

APRENDIZADO DE MÁQUINA

Aprendizado para
Agrupamento



TÓPICOS

- Modelos descritivos
- Conceitos principais de agrupamento
- Etapas da análise de agrupamento

MODELOS DESCRIPTIVOS

Não há um elemento externo para guiar o aprendizado

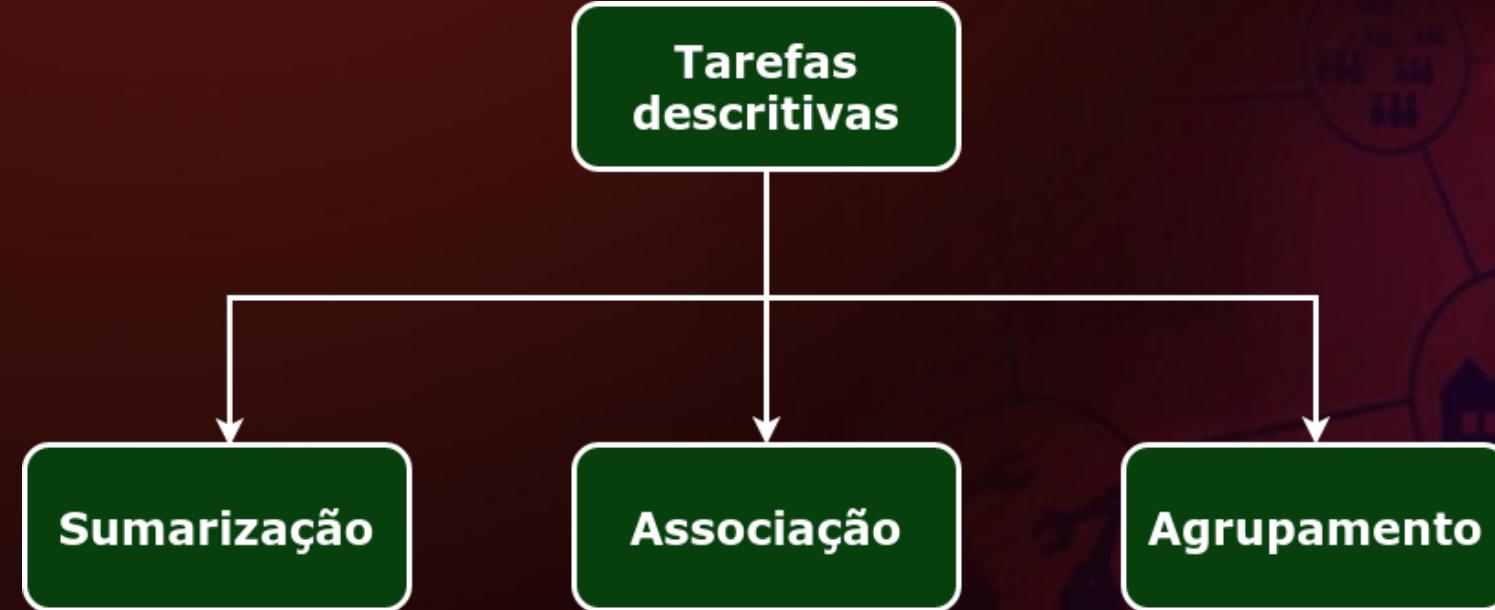
Utiliza-se as propriedades intrínsecas dos objetos para decidir suas semelhanças e diferenças

Aprendizado não supervisionado

Agrupamento de dados

Métodos de análise de dados com o objetivo de descobrir grupos homogêneos

MODELOS DESCRIPTIVOS



SUMARIZAÇÃO

Descrever as características dos dados

Medidas estatísticas:

- **Média, mediana, moda, desvio padrão...**

Técnicas de visualização mais simples:

