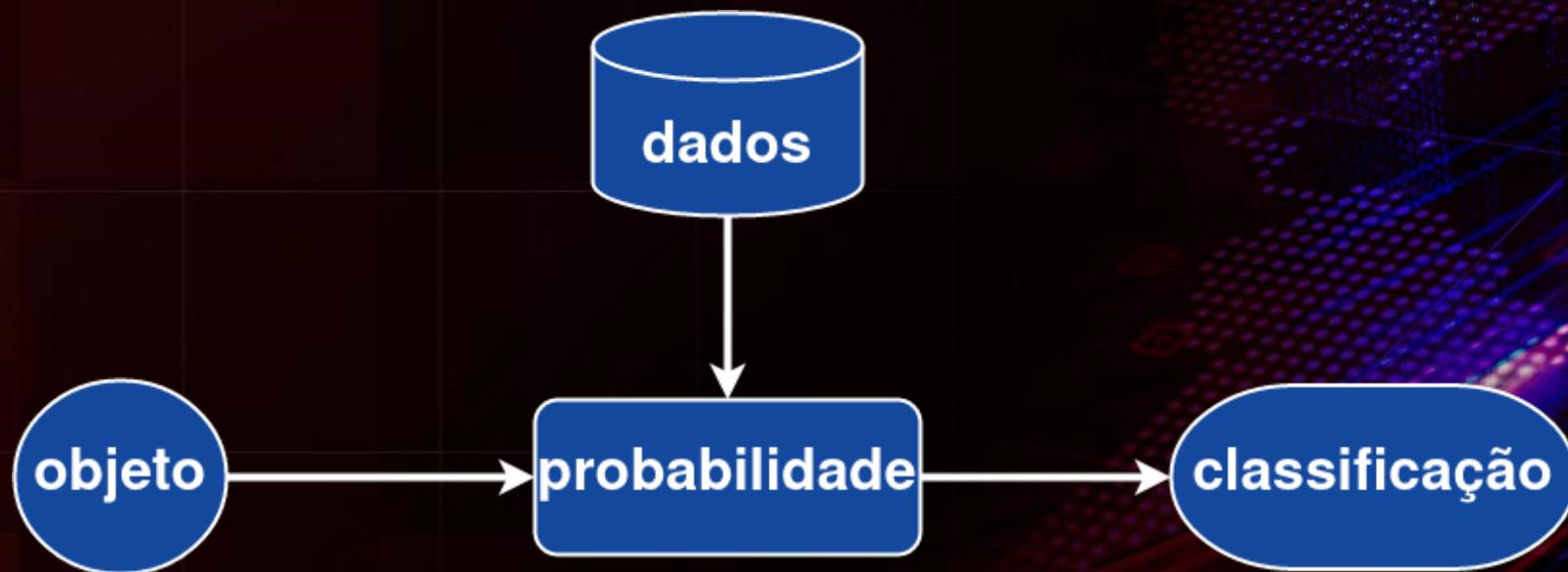
Baseados em
árvores

Conexionistas

Baseados em
distância

Baseados em
função

Probabilísticos

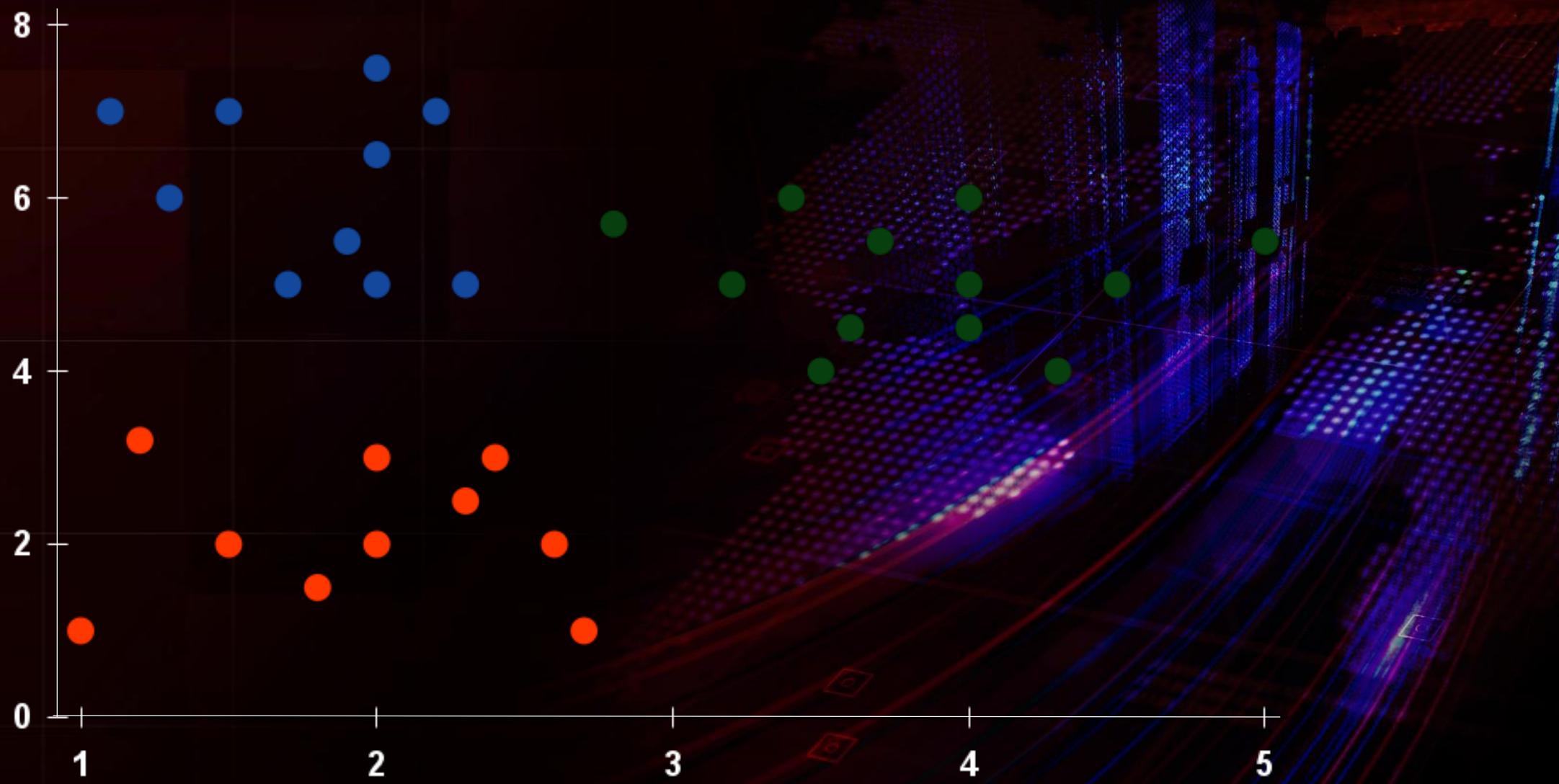


naïve bayes

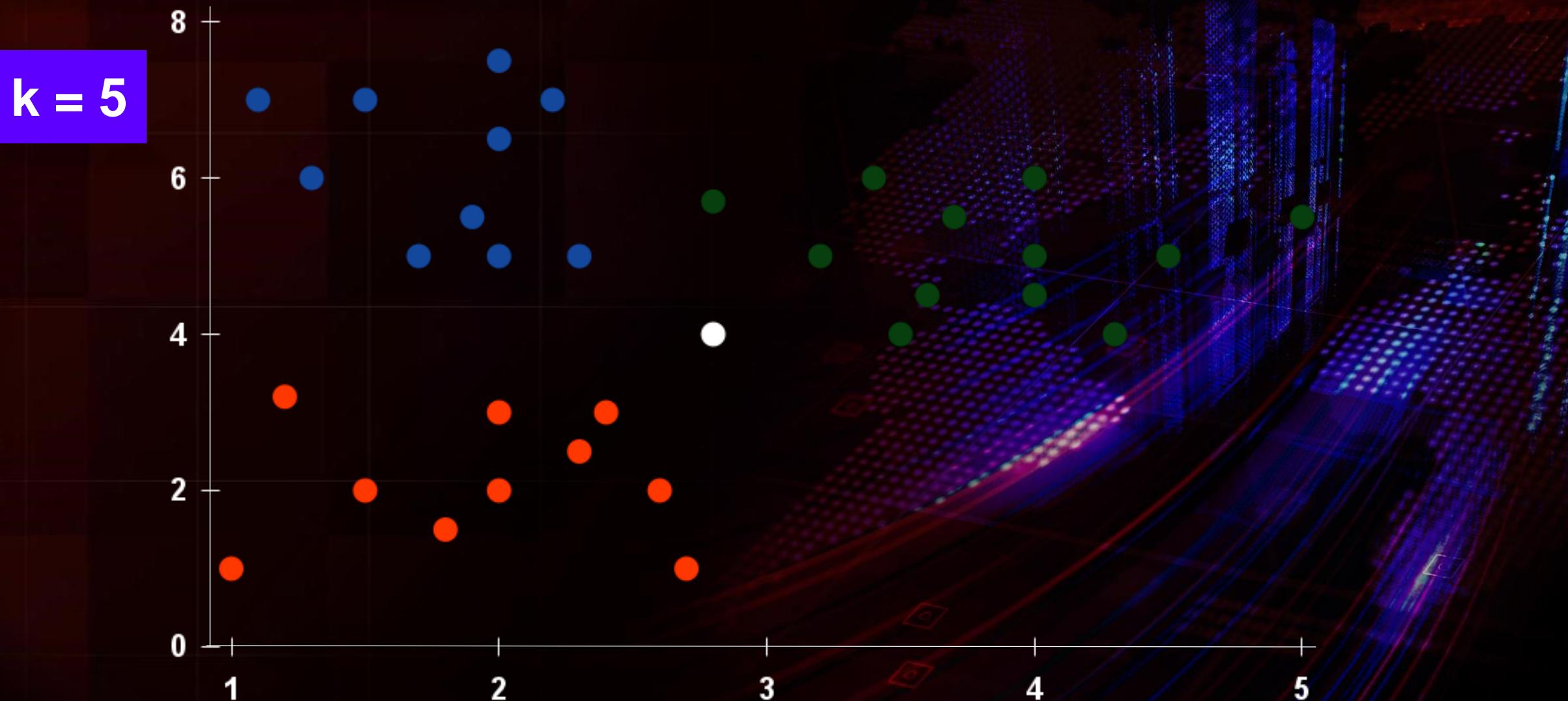
k-NN (k-VIZINHOS MAIS PRÓXIMOS)

- Baseado na distância dos k vizinhos
- Objeto vai pertencer à classe da maioria dos vizinhos
- Usa medidas de distância / similaridade
 - Ex: Euclidiana
- Algoritmo preguiçoso
- Definição de k:
 - Heurística / tentativa e erro

