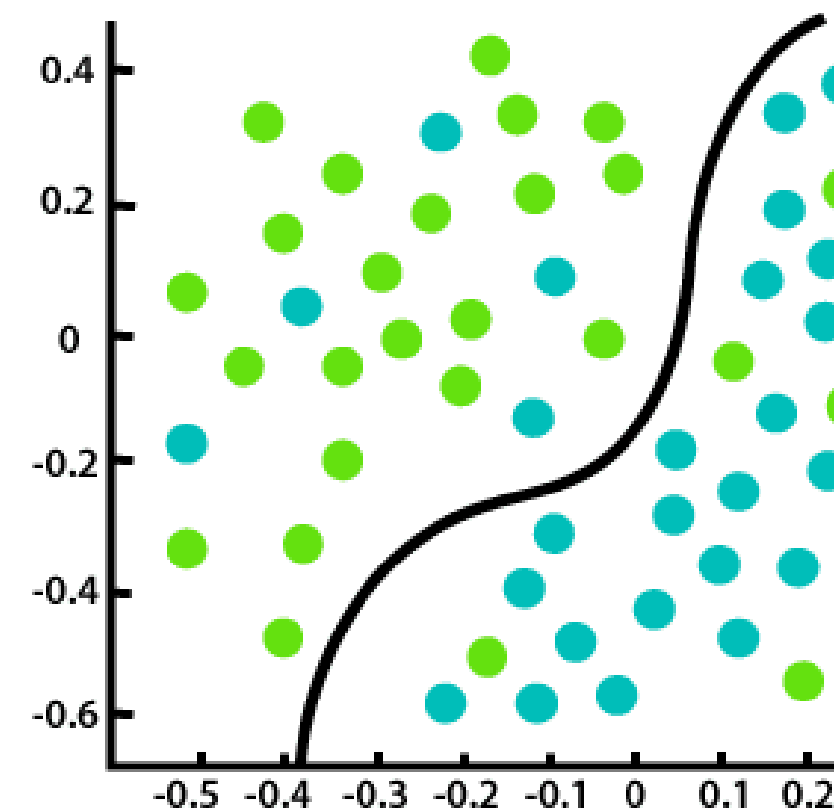


Período 2

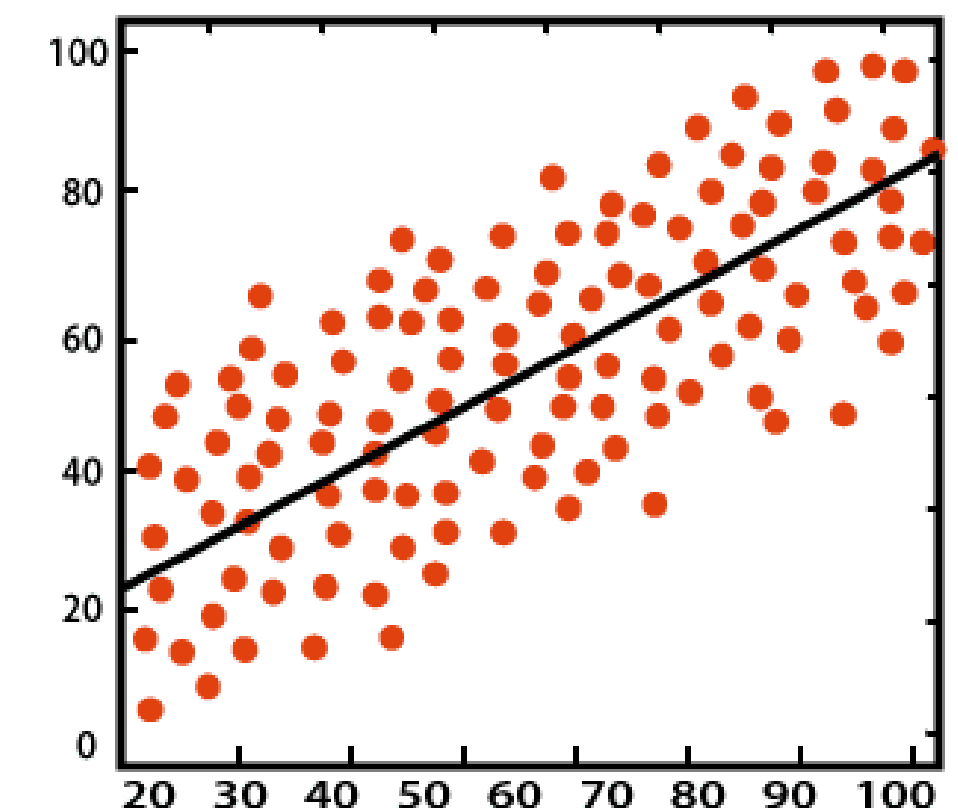
Clusterização

Aprendizado Supervisionado

- Na classificação, se quer prever a classe ou categoria de uma observação.
- Na regressão, se quer prever um valor contínuo, como o preço de um produto ou a temperatura



