

APRENDIZADO DE MÁQUINAS

Algoritmos de Agrupamento



TÓPICOS

Tipos de algoritmos:

- Hierárquicos
- Particionais
- Baseados em densidade
- Baseados em grafos
- Baseados em redes neurais
- Baseados em grid

CATEGORIAS DE AGRUPAMENTO

- **Hierárquicos**
 - Aglomerativos
 - Divisivos
- **Particionais**
 - Exclusivos
 - Não exclusivos

AGRUPAMENTO HIERÁRQUICO

- **Hierárquicos aglomerativos**
 - Cada objeto pertencendo a um grupo
 - Objetos são agrupados até um critério de parada
- **Hierárquicos divisivos**
 - Todos os objetos pertencem ao mesmo grupo
 - Objetos são divididos até um critério de parada

AGRUPAMENTO PARTICIONAL

- **Particionais exclusivos**
 - Cada objeto pertence a um único grupo
- **Particionais não exclusivos**
 - Cada objeto pode pertencer a mais de um grupo
 - Agrupamento *fuzzy*

CATEGORIAS DE AGRUPAMENTO

- **Monotéticos ou politéticos**
 - Número de atributos usados para calcular a distância
- ***Hard* ou *fuzzy***
 - Pertencimento integral ou parcial de um objeto aos grupos
- **Determinístico ou estocástico**
 - Resultado do agrupamento

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

- Métricas de ligação:
 - Distância entre *clusters*
 - Nível de granularidade flexível
 - Critério de parada indeterminado
 - Não melhora os *clusters*

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Hierárquicos

Particionais

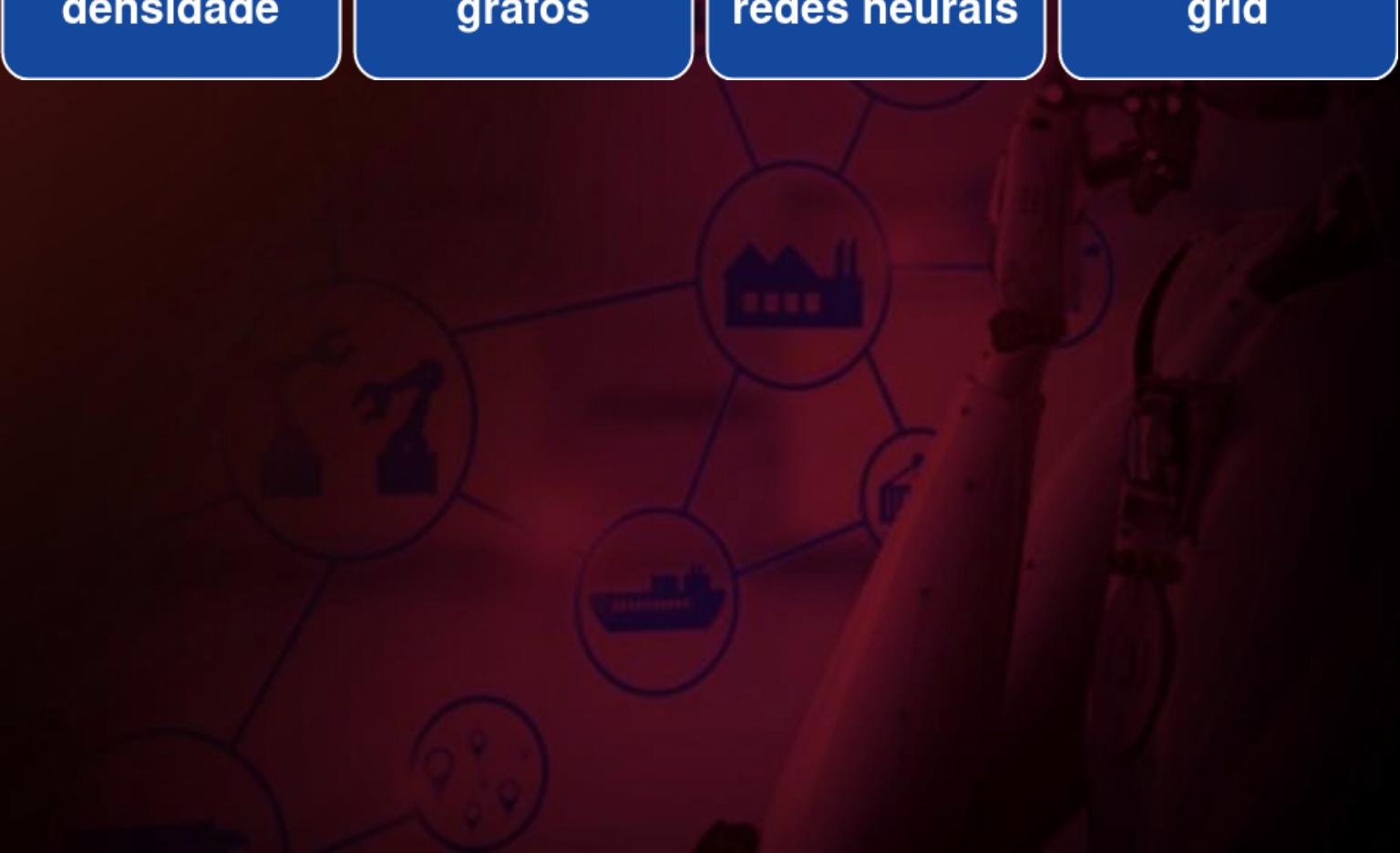
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

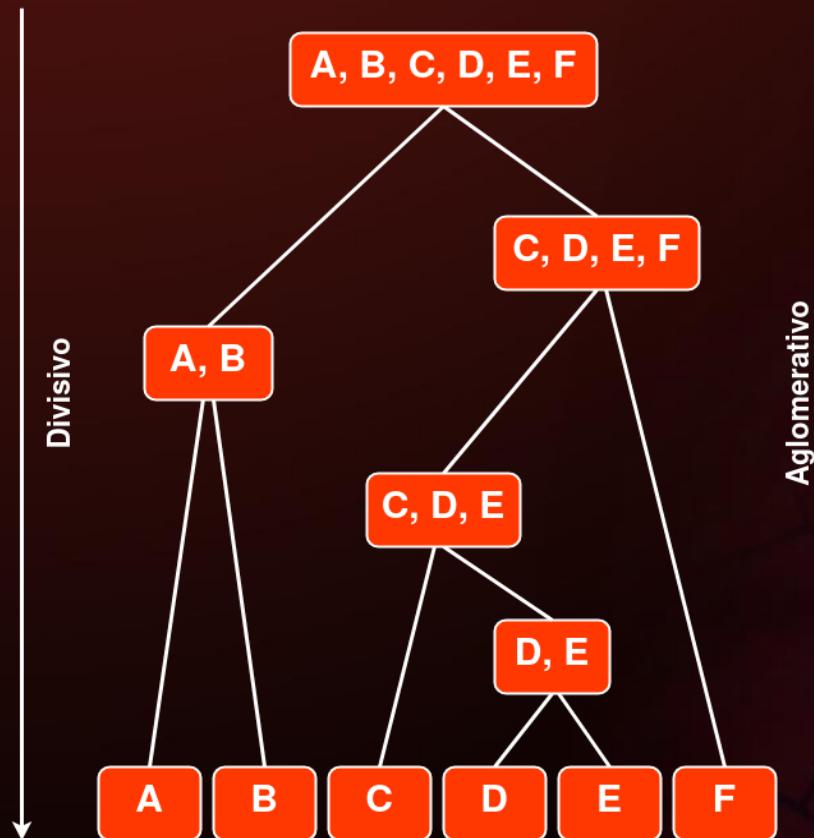
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