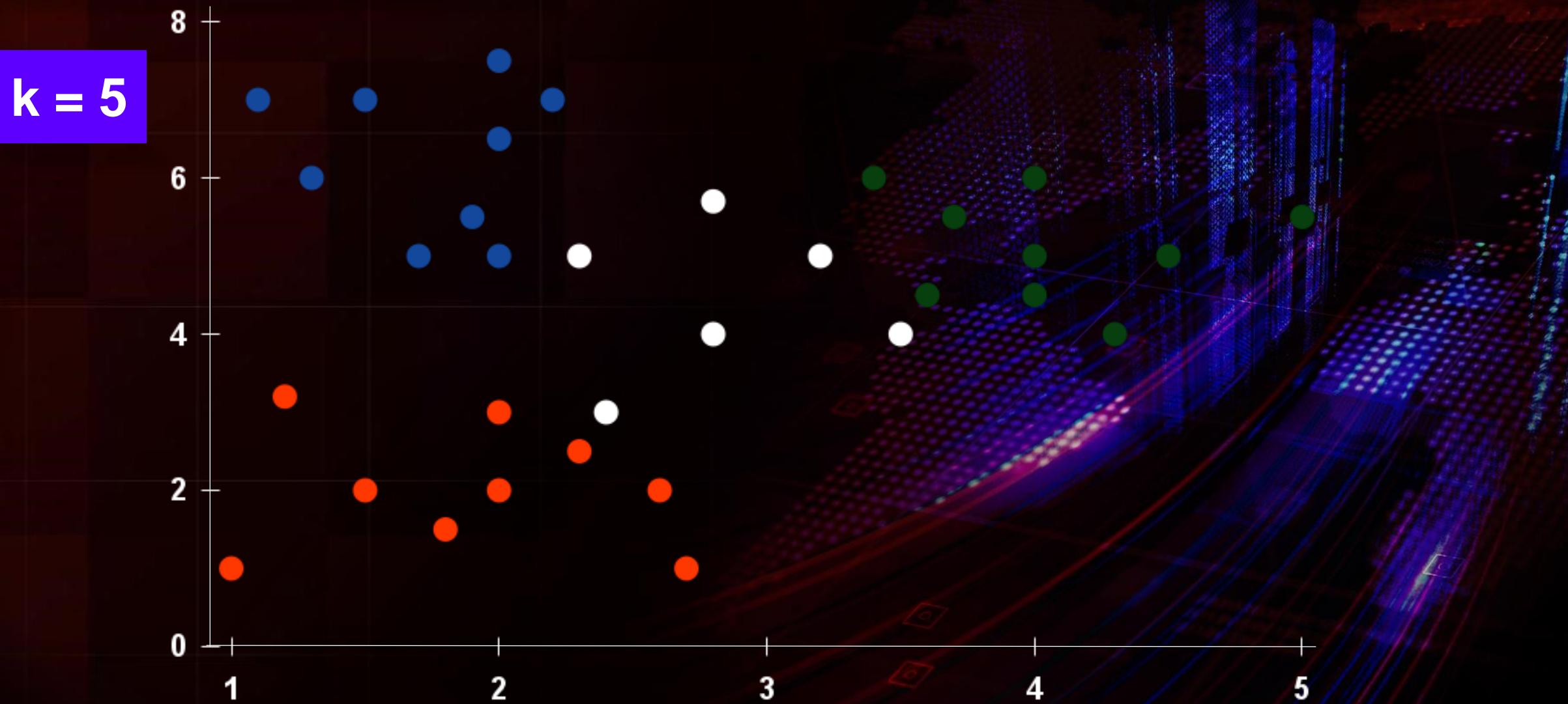
k-NN - EXEMPLO



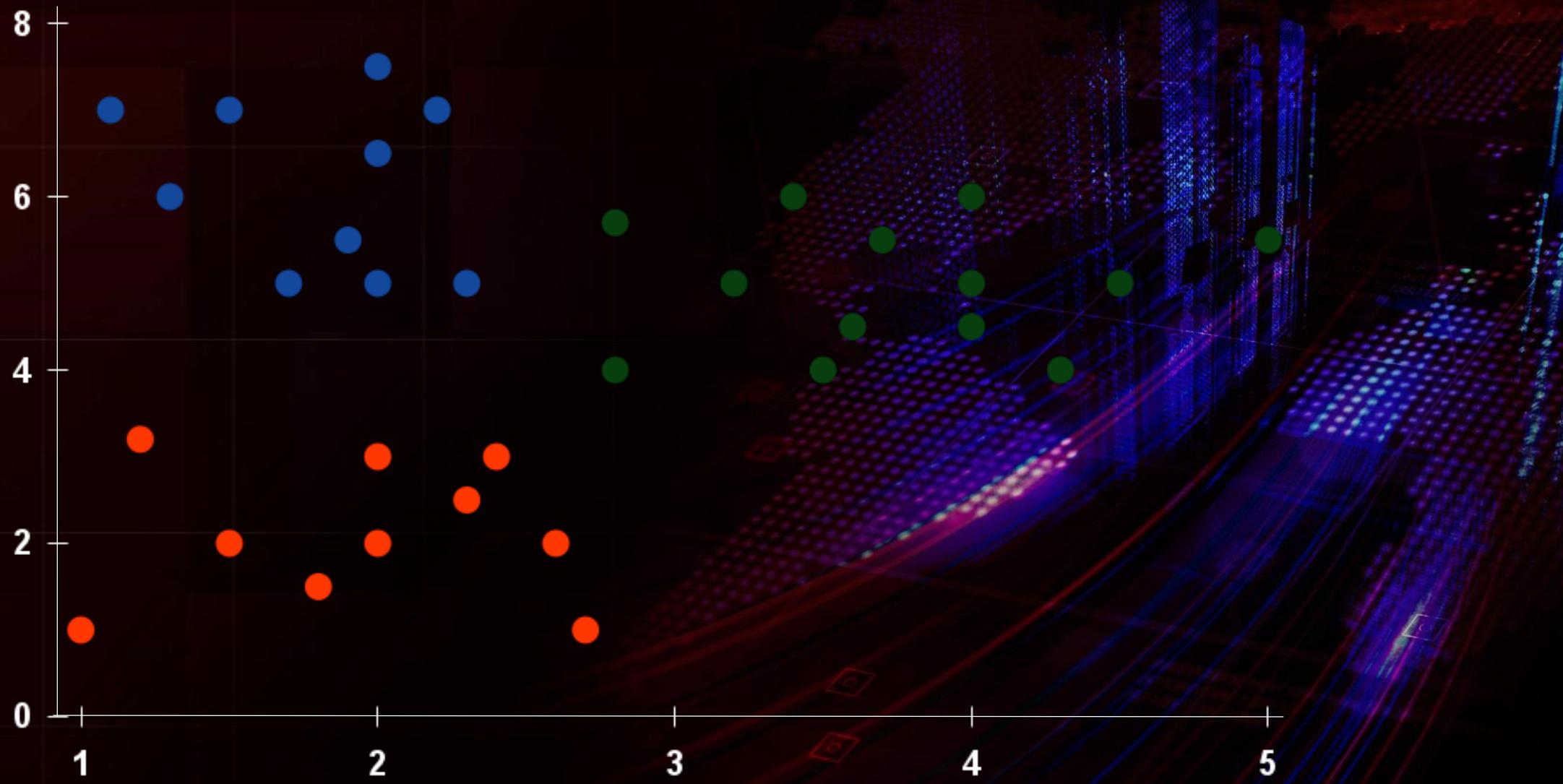
k-NN - EXEMPLO