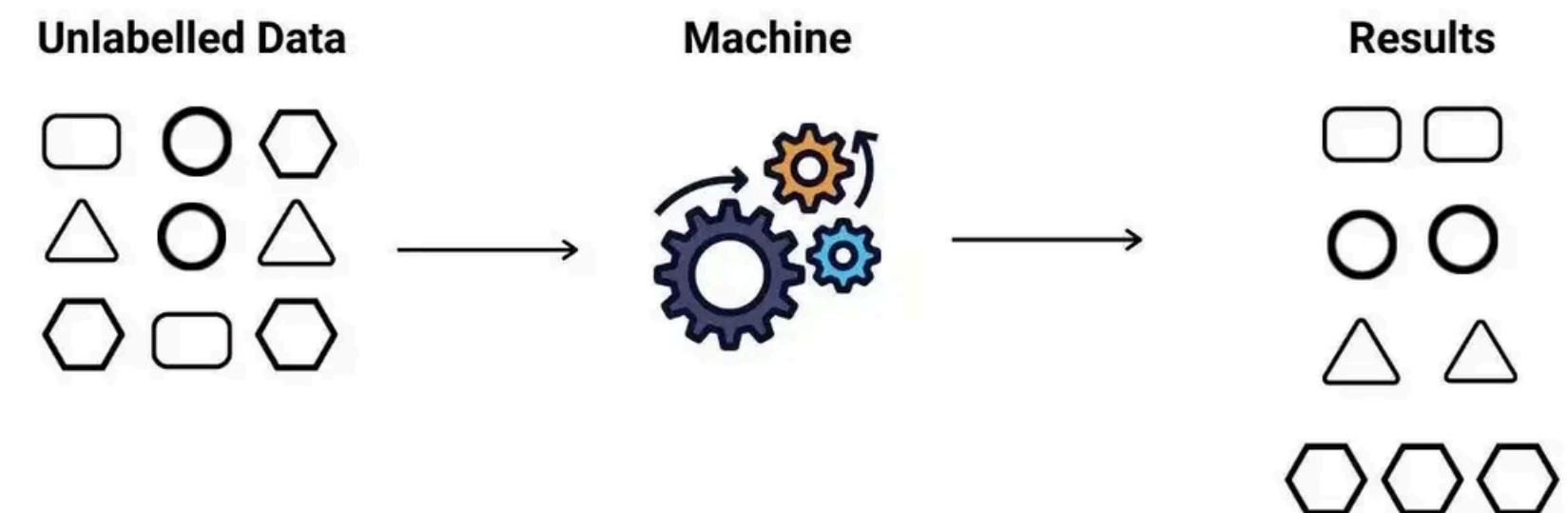
Classificação



Regressão

Aprendizado Não Supervisionado

- Usa algoritmos de machine learning para analisar e agrupar conjuntos de dados não rotulados. Esses algoritmos descobrem padrões ocultos ou agrupamentos de dados sem a necessidade de intervenção humana.



Introdução

O que é Clusterização?