Aglomerativo



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

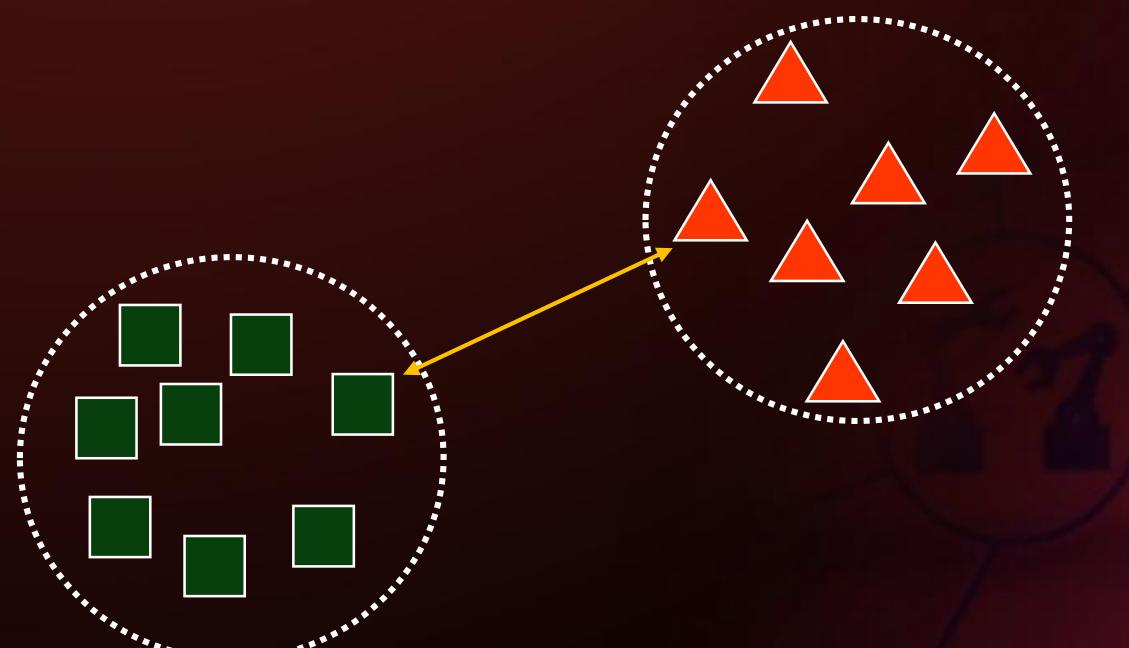
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



Ligaçāo simples
(single-link)

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

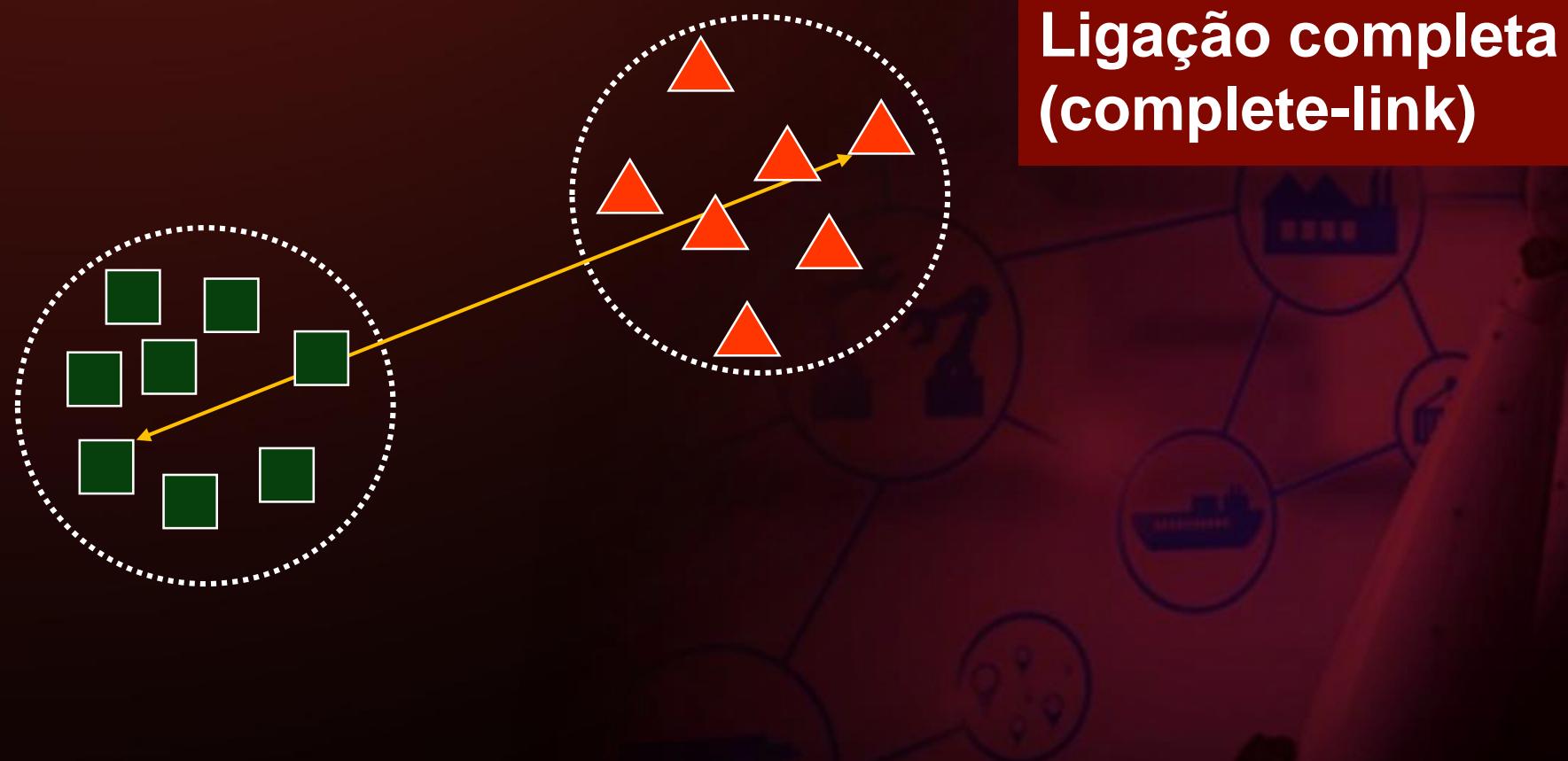
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

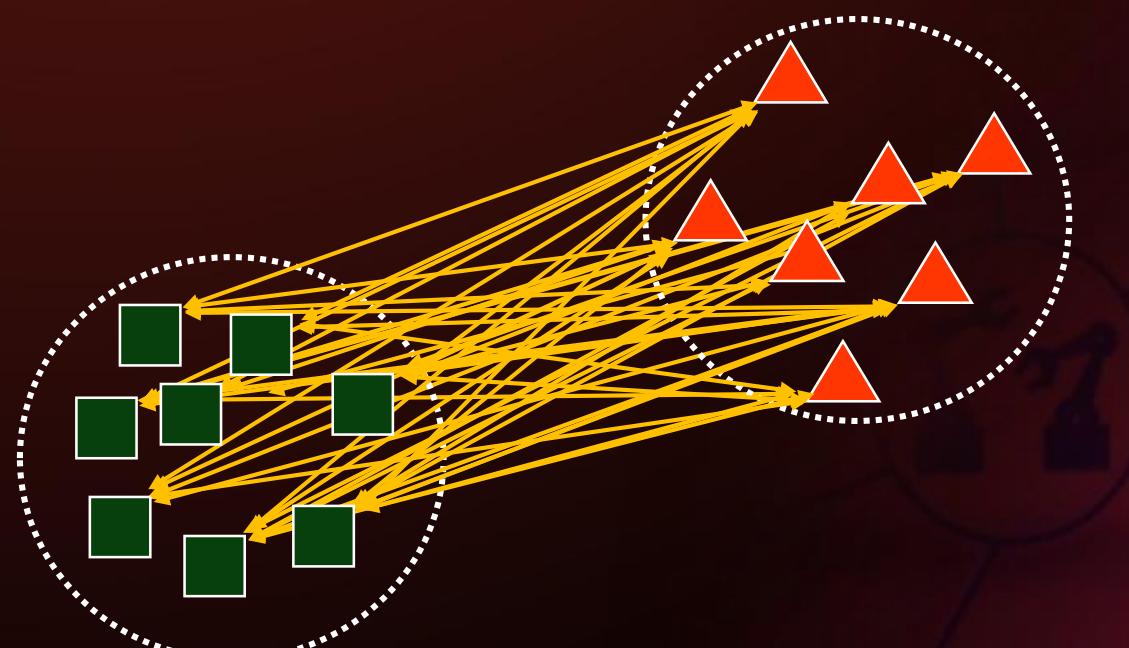
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