- **Histograma, gráfico de dispersão, gráfico de setores...**

ASSOCIAÇÃO

Identificar padrões frequentes de associações entre atributos

Mineração de padrões frequentes

Regras de associação

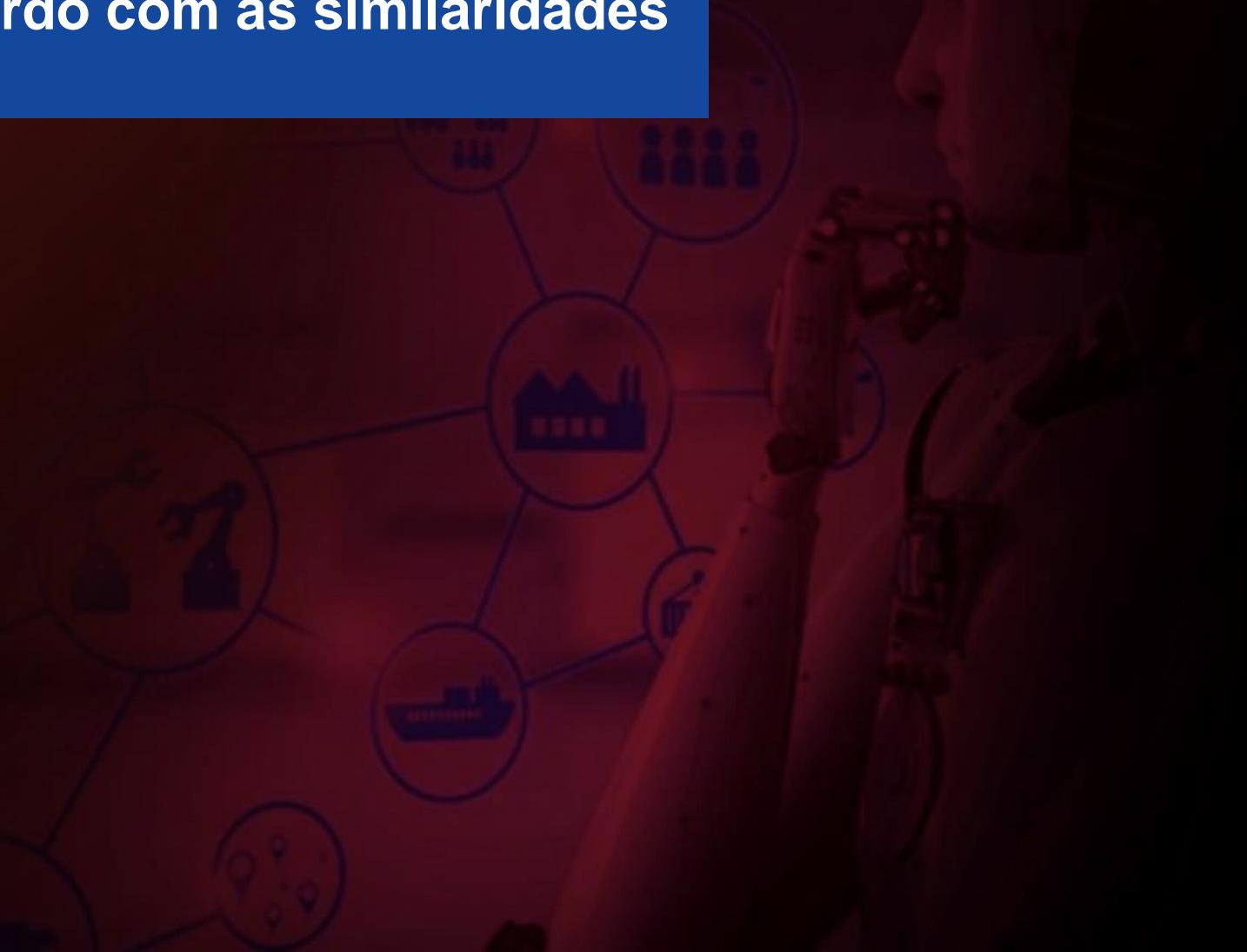
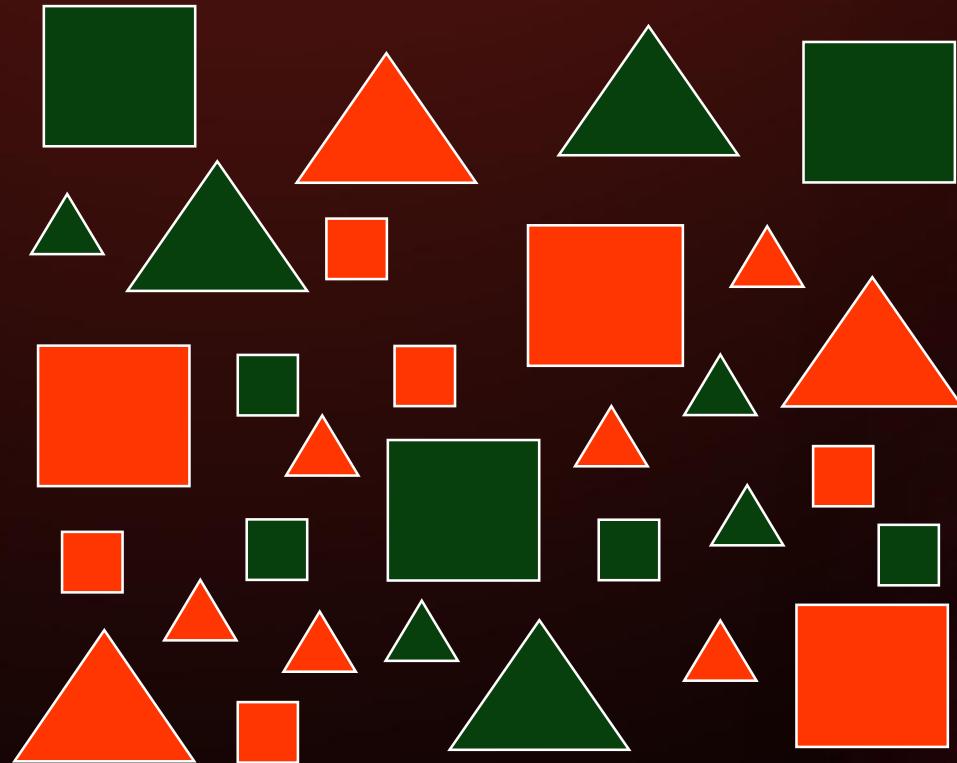
Ex: produtos frequentemente comprados juntos