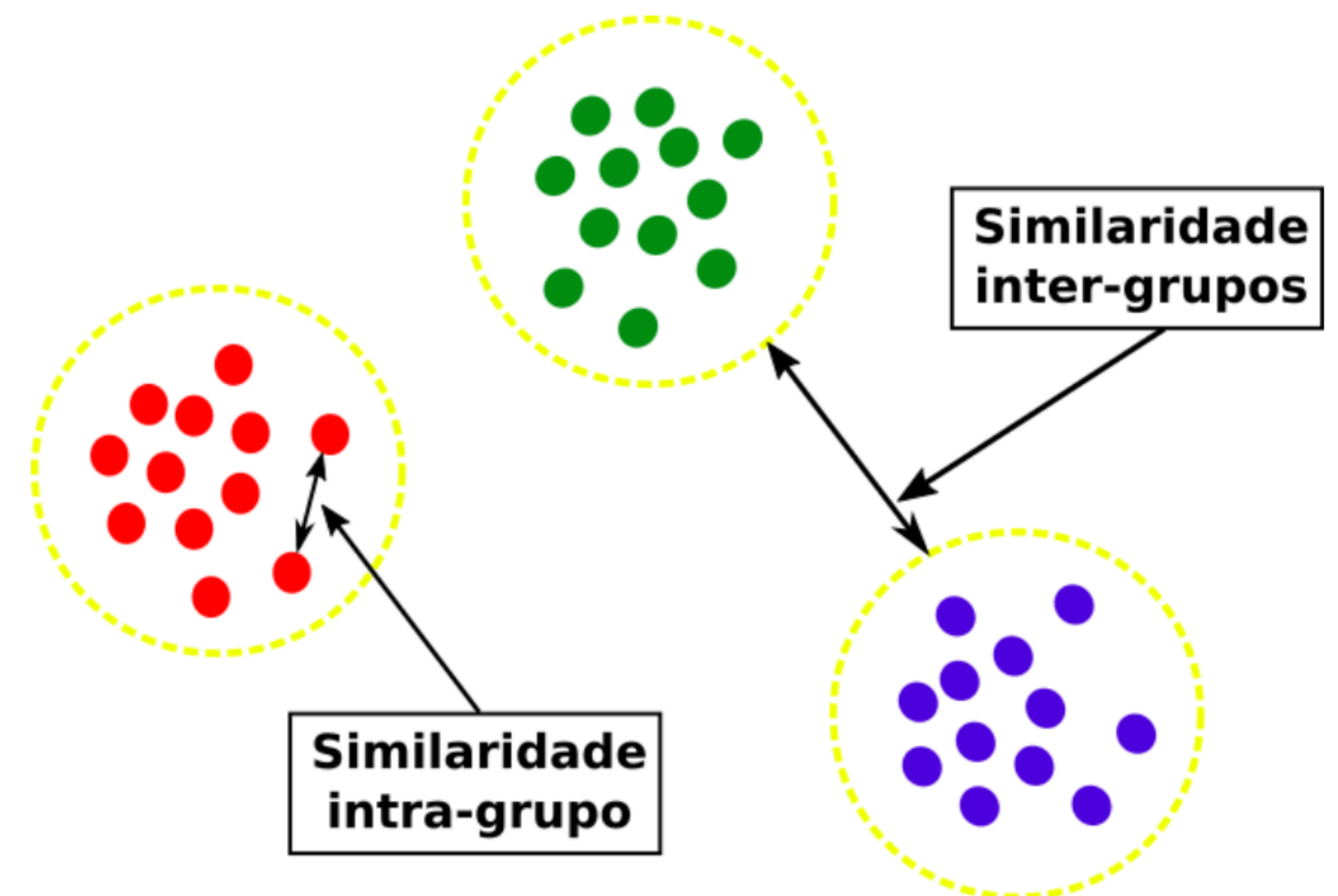
Clusterização (ou clustering) é uma técnica de aprendizado não supervisionado usada para agrupar dados semelhantes em grupos chamados clusters, sem a necessidade de rótulos prévios.

Aplicações:

- Encontrar padrões e estrutura nos dados
- Identificar grupos naturais dentro do conjunto de dados
- Melhorar a organização e análise das informações

Objetivo

Os grupos são formados de maneira a maximizar a similaridade entre os elementos de um grupo (**similaridade intra-grupo**) e minimizar a similaridade entre elementos de grupos diferentes (**similaridade inter-grupos**)



Por que usar Clusterização?

Identificação automática de padrões sem supervisão

Redução de dimensionalidade e simplificação de dados

Auxilia na tomada de decisão baseada em grupos

Por que usar Clusterização?

Identificação automática de padrões sem supervisão

Redução de dimensionalidade e simplificação de dados

Auxilia na tomada de decisão baseada em grupos

Cliente	Renda Mensal (R\$)	Gastos Menais (R\$)
João	10.000	6.000
Maria	3.000	1.000
Andressa	5.000	2.500
Pedro	2.000	500

Por que usar Clusterização?

Identificação automática de padrões sem supervisão

Redução de dimensionalidade e simplificação de dados

Auxilia na tomada de decisão baseada em grupos

Cliente	Renda Mensal (R\$)	Gastos Menais (R\$)	Cluster (Exemplo)
João	10.000	6.000	Gastador
Maria	3.000	1.000	Econômico
Andressa	5.000	2.500	Moderada
Pedro	2.000	500	Econômico

Funcionamento do K-means

1. Inicialização:

- Escolha K centróides aleatórios do conjunto de dados.

2. Atribuição:

- Atribua cada ponto de dado ao centróide mais próximo.