Ligaçāo média
(average-link)

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

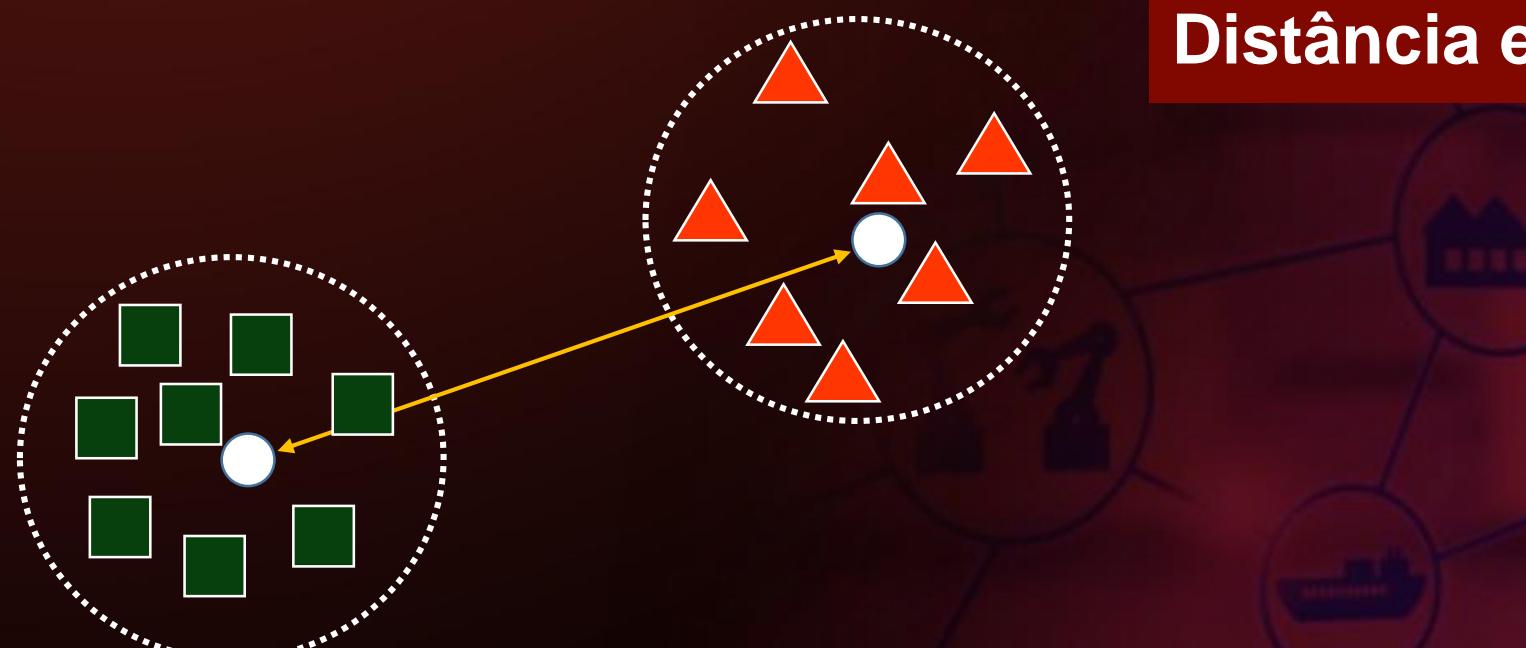
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



Distância euclidiana

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

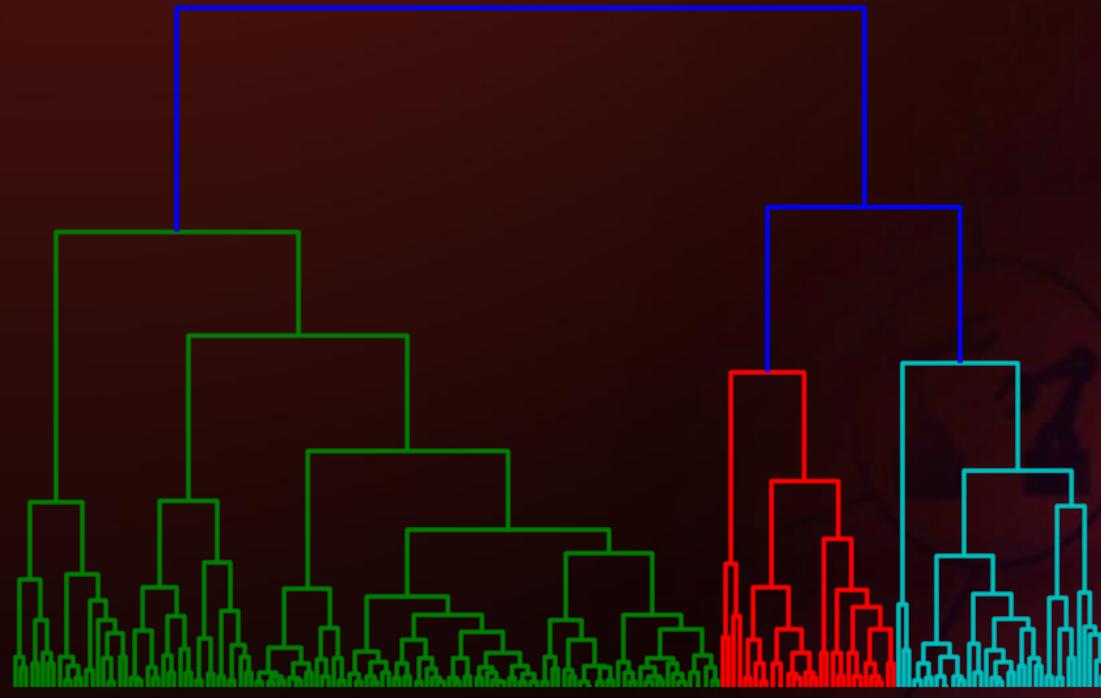
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



Dendograma

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

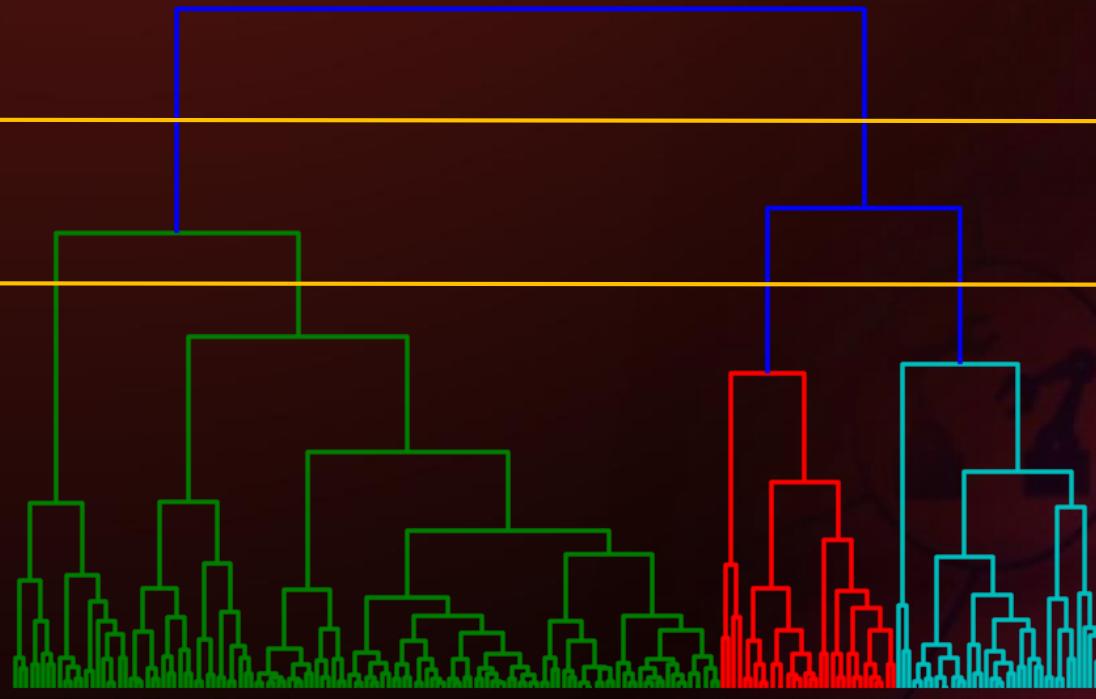
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