k-NN - EXEMPLO

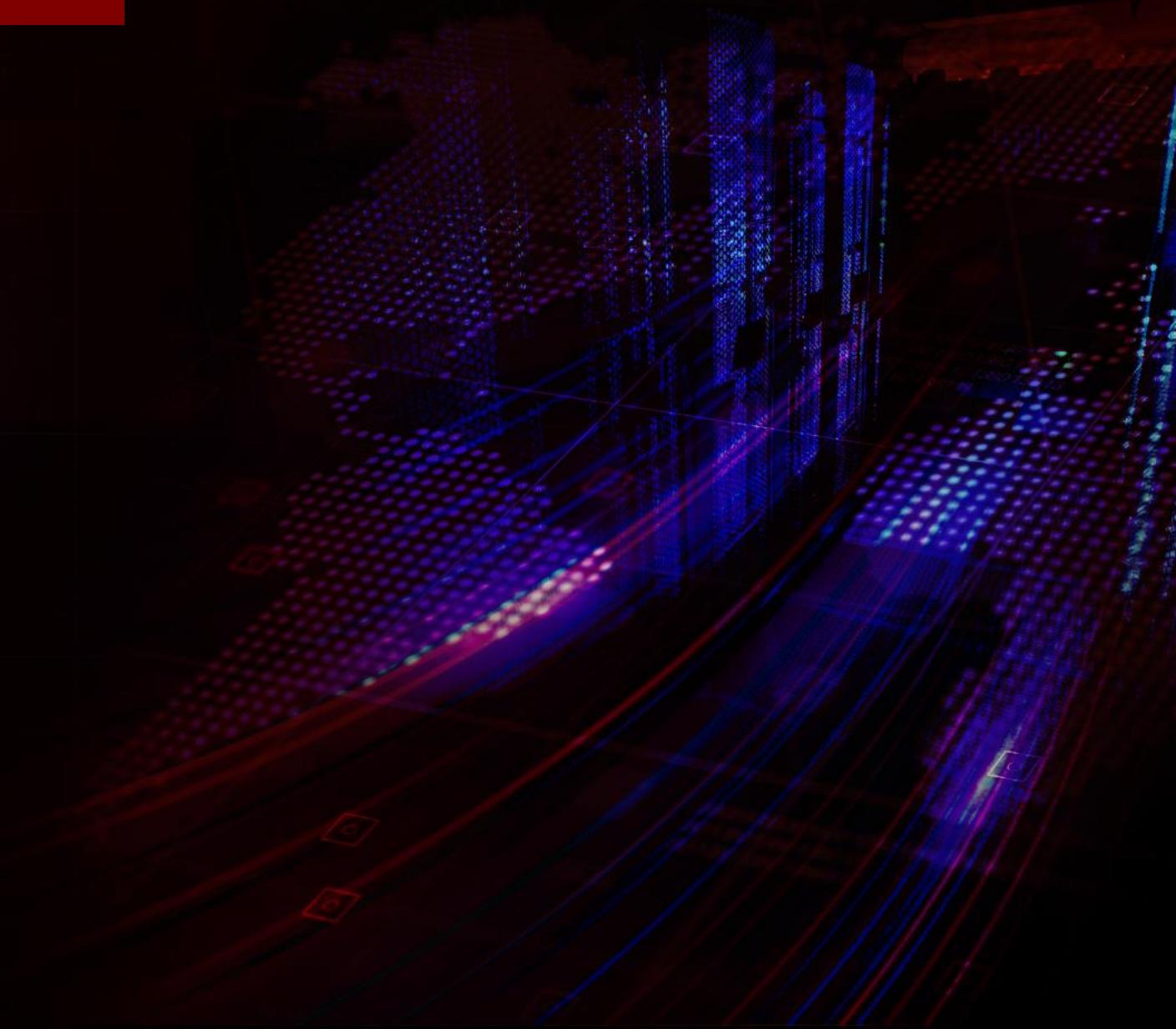


k-NN - EXEMPLO

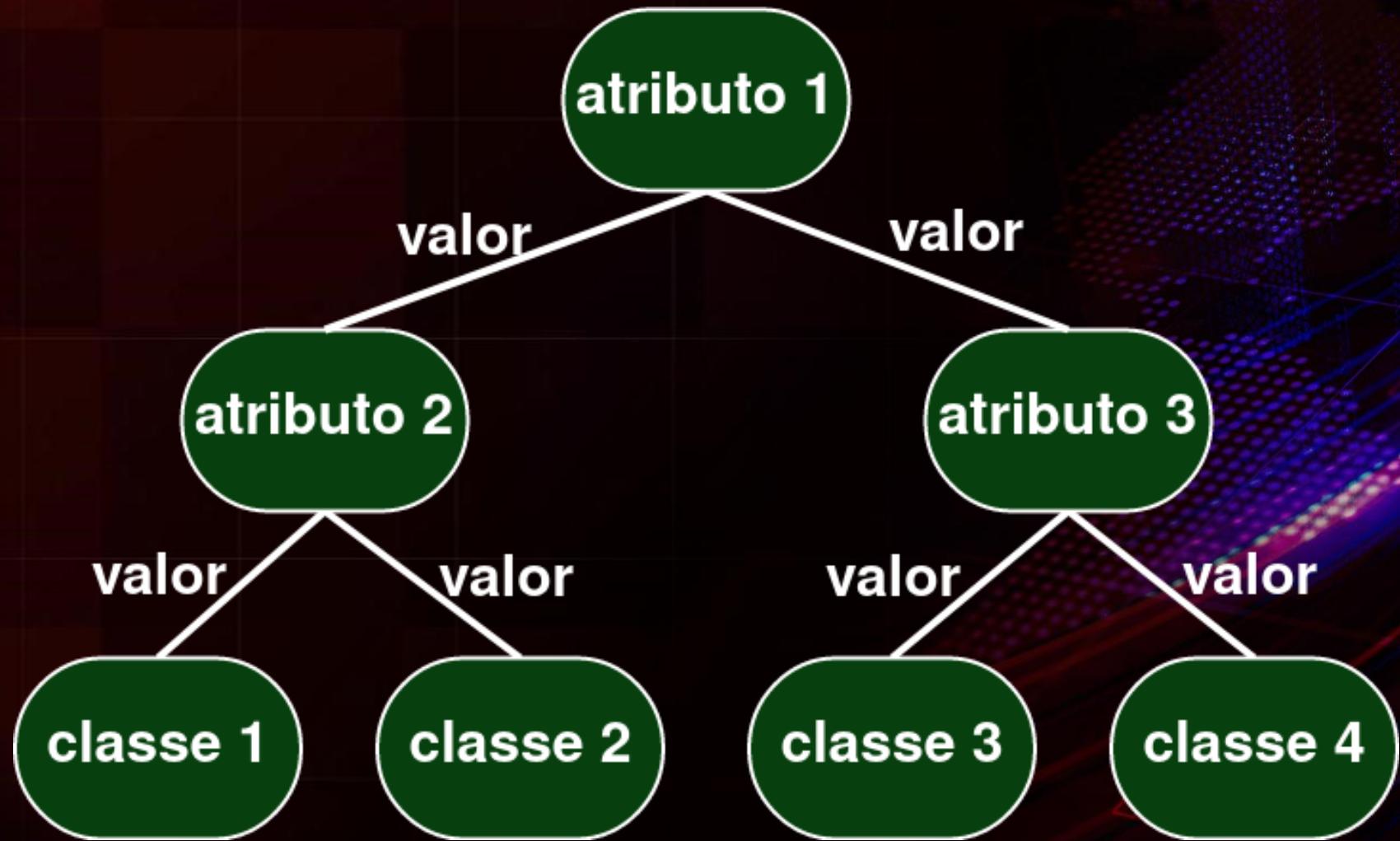