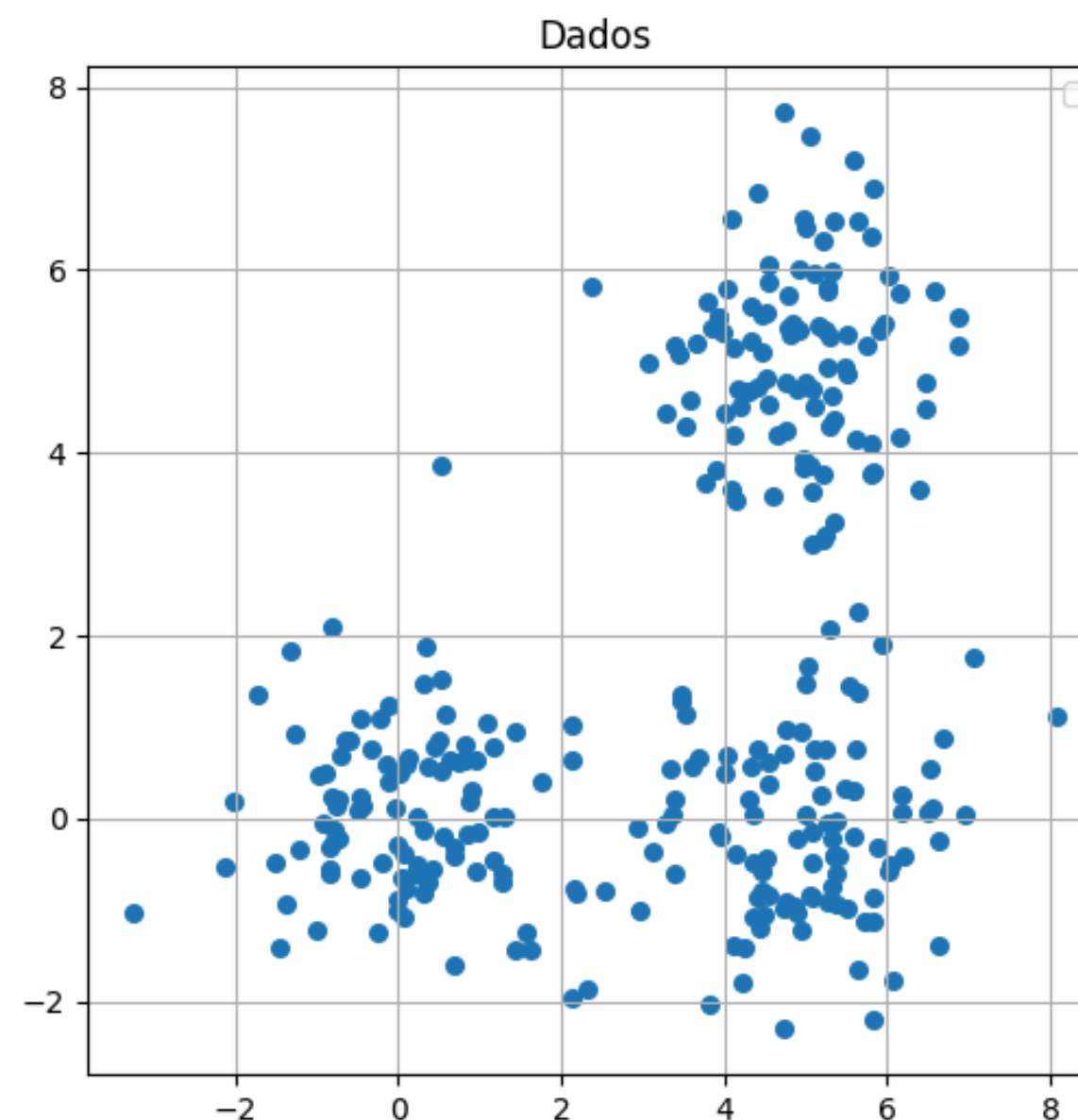
3. Recalcular Centroides:

- Recalcule os centroides como a média dos pontos de cada grupo.