$k = 2$

$k = 4$

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

- Agrupamento é feito de forma iterativa:
 - Partição inicial aleatória
- Critério de agrupamento:
 - Erro quadrático:
 - Minimizar distância intracluster
 - Maximizar distância intercluster
- Conjunto de *clusters k* definido pelo usuário



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

- Pode-se escolher diferentes medidas de similaridade
- Critério de parada:
 - Centroide não é mais alterado

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

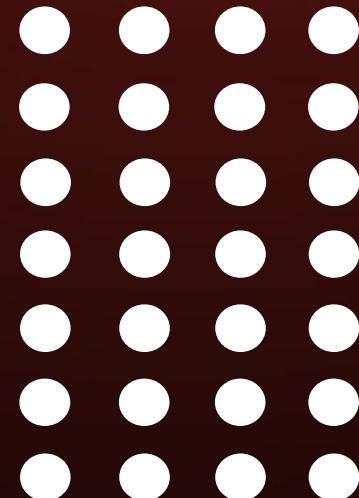
Particionais
baseados em
erro quadrático

Baseados em
densidade

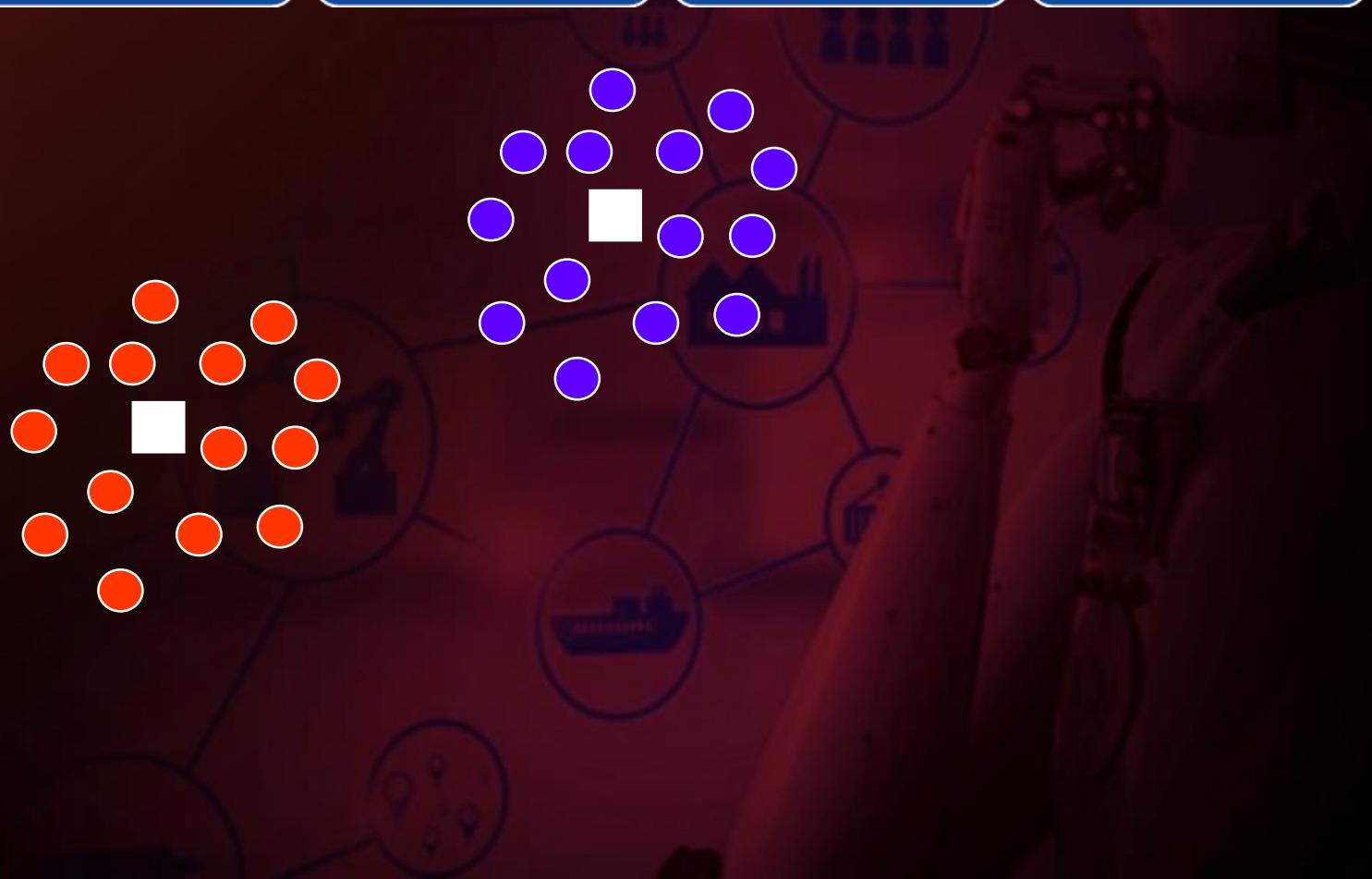
Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



k-médias



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

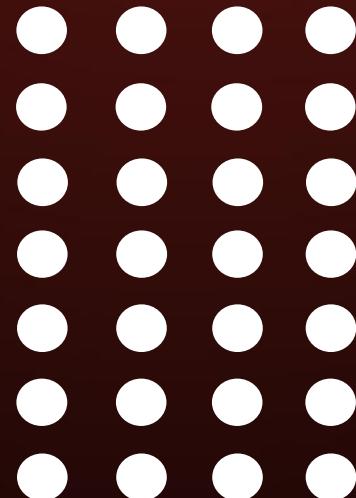
Particionais
baseados em
erro quadrático

Baseados em
densidade

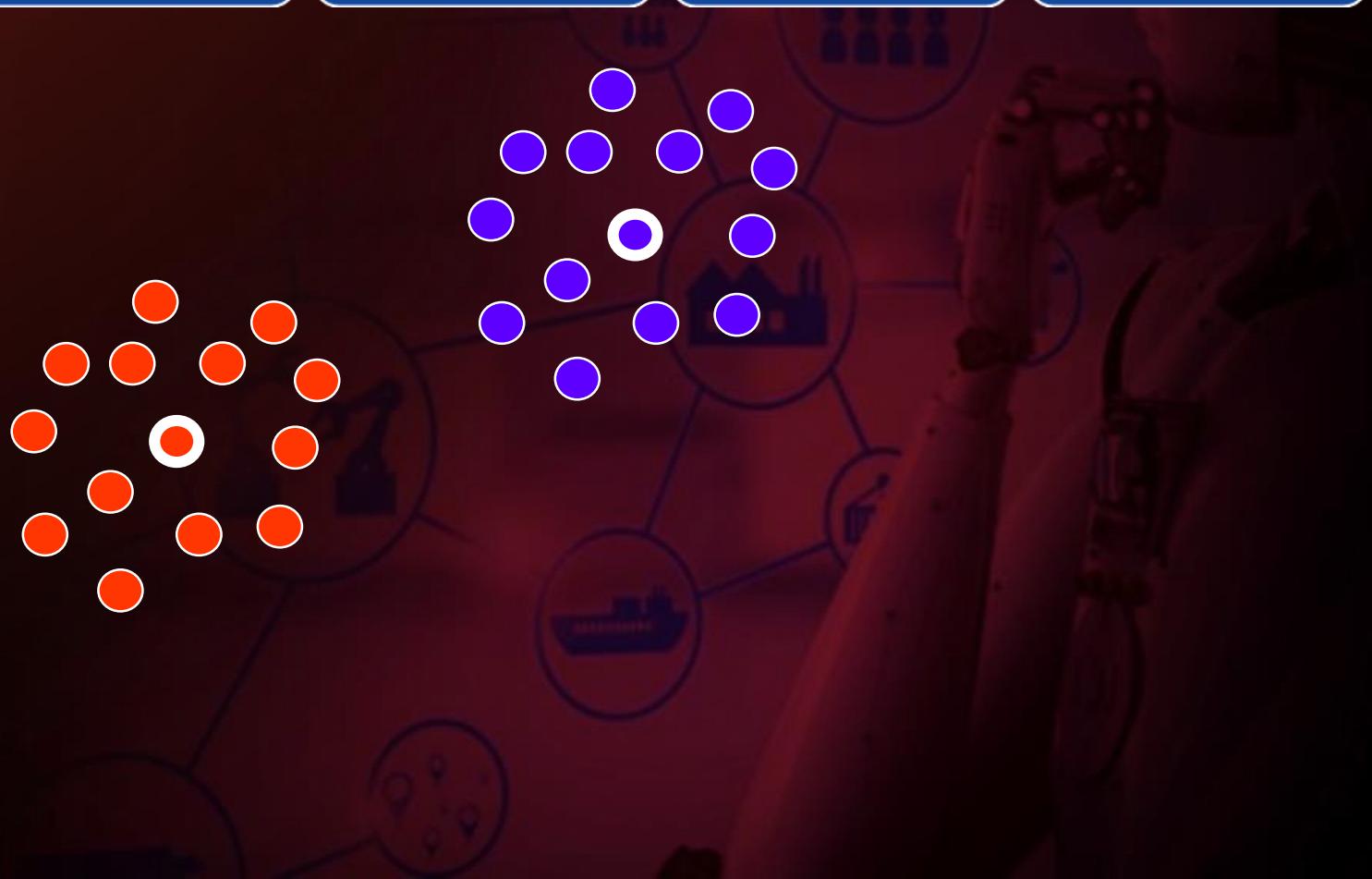
Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