Apriori, FP-Growth



AGRUPAMENTO

Identificação de grupos de acordo com as similaridades entre os objetos



USOS DE AGRUPAMENTO DE DADOS

Medicina: identificar categorias de diagnósticos

Biologia: taxonomia de animais e plantas

Marketing: identificar grupos de clientes

Análise de imagens: detecção facial

AGRUPAMENTO X CLASSIFICAÇÃO

Agrupamento



Aprendizado não supervisionado

Rótulos

subconjuntos

Treinamento

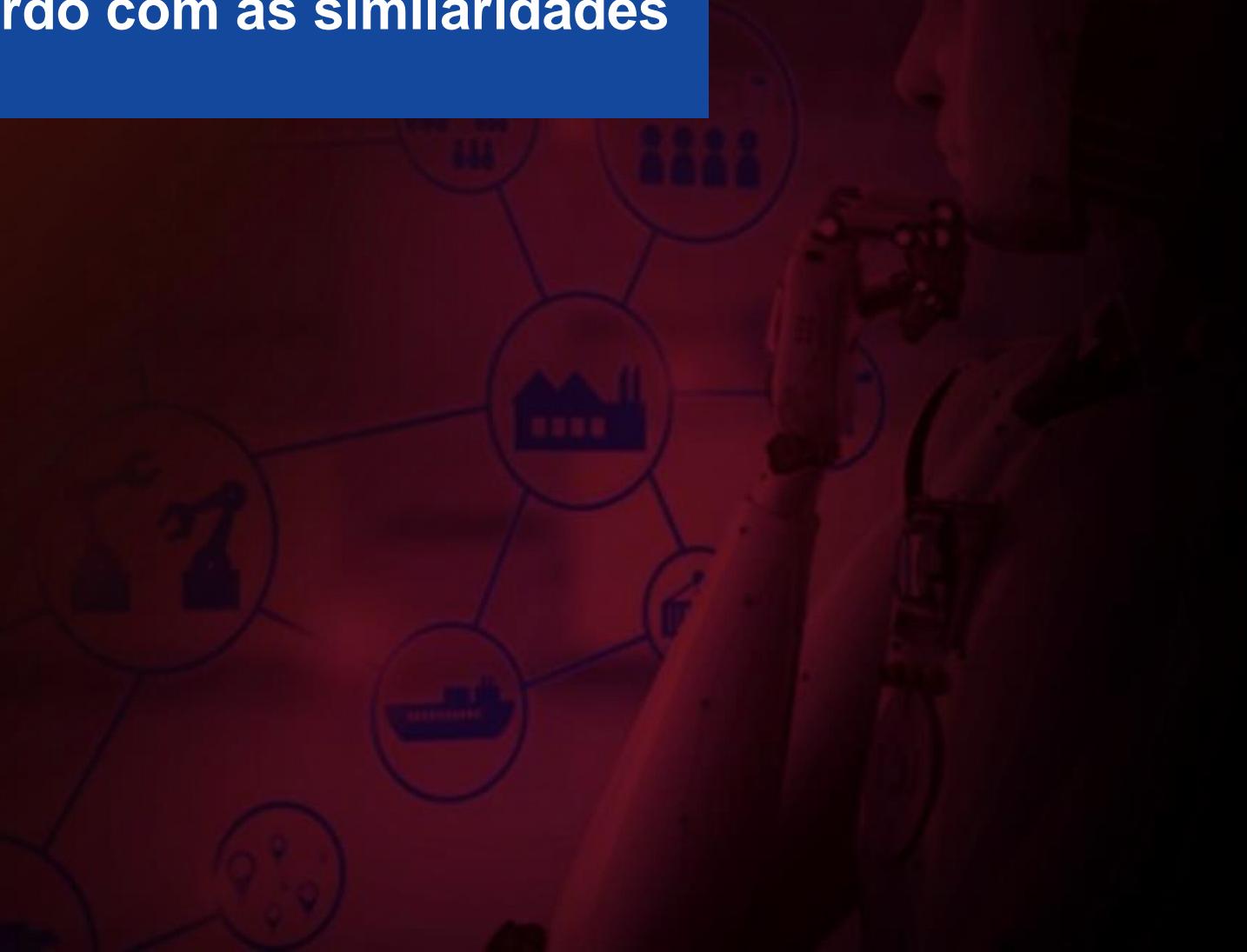
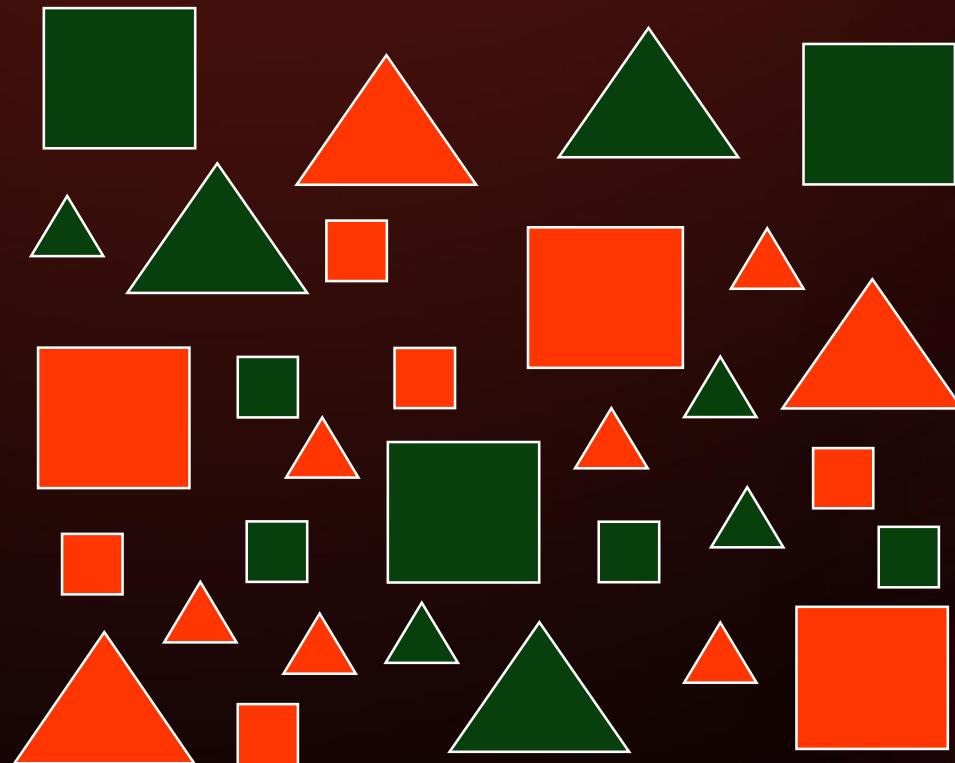
Classificação