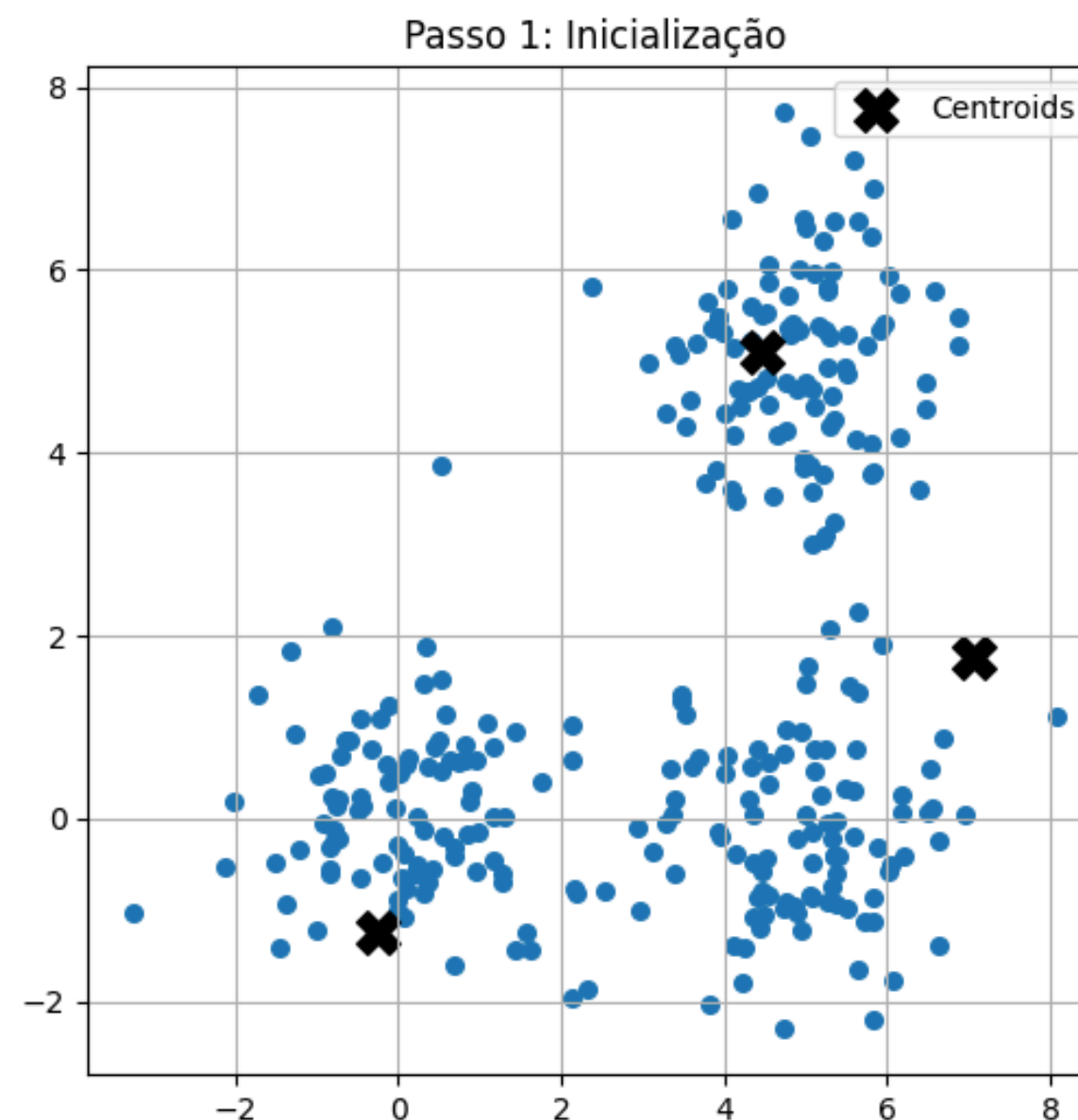
4. Repetir:

- Repita os passos 2 e 3 até que os centróides não mudem mais (convergência).