k-medoides



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

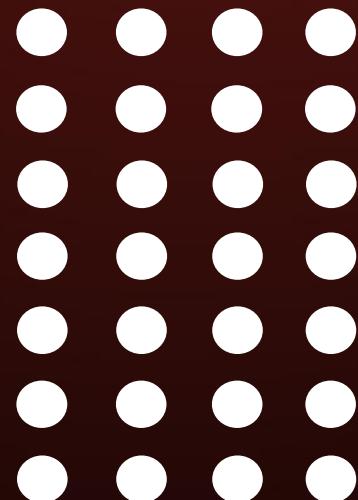
Particionais
baseados em
erro quadrático

Baseados em
densidade

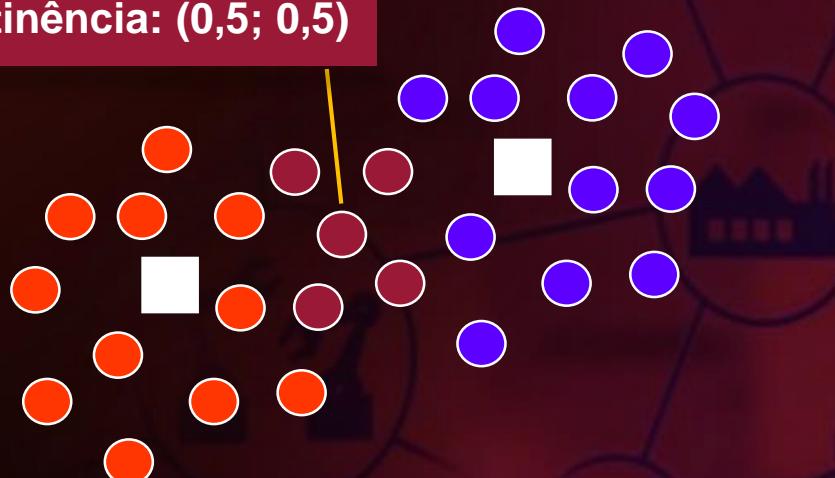
Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



Grau de pertinência: (0,5; 0,5)



fuzzy c-médias

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

- Consideram que os *clusters* são regiões de alta densidade de objetos
- A separação dos *clusters* ocorre em regiões de baixa densidade
- Gera *clusters* de formas arbitrárias
- Não é necessário definir o número de *clusters*

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

DENCLUE: DENsity CLUstEring

- Modela a densidade global de um conjunto de pontos:
 - Soma das funções de influência de cada ponto
 - Ex: distância euclidiana
 - Picos locais são atratores de pontos:
 - Formação de *clusters*

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

DENCLUE: DENsity CLUstEring

- Pré-agrupamento:
 - Mapeamento do espaço
- Agrupamento:
 - Identifica os atratores de densidade (picos locais)
 - Identifica os pontos atraídos

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

DENCLUE



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

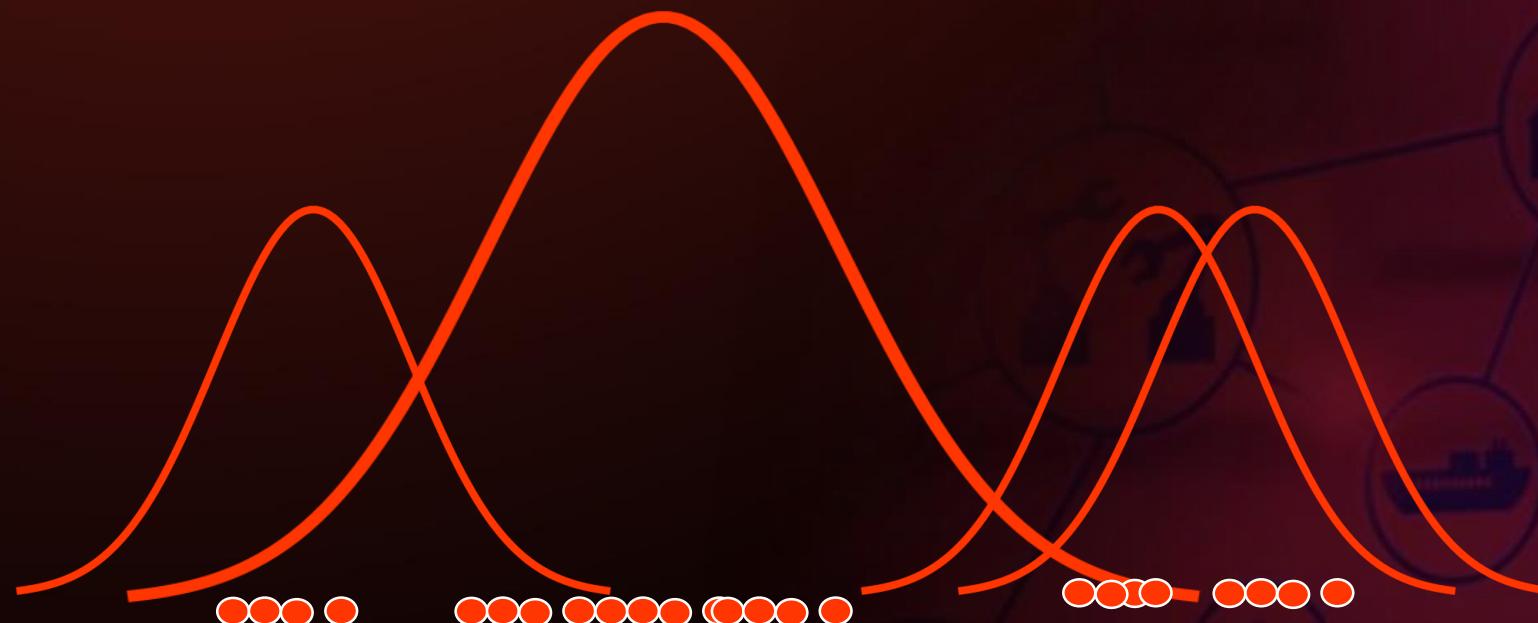
Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

DENCLUE



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

- **Grafo de proximidade:**
 - Representa um conjunto de dados
 - Cada nó é um objeto conectado aos outros nós
 - Aresta representa similaridade ou distância
- **Agrupamento:**
 - Remoção ou inclusão de nós