Aprendizado supervisionado

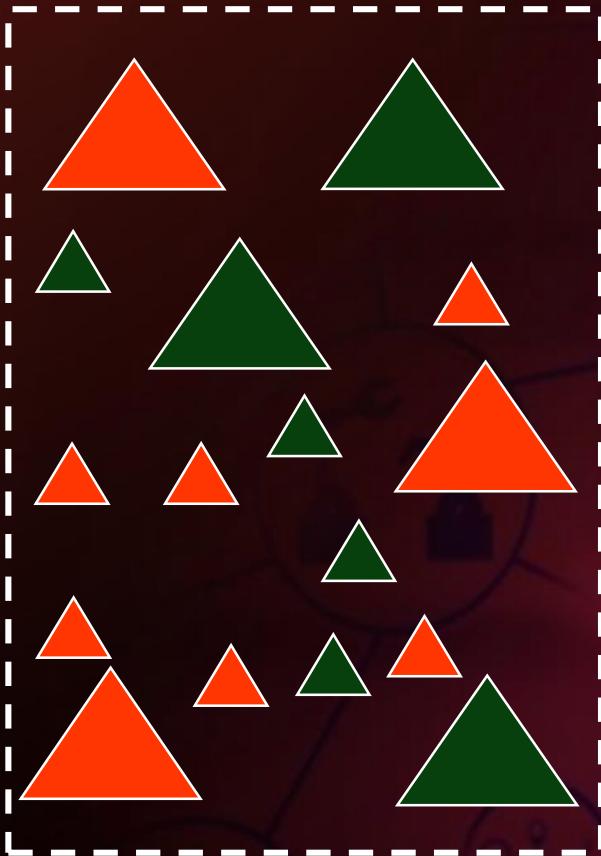
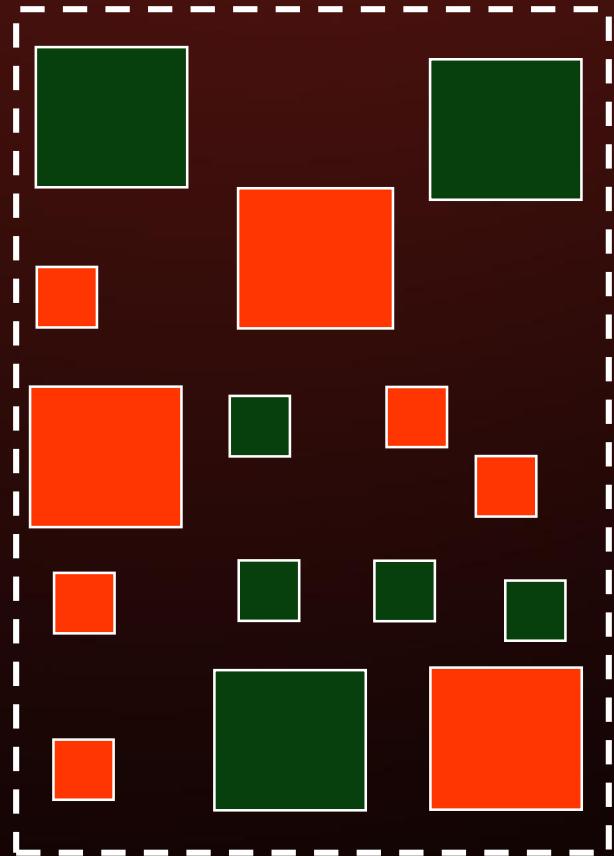
AGRUPAMENTO