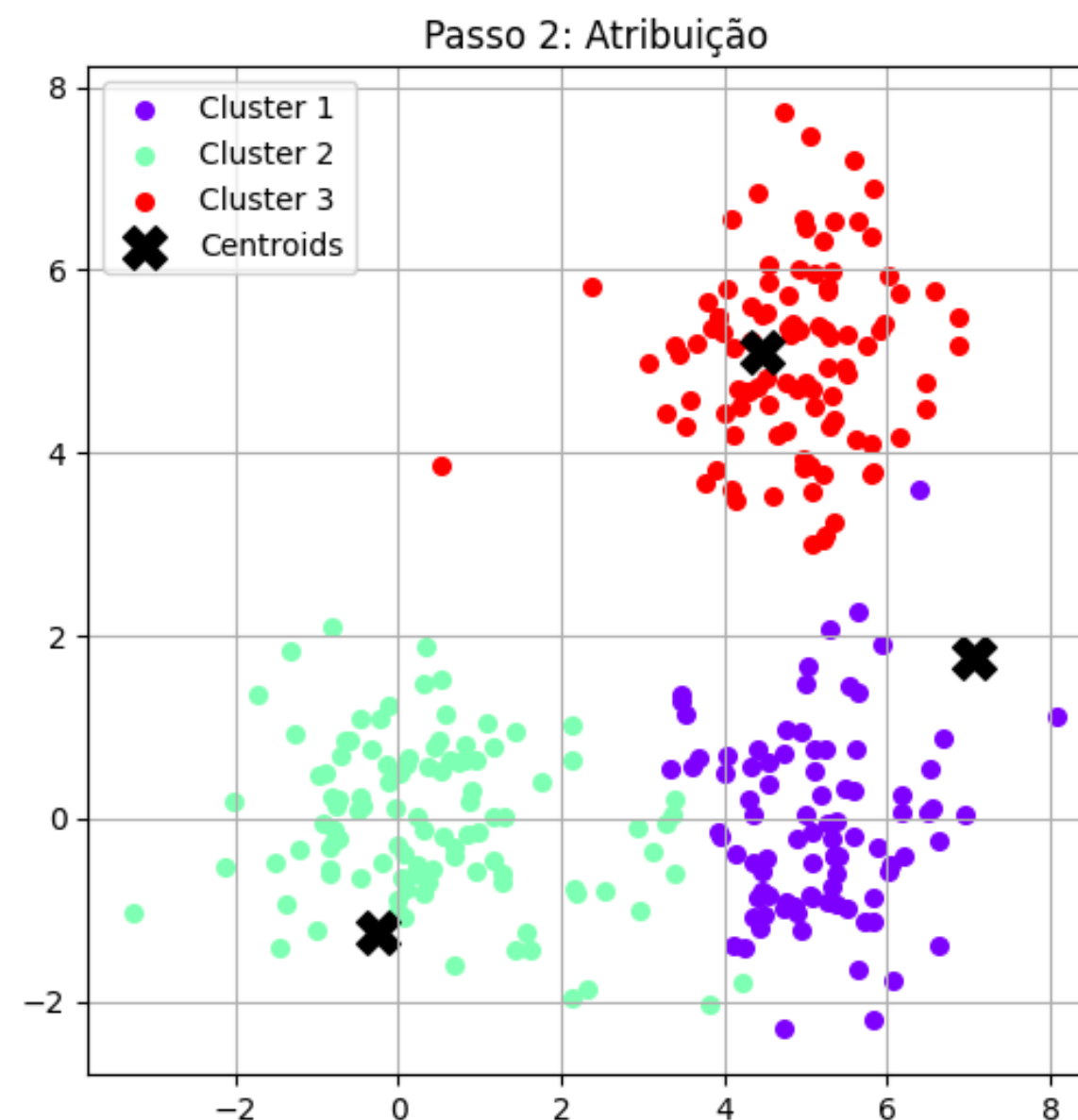
Conjunto de Dados



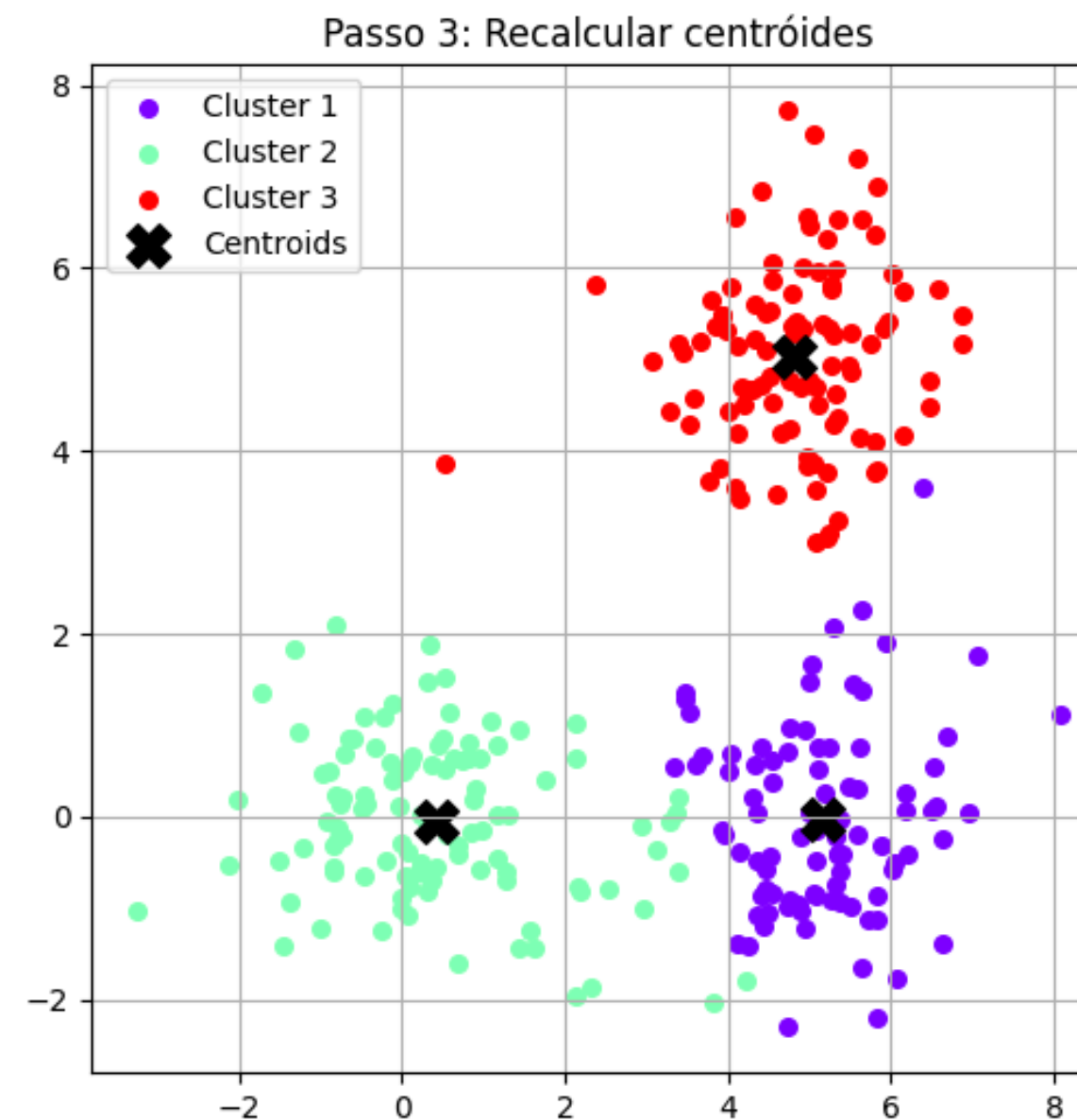