ÁRVORES DE DECISÃO

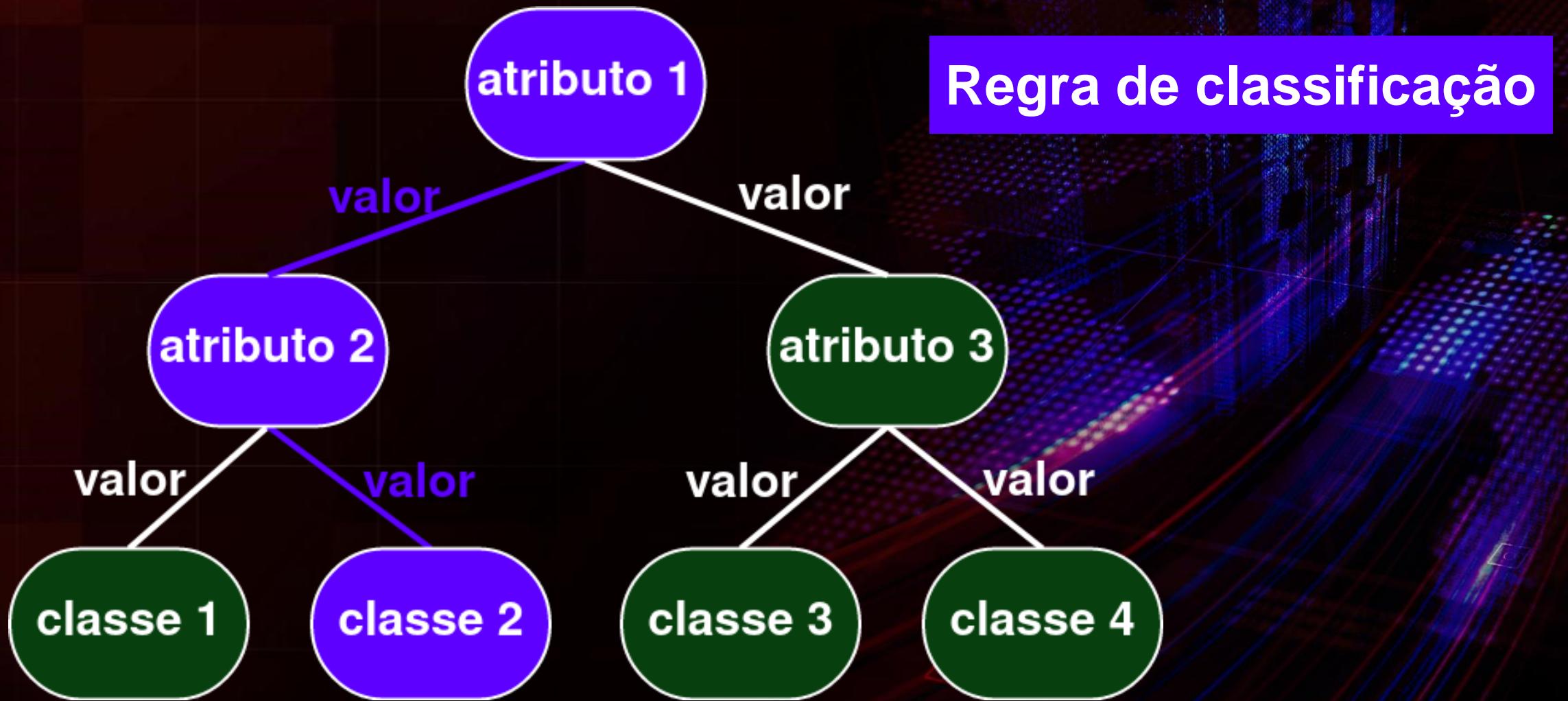
- Estrutura de árvore
- Modelo caixa branca
 - Fácil compreensão