Identificação de grupos de acordo com as similaridades entre os objetos



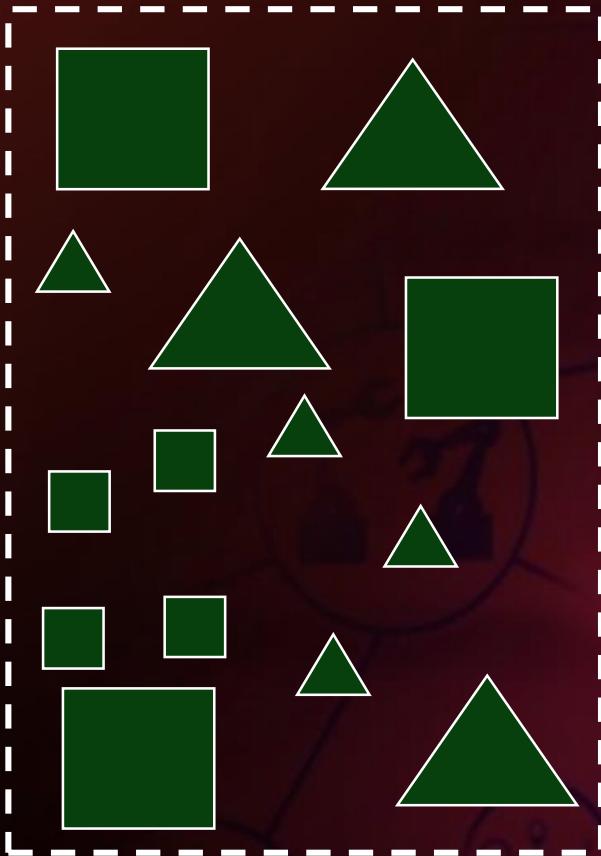
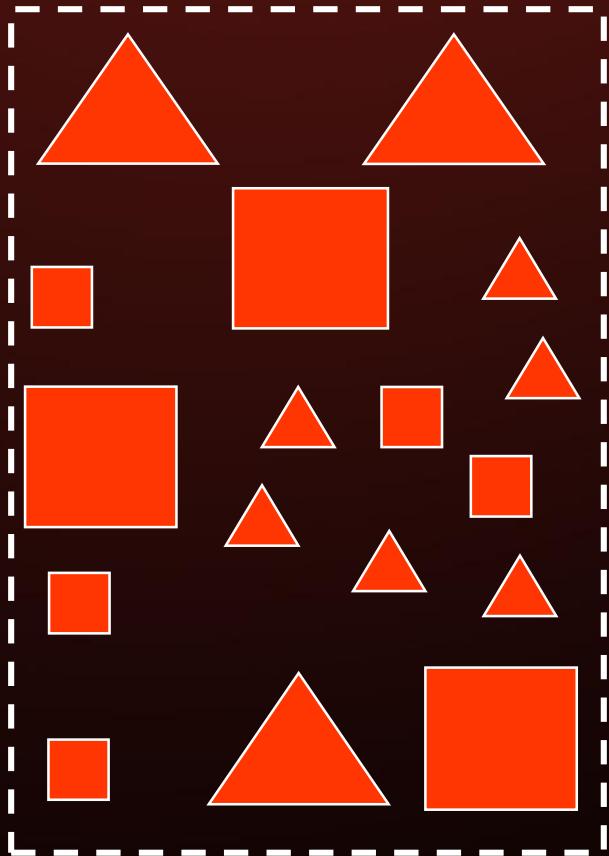
AGRUPAMENTO