Inicialização



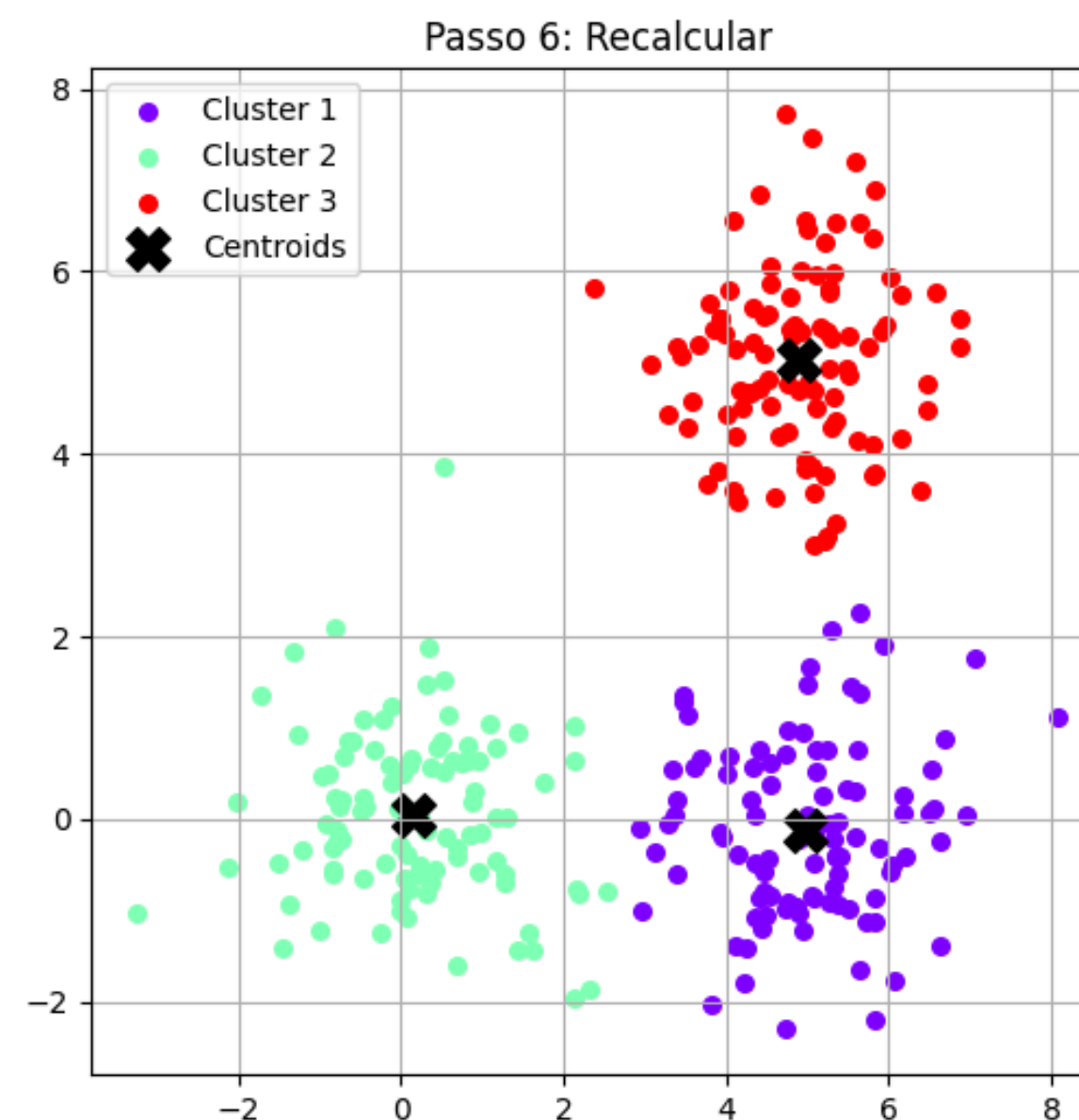
Atribuição