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

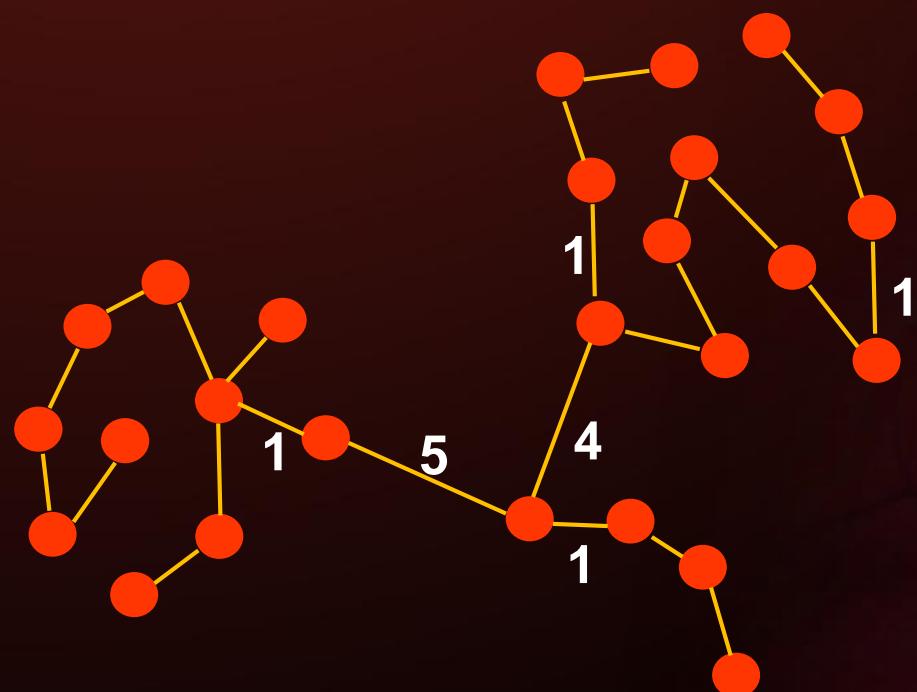
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

CLICK: CLuster Identification via Connectivity Kernels

- Dados são representados por um grafo de similaridade
- Partição recursiva em dois subconjuntos
- Critério de parada determina os subconjuntos finais:
 - *Clusters*

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

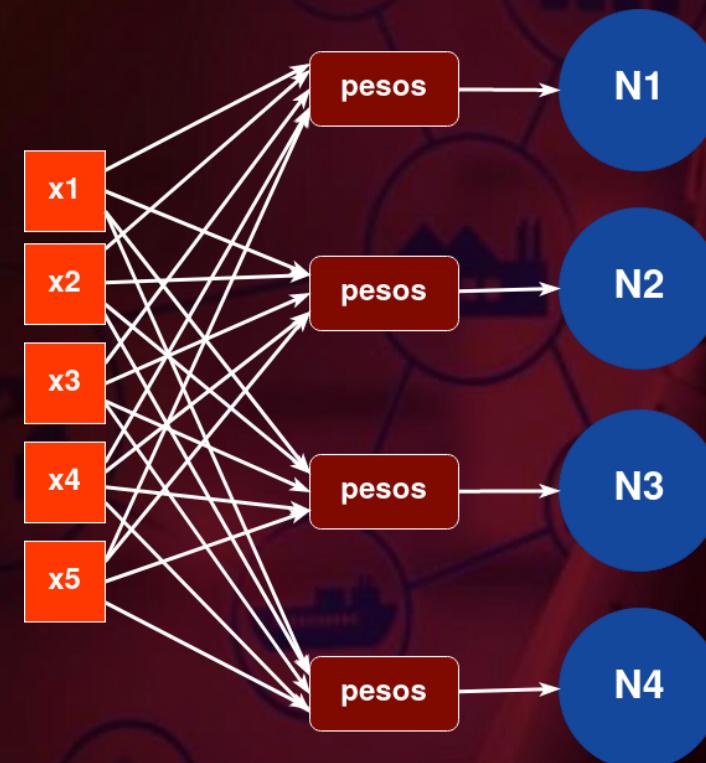
Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

- Redes neurais auto-organizáveis
 - Camada de entrada
 - Matriz de pesos
 - Acionam determinados neurônios em uma única camada



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

SOM: Self-Organizing Maps

- Camada única de neurônios (mapa)
- Para cada objeto:
 - Neurônio vencedor e seus vizinhos têm seus pesos ajustados
 - Objetos distintos acionam regiões distintas do mapa
- Saída: Vetores de referência que representam *clusters*

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

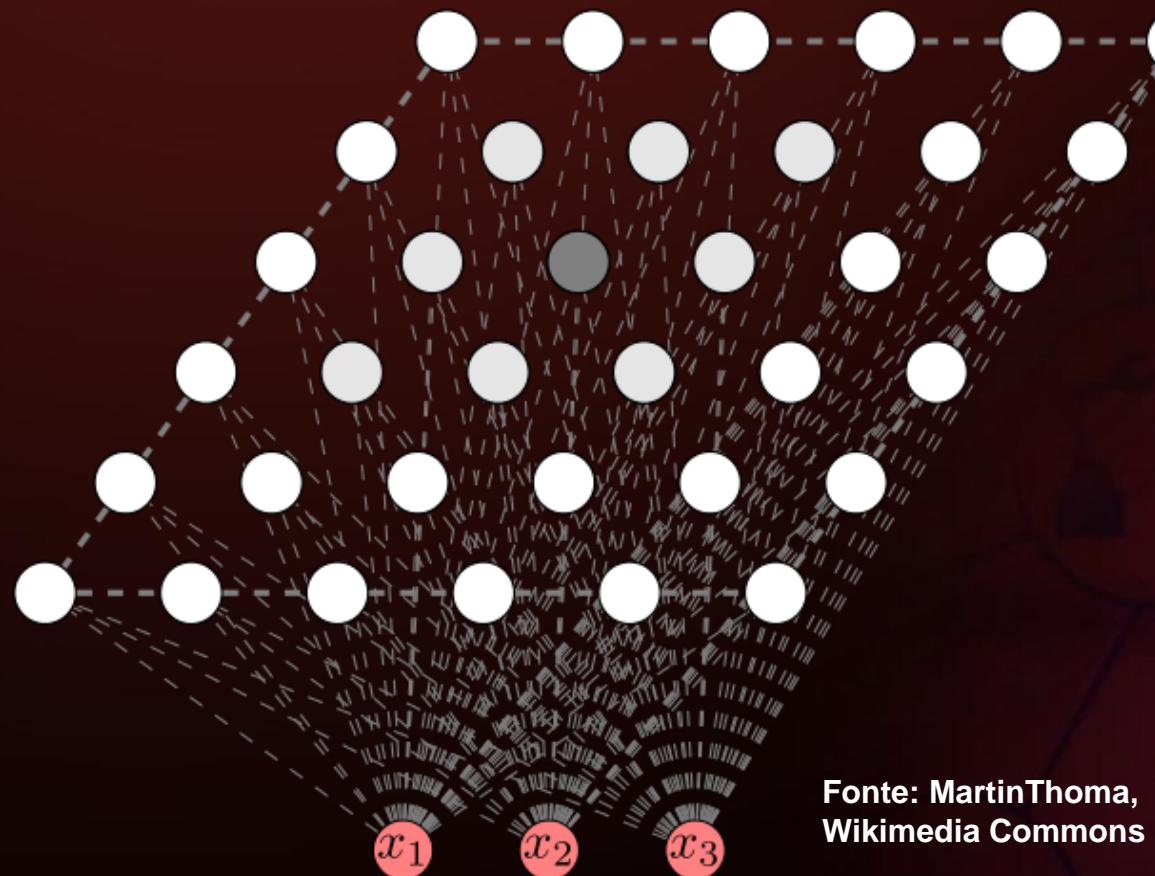
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



SOM: Self-Organizing Maps

Fonte: MartinThoma,
Wikimedia Commons

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

- Subdivisão do espaço dos objetos em grades
- Eficientes para grandes conjuntos de dados
- Funcionam bem para *clusters* arbitrários
- Bons para identificar *outliers*