Agrupamento pela forma



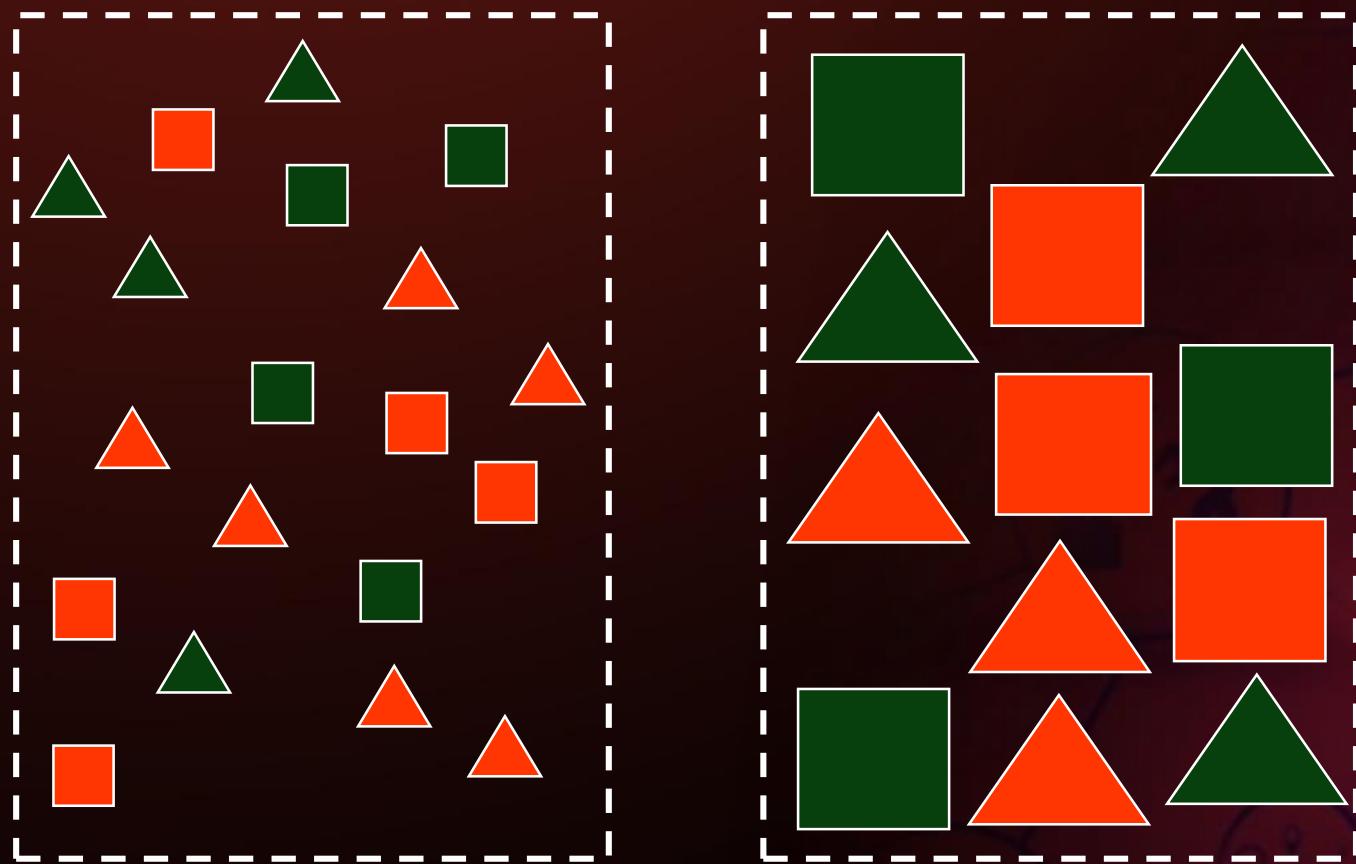
AGRUPAMENTO

Agrupamento pela cor



AGRUPAMENTO

Agrupamento pelo tamanho



CLUSTER

Grupo de objetos similares

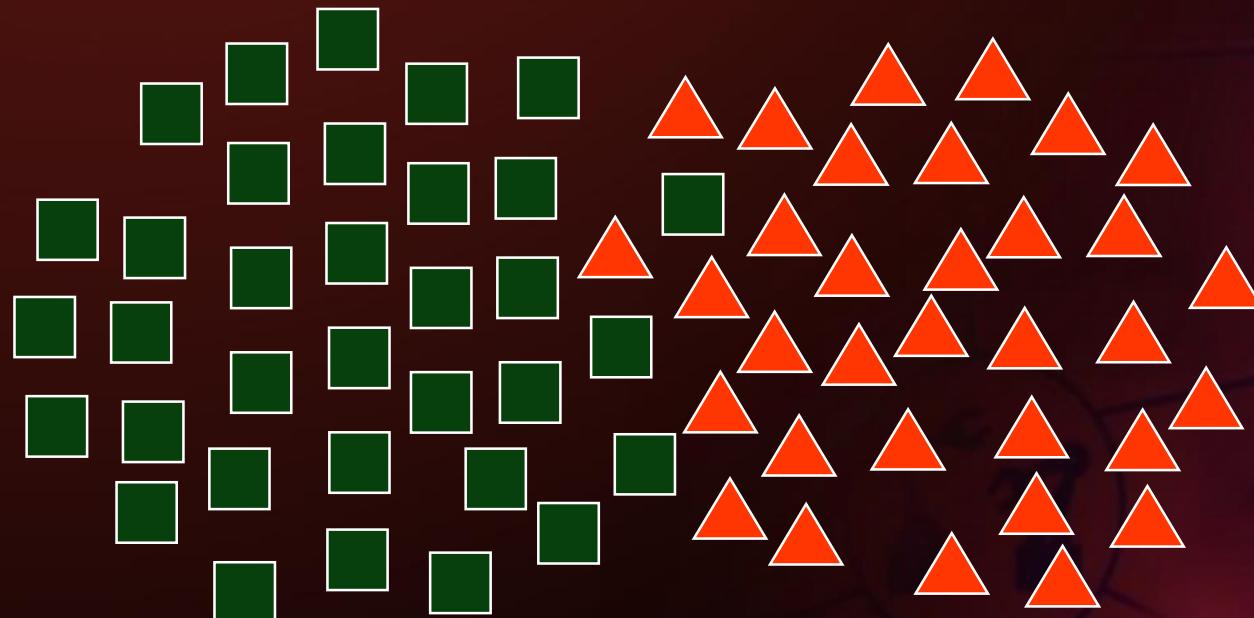
- **Cluster bem separado:**
 - Objetos similares mais próximos do que objetos distintos
- **Cluster baseado em centro:**
 - Objeto que representa o conjunto (real ou calculado)
- **Cluster contínuo ou encadeado:**
 - Objetos similares mais próximos do que objetos distintos
- **Cluster baseado em densidade:**
 - Regiões mais densas com objetos similares
- **Cluster baseado em similaridade:**
 - Conjunto de pontos similares

CRITÉRIOS DE AGRUPAMENTO

- **Compactação:**
 - Critério de homogeneidade, compactação
 - *Clusters* esféricos ou bem separados
 - Ex: k-médias
- **Encadeamento (ligação):**
 - Critério local, de vizinhança
 - *Clusters* com formatos arbitrários
 - Ex: hierárquico de ligação simples
- **Separação espacial:**
 - Distância natural entre os *clusters*
 - Simples
 - Uso em conjunto com os demais critérios