ÁRVORES DE DECISÃO



ÁRVORES DE DECISÃO



ÁRVORES DE DECISÃO - CONSTRUÇÃO

- **Selecione um atributo**
- **Crie um ramo para cada valor**
- **Repita o processo recursivamente para cada ramo**
 - **Use apenas os objetos que atingem o ramo**
 - **Pare quando todos os objetos atingirem a mesma folha**

Como escolher o atributo?

ESCOLHA DO ATRIBUTO

- Calcule o ganho de informação dos atributos:
 - **Ganho = informação - entropia**
- Escolha o atributo com maior ganho de informação
- Crie ramos para cada valor do atributo
- Repita o processo para os demais atributos

ÁRVORES DE DECISÃO - VARIAÇÕES

- ID3 (dicotomizador iterativo 3)
 - Atributos categóricos
- C4.5: evolução do ID3:
 - Atributos discretos, contínuos discretizados, nulos
 - Poda: substitui nós internos por folhas
- C5.0: menor uso de memória, menos regras
- CART (árvores de classificação e regressão)

ÁRVORES DE DECISÃO - EXEMPLO

Compra	Manutenção	Portas	Passageiros	Bagageiro	Segurança	Aceitabilidade
muito alto	muito alto	2	mais	pequeno	baixa	inaceitável
muito alto	baixo	3	4	pequeno	baixa	inaceitável
muito alto	médio	4	mais	grande	média	aceitável
muito alto	baixo	5+	mais	médio	média	aceitável
médio	baixo	3	4	pequeno	alta	bom
baixo	médio	2	4	pequeno	alta	bom
médio	médio	2	4	grande	alta	ótimo
médio	baixo	4	mais	grande	alta	ótimo

Base: carros

ÁRVORES DE DECISÃO - EXEMPLO

Compra	Manutenção	Portas	Passageiros	Bagageiro	Segurança	Aceitabilidade
muito alto	muito alto	2	mais	pequeno	baixa	inaceitável
muito alto	baixo	3	4	pequeno	baixa	inaceitável
muito alto	médio	4	mais	grande	média	aceitável
muito alto	baixo	5+	mais	médio	média	aceitável
médio	baixo	3	4	pequeno	alta	bom
baixo	médio	2	4	pequeno	alta	bom
médio	médio	2	4	grande	alta	ótimo
médio	baixo	4	mais	grande	alta	ótimo

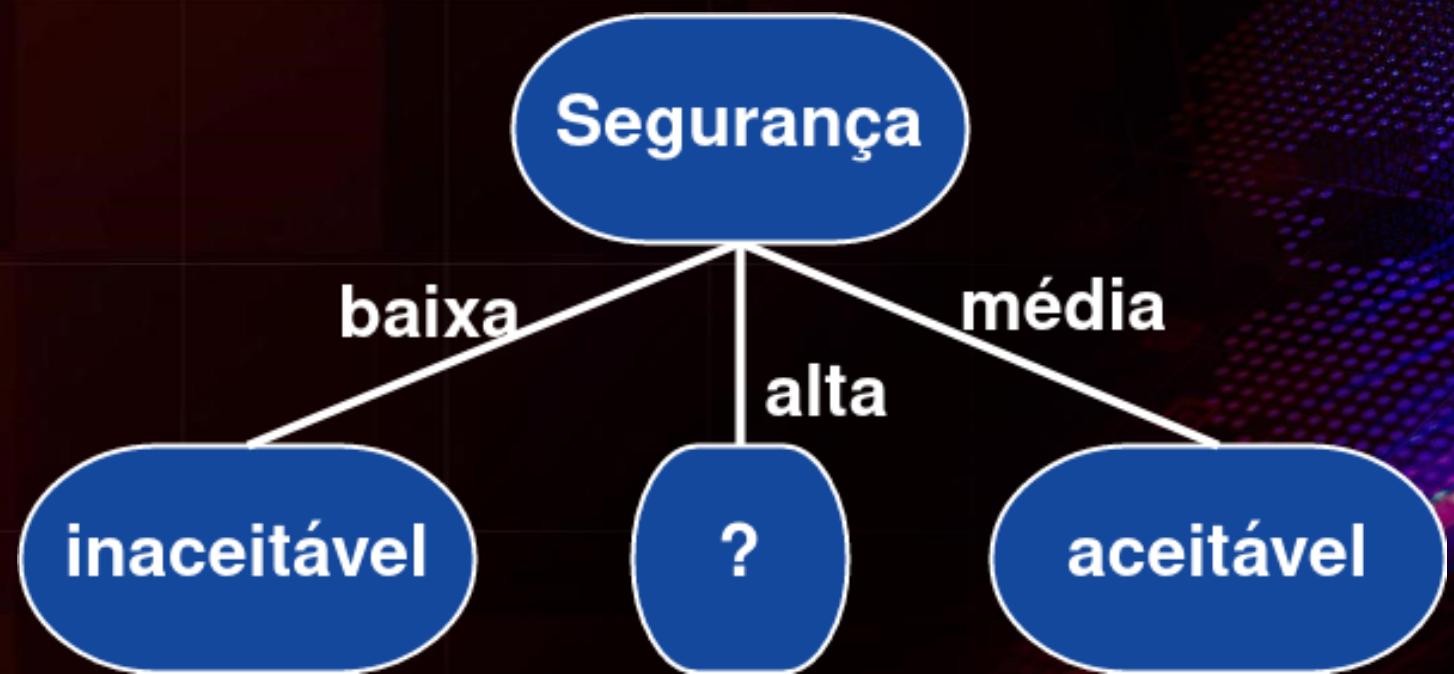
Base: carros

ÁRVORES DE DECISÃO - EXEMPLO

	Compra	Manutenção	Portas	Passageiros	Bagageiro	Segurança
Entropia	0,844	1,594	1,094	1,500	0,844	0,500
Ganho	1,156	0,406	0,906	0,500	1,156	1,500