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid

CLIQUE: CLustering In QUEst

- Determina as grades com base em densidade
 - Particiona o espaço em células
 - Células adjacentes formam *clusters*
 - Dados podem ser agrupados por diferentes perspectivas (atributos)
 - Bom desempenho computacional

TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

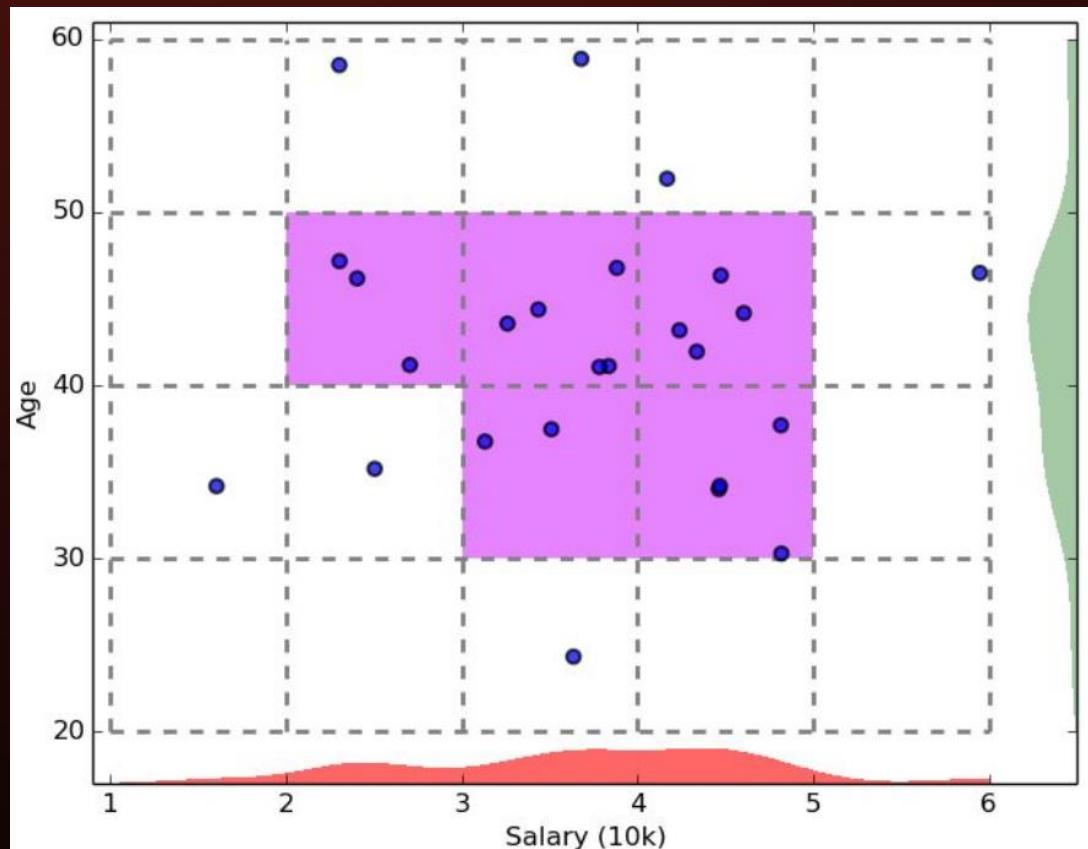
Particionais
baseados em
erro quadrático

Baseados em
densidade

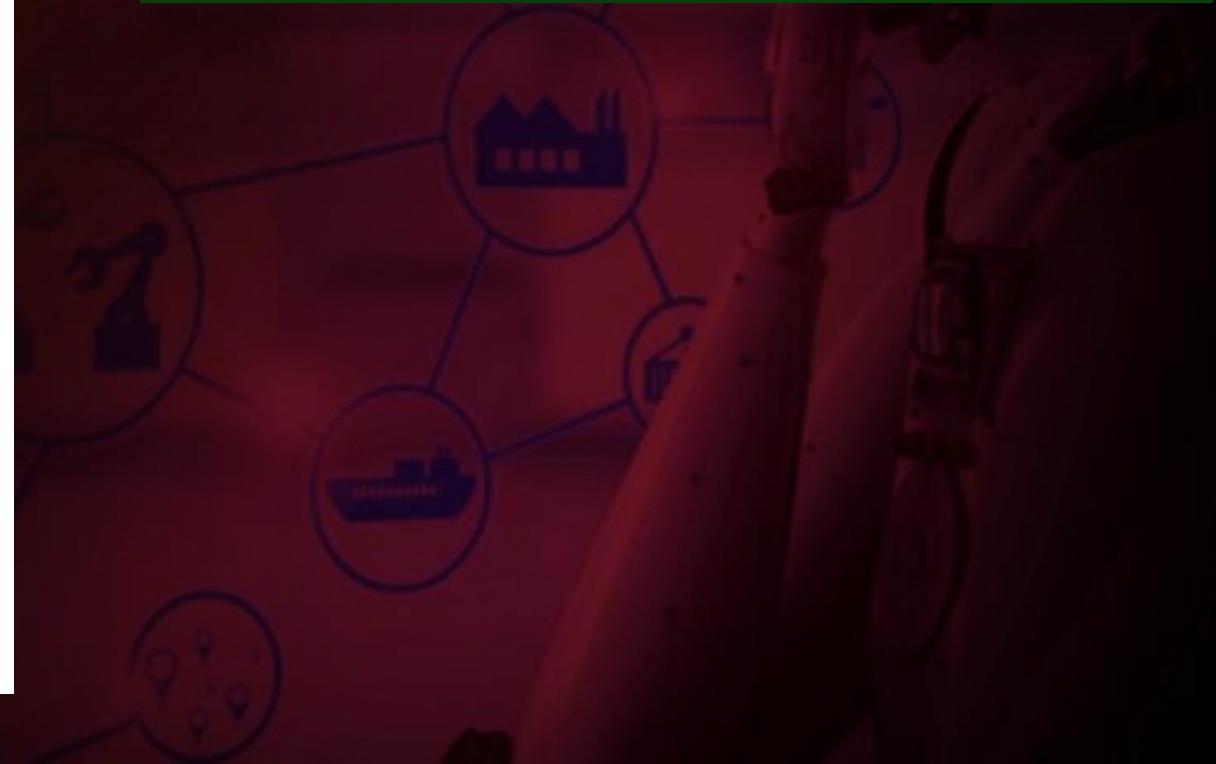
Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



CLIQUE: CLustering In QUEst



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

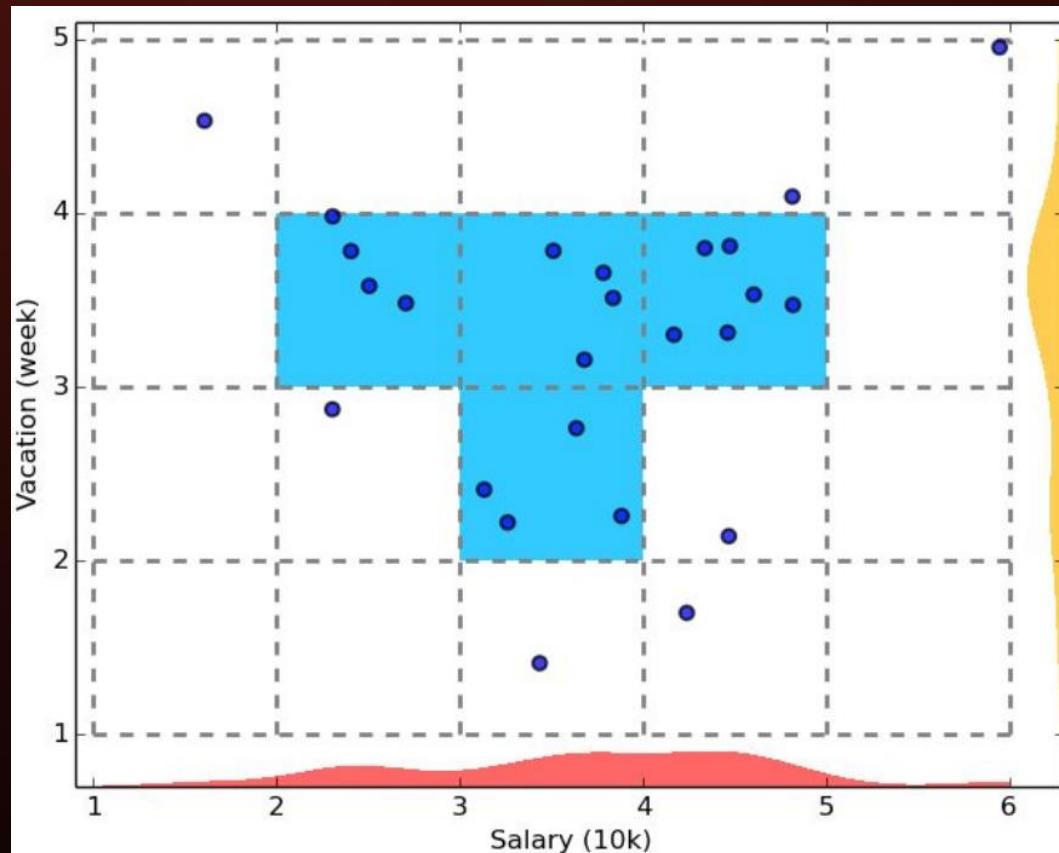
Particionais
baseados em
erro quadrático