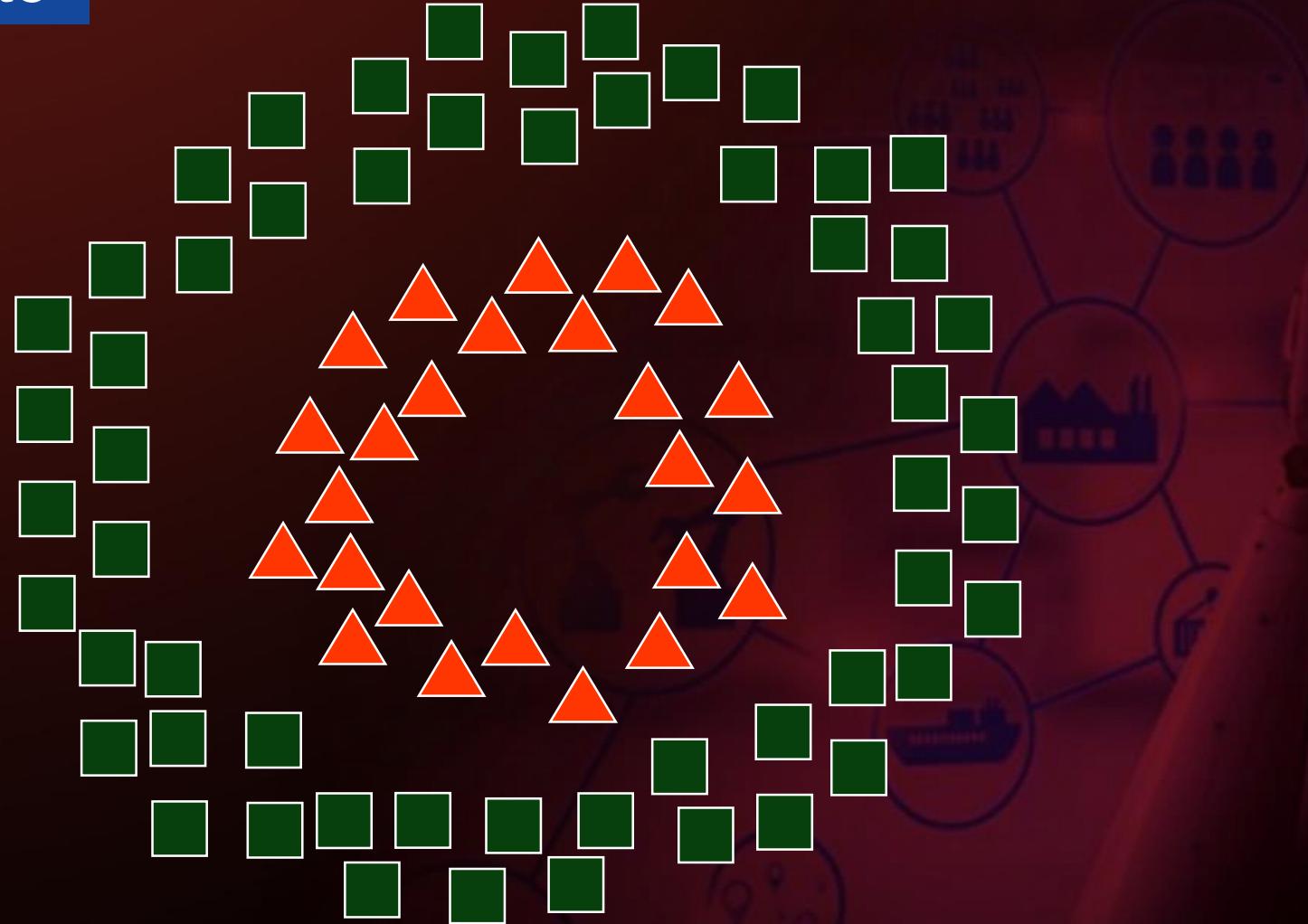
CRITÉRIOS DE AGRUPAMENTO

Compactação



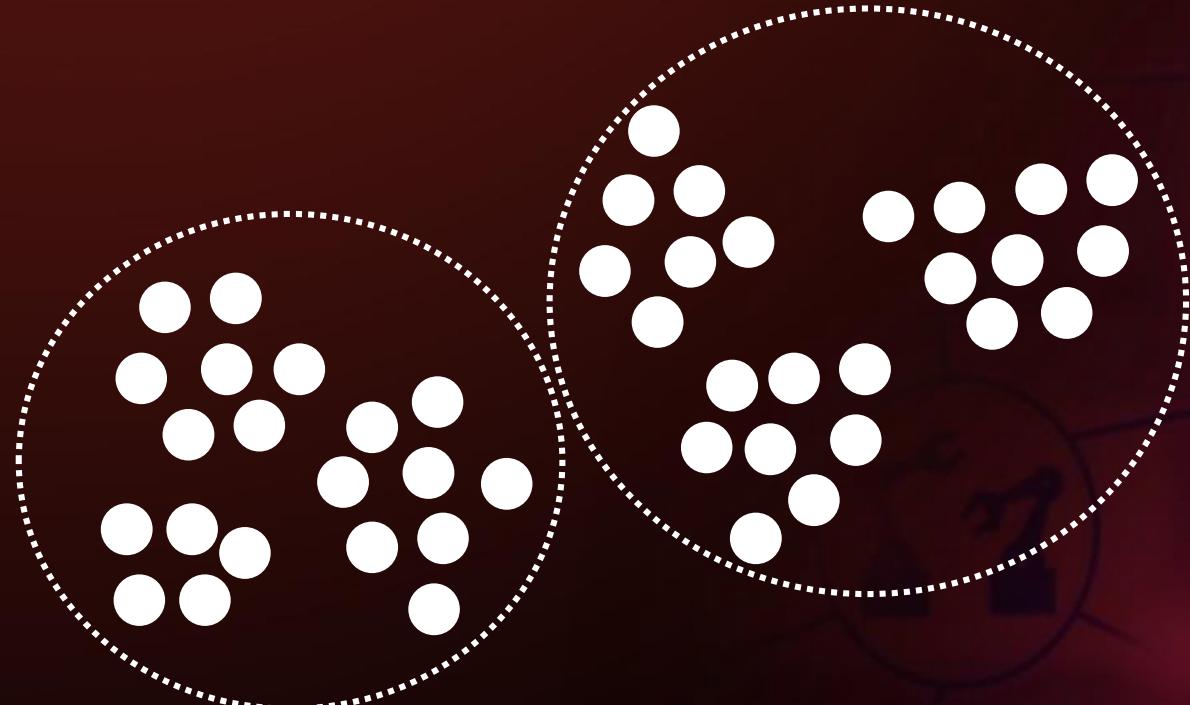
CRITÉRIOS DE AGRUPAMENTO

Encadeamento



REFINAMENTO

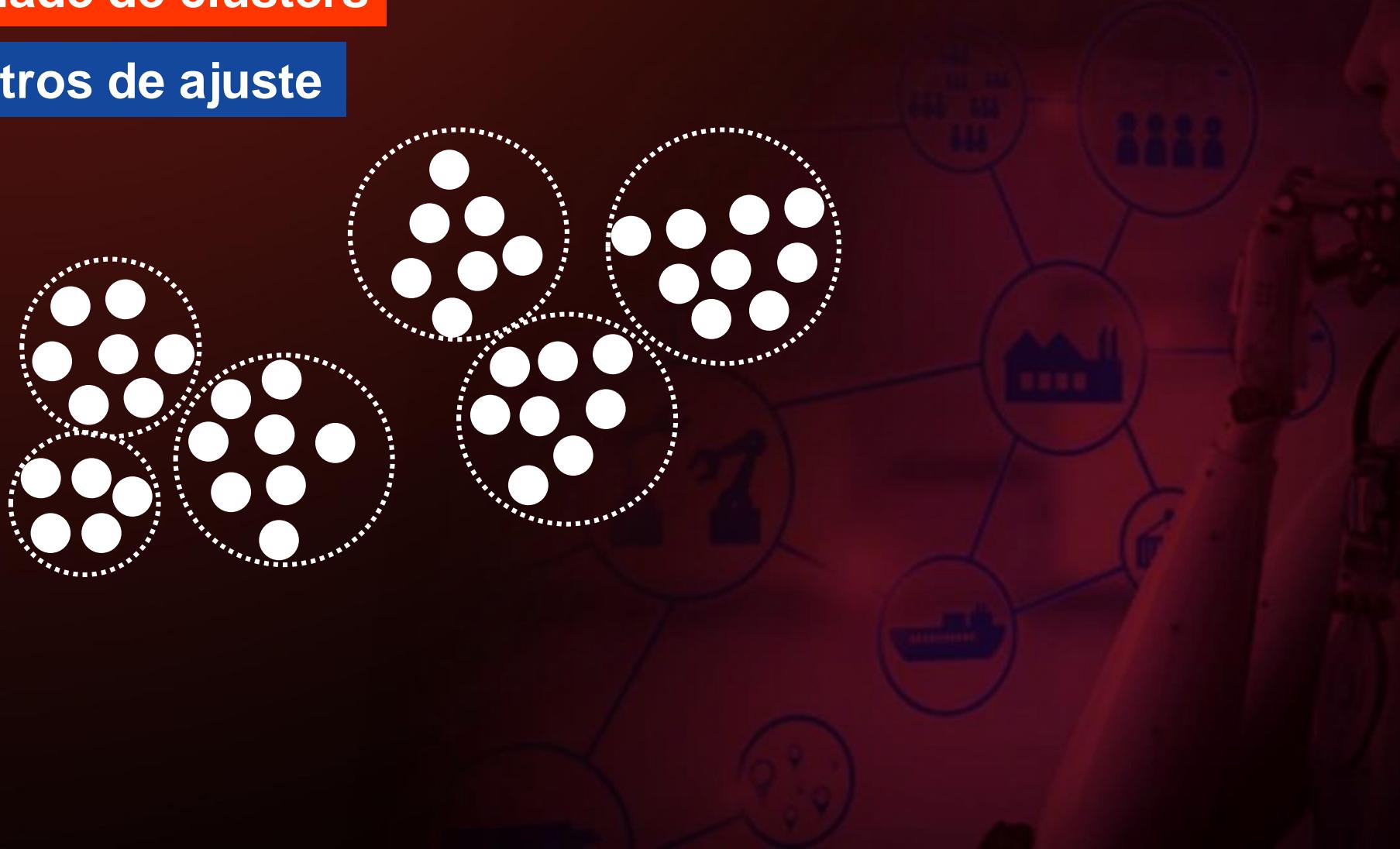
Quantidade de *clusters*



REFINAMENTO

Quantidade de clusters

Parâmetros de ajuste



ETAPAS DE ANÁLISE DE AGRUPAMENTO



ETAPAS DE ANÁLISE DE AGRUPAMENTO



Conjuntos de dados
a serem analisados

ETAPAS DE ANÁLISE DE AGRUPAMENTO



Ajuda a guiar as escolhas técnicas

ETAPAS DE ANÁLISE DE AGRUPAMENTO



Pré-processamento

Normalização dos dados

Magnitude

Matriz de similaridade

ETAPAS DE ANÁLISE DE AGRUPAMENTO



Escolher a medida de distância

Atributos quantitativos

- Métrica de Minkowski

ETAPAS DE ANÁLISE DE AGRUPAMENTO



Escolher a medida de distância

Atributos quantitativos

- Métrica de Minkowski

Distância de Manhattan:

- Dados binários

ETAPAS DE ANÁLISE DE AGRUPAMENTO



Escolher a medida de distância

Atributos quantitativos

- Métrica de Minkowski

Distância euclidiana:

- Dados gerais

ETAPAS DE ANÁLISE DE AGRUPAMENTO



Escolher a medida de distância

Atributos quantitativos

- Métrica de Minkowski

Distância de Chebyschev:

- Distância máxima entre atributos

ETAPAS DE ANÁLISE DE AGRUPAMENTO



Etapa central do processo

Escolha do algoritmo de agrupamento

ETAPAS DE ANÁLISE DE AGRUPAMENTO



Avaliar se os *clusters* formados s o representativos

ETAPAS DE ANÁLISE DE AGRUPAMENTO



Examinar e rotular os *clusters*

ETAPAS DE ANÁLISE DE AGRUPAMENTO



Analisar os resultados obtidos

Uso dos dados para os fins desejados

REFERÊNCIA

Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina. Katti Facelli et al. SAGAH, 2021.