Fonte: Castro e Ferrari (2016)

ÁRVORES DE DECISÃO - EXEMPLO



ÁRVORES DE DECISÃO - EXEMPLO

Compra	Manutenção	Portas	Passageiros	Bagageiro	Segurança	Aceitabilidade
muito alto	muito alto	2	mais	pequeno	baixa	inaceitável
muito alto	baixo	3	4	pequeno	baixa	inaceitável
muito alto	médio	4	mais	grande	média	aceitável
muito alto	baixo	5+	mais	médio	média	aceitável
médio	baixo	3	4	pequeno	alta	bom
baixo	médio	2	4	pequeno	alta	bom
médio	médio	2	4	grande	alta	ótimo
médio	baixo	4	mais	grande	alta	ótimo

Fonte: Castro e Ferrari (2016)

ÁRVORES DE DECISÃO - EXEMPLO

Compra	Manutenção	Portas	Passageiros	Bagageiro	Segurança	Aceitabilidade
muito alto	muito alto	2	mais	pequeno	baixa	inaceitável
muito alto	baixo	3	4	pequeno	baixa	inaceitável
muito alto	médio	4	mais	grande	média	aceitável
muito alto	baixo	5+	mais	médio	média	aceitável
médio	baixo	3	4	pequeno	alta	bom
baixo	médio	2	4	pequeno	alta	bom
médio	médio	2	4	grande	alta	ótimo
médio	baixo	4	mais	grande	alta	ótimo

Fonte: Castro e Ferrari (2016)

ÁRVORES DE DECISÃO - EXEMPLO

Compra	Manutenção	Portas	Passageiros	Bagageiro	Segurança	Aceitabilidade
muito alto	muito alto	2	mais	pequeno	baixa	inaceitável
muito alto	baixo	3	4	pequeno	baixa	inaceitável
muito alto	médio	4	mais	grande	média	aceitável
muito alto	baixo	5+	mais	médio	média	aceitável
médio	baixo	3	4	pequeno	alta	bom
baixo	médio	2	4	pequeno	alta	bom
médio	médio	2	4	grande	alta	ótimo
médio	baixo	4	mais	grande	alta	ótimo

Fonte: Castro e Ferrari (2016)

ÁRVORES DE DECISÃO - EXEMPLO

	Compra	Manutenção	Portas	Passageiros	Bagageiro
Entropia	0,689	1,000	0,500	0,689	0,000
Ganho	0,311	0,000	0,500	0,311	1,000

Fonte: Castro e Ferrari (2016)

ÁRVORES DE DECISÃO - EXEMPLO



ÁRVORES DE DECISÃO - EXEMPLO

Compra	Manutenção	Portas	Passageiros	Bagageiro	Segurança	Aceitabilidade
muito alto	muito alto	2	mais	pequeno	baixa	inaceitável
muito alto	baixo	3	4	pequeno	baixa	inaceitável
muito alto	médio	4	mais	grande	média	aceitável
muito alto	baixo	5+	mais	médio	média	aceitável
médio	baixo	3	4	pequeno	alta	bom
baixo	médio	2	4	pequeno	alta	bom
médio	médio	2	4	grande	alta	ótimo
médio	baixo	4	mais	grande	alta	ótimo

Fonte: Castro e Ferrari (2016)

ÁRVORES DE DECISÃO - EXEMPLO

Compra	Manutenção	Portas	Passageiros	Bagageiro	Segurança	Aceitabilidade
muito alto	muito alto	2	mais	pequeno	baixa	inaceitável
muito alto	baixo	3	4	pequeno	baixa	inaceitável
muito alto	médio	4	mais	grande	média	aceitável
muito alto	baixo	5+	mais	médio	média	aceitável
médio	baixo	3	4	pequeno	alta	bom
baixo	médio	2	4	pequeno	alta	bom
médio	médio	2	4	grande	alta	ótimo
médio	baixo	4	mais	grande	alta	ótimo

Fonte: Castro e Ferrari (2016)

ÁRVORES DE DECISÃO - EXEMPLO



ONE-RULE (1R)

- Regras que testam um único atributo
- Baixo custo computacional
- Boa acurácia para bases simples

ONE-RULE (1R)

- **Gere um conjunto de regras, criando uma ramificação para cada valor do atributo**
- **Classifique os ramos em ordem de uso da classe mais frequente**
- **Calcule a taxa de erro de classificação**

ONE-RULE (1R)

- Avalie a taxa de erro do conjunto de regras de cada atributo
- Escolha o conjunto de regras do atributo de menor taxa de erro

ONE-RULE (1R)

Base: jogo da velha (958 objetos)

Valores: X, O, b

SE	SC	SD
CE	CC	CD
IE	IC	ID



ONE-RULE (1R)

SE	SC	SD	CE	CC	CD	IE	IC	ID	Vitória
X	X	X	X	O	O	X	O	O	Sim
X	X	X	X	O	O	O	X	O	Sim
X	X	X	X	O	O	O	O	X	Sim
X	X	X	X	O	O	O	b	b	Sim
X	X	X	X	O	O	b	O	b	Sim
O	X	X	X	O	O	O	X	X	Não
O	X	O	X	X	O	X	O	X	Não
O	X	O	X	O	X	X	O	X	Não
O	X	O	O	X	X	X	O	X	Não
O	O	X	X	X	O	O	X	X	Não