Baseados em
densidade

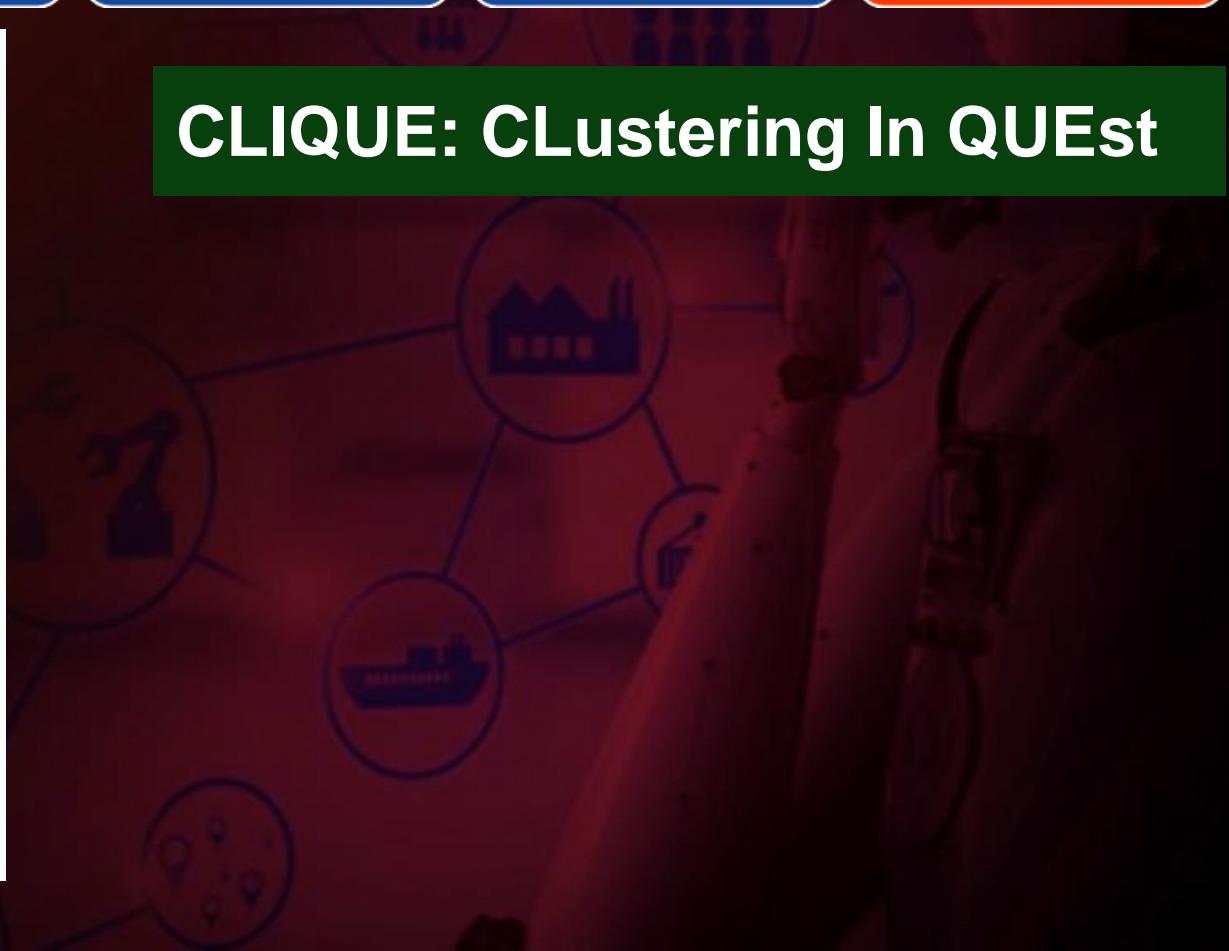
Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



CLIQUE: CLustering In QUEst



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Hierárquicos

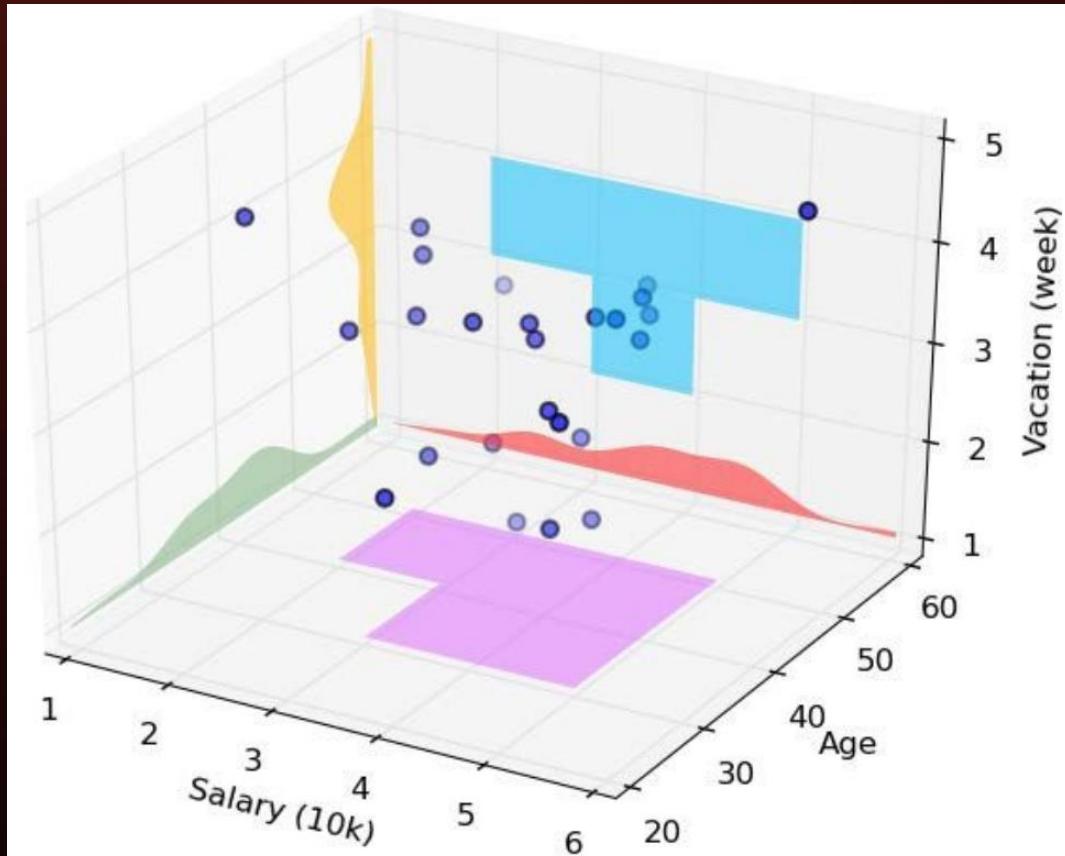
Particionais
baseados em
erro quadrático

Baseados em
densidade

Baseados em
grafos

Baseados em
redes neurais

Baseados em
grid



CLIQUE: CLustering In QUEst

REFERÊNCIA

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