Fonte: Castro e Ferrari (2016)

ONE-RULE (1R)

Atributo valor	Classe	# objetos atributo valor	# objetos atributo valor classe	Erro por valor	Erro por atributo
SE - X	Sim	418	295	123	332
SE - b	Sim	205	142	63	
SE - O	Sim	335	189	146	
...	
CC - X	Sim	458	366	92	288
CC - b	Sim	160	112	48	
CC - O	Não	340	192	148	
...	
ID - X	Sim	418	295	123	332
ID - b	Sim	205	142	63	
ID - O	Sim	335	189	146	

ONE-RULE (1R)

Atributo valor	Classe	# objetos atributo valor	# objetos atributo valor classe	Erro por valor	Erro por atributo
SE - X	Sim	418	295	123	332
SE - b	Sim	205	142	63	
SE - O	Sim	335	189	146	
...
CC - X	Sim	458	366	92	288
CC - b	Sim	160	112	48	
CC - O	Não	340	192	148	
...
ID - X	Sim	418	295	123	332
ID - b	Sim	205	142	63	
ID - O	Sim	335	189	146	

ONE-RULE (1R)

Se $CC = X$ ou $CC = b$, então Vitória = Sim

Se $CC = O$, então Vitória = Não

NAÏVE BAYES

- Baseado no Teorema de Bayes
- Independência condicional de classe
 - Atributos contribuem de forma independente
 - Cálculos mais simples
 - Naïve = ingênuo

TEOREMA DE BAYES

- Probabilidade condicional

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

- $P(A|B)$, $P(B|A)$: probabilidades *a posteriori* de A e B
- $P(A)$, $P(B)$: probabilidades *a priori* de A e B

TEOREMA DE BAYES

Ex: qual a probabilidade de ter incêndio onde há fumaça?

- $P(\text{incêndio}) = 1\%$ e $P(\text{fumaça}) = 10\%$
- 90% dos incêndios tem fumaça: $P(\text{fumaça} | \text{incêndio})$

$$P(\text{incêndio} | \text{fumaça}) = \frac{P(\text{fumaça} | \text{incêndio}) * P(\text{incêndio})}{P(\text{fumaça})}$$

$$P(\text{incêndio} | \text{fumaça}) = \frac{90\% * 1\%}{10\%} = 9\%$$

NAÏVE BAYES

- Cada objeto é um vetor com m atributos
- Calcular a probabilidade *a posteriori* (atributo-valor-classe)
 - Multiplicar as probabilidades *a posteriori* de cada atributo, pela probabilidade *a priori* de cada classe
- Escolher a classe de maior probabilidade

NAÏVE BAYES - EXEMPLO

dinheiro	fome	comer
pouco	muita	não
pouco	nenhuma	não
suficiente	nenhuma	não
suficiente	muita	sim
suficiente	pouca	não
muito	nenhuma	não
muito	pouca	sim
muito	muita	sim
pouco	pouca	não
suficiente	pouca	sim

Comer ou não no restaurante?

NAÏVE BAYES - EXEMPLO

dinheiro	fome	comer
pouco	muita	não
pouco	nenhuma	não
suficiente	nenhuma	não
suficiente	muita	sim
suficiente	pouca	não
muito	nenhuma	não
muito	pouca	sim
muito	muita	sim
pouco	pouca	não
suficiente	pouca	sim

Comer ou não no restaurante?

NAÏVE BAYES - EXEMPLO

dinheiro	fome	comer
pouco	muita	não
pouco	nenhuma	não
suficiente	nenhuma	não
suficiente	muita	sim
suficiente	pouca	não
muito	nenhuma	não
muito	pouca	sim
muito	muita	sim
pouco	pouca	não
suficiente	pouca	sim

Probabilidade *a priori*: classes

$$P(\text{sim}) = 4/10 = 0,4$$

$$P(\text{não}) = 6/10 = 0,6$$

NAÏVE BAYES - EXEMPLO

Probabilidade a posteriori: atributo-valor-classe

atributo	valor	classe	
		sim	não
dinheiro	pouco	0	1
dinheiro	suficiente	0.5	0.5
dinheiro	muito	0.66	0.33
fome	nenhuma	0	1
fome	pouca	0.5	0.5
fome	muita	0.66	0.33

NAÏVE BAYES - EXEMPLO

Objeto x:

- **dinheiro: muito**
- **fome: muita**

Probabilidade do objeto pertencer a cada classe:

$$P(\text{sim} | x) = P(\text{dinheiro} | \text{sim}) * P(\text{fome} | \text{sim}) * P(\text{sim})$$

$$P(\text{não} | x) = P(\text{dinheiro} | \text{não}) * P(\text{fome} | \text{não}) * P(\text{não})$$

NAÏVE BAYES - EXEMPLO

atributo	valor	classe	
		sim	não
dinheiro	pouco	0	1
dinheiro	suficiente	0.5	0.5
dinheiro	muito	0.66	0.33
fome	nenhuma	0	1
fome	pouca	0.5	0.5
fome	muita	0.66	0.33

$$\begin{aligned}P(\text{sim}) &= 0,4 \\P(\text{não}) &= 0,6\end{aligned}$$

Objeto x vai comer no restaurante

Probabilidade do objeto pertencer a cada classe:

$$\begin{aligned}P(\text{sim} | x) &= \\(0,66 * 0,66) * 0,4 &= 0,17\end{aligned}$$

$$\begin{aligned}P(\text{não} | x) &= \\(0,33 * 0,33) * 0,6 &= 0,06\end{aligned}$$

REFERÊNCIAS

**Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações:
Cap. 5: Classificação de dados.**

Leandro Nunes de Castro e Daniel Gomes Ferrari. Editora Saraiva, 2016.