Recalcular os Centróides



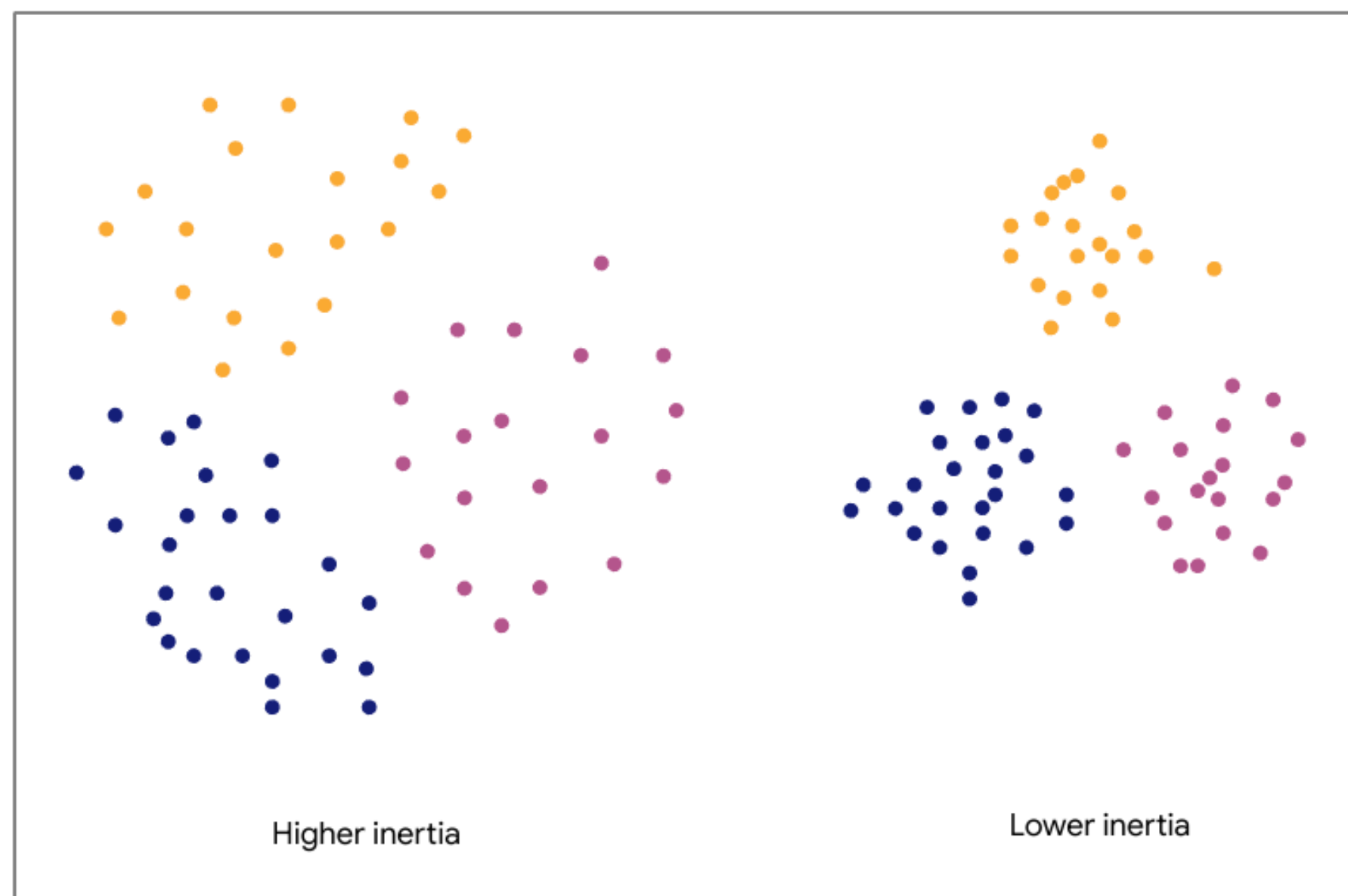
Resultado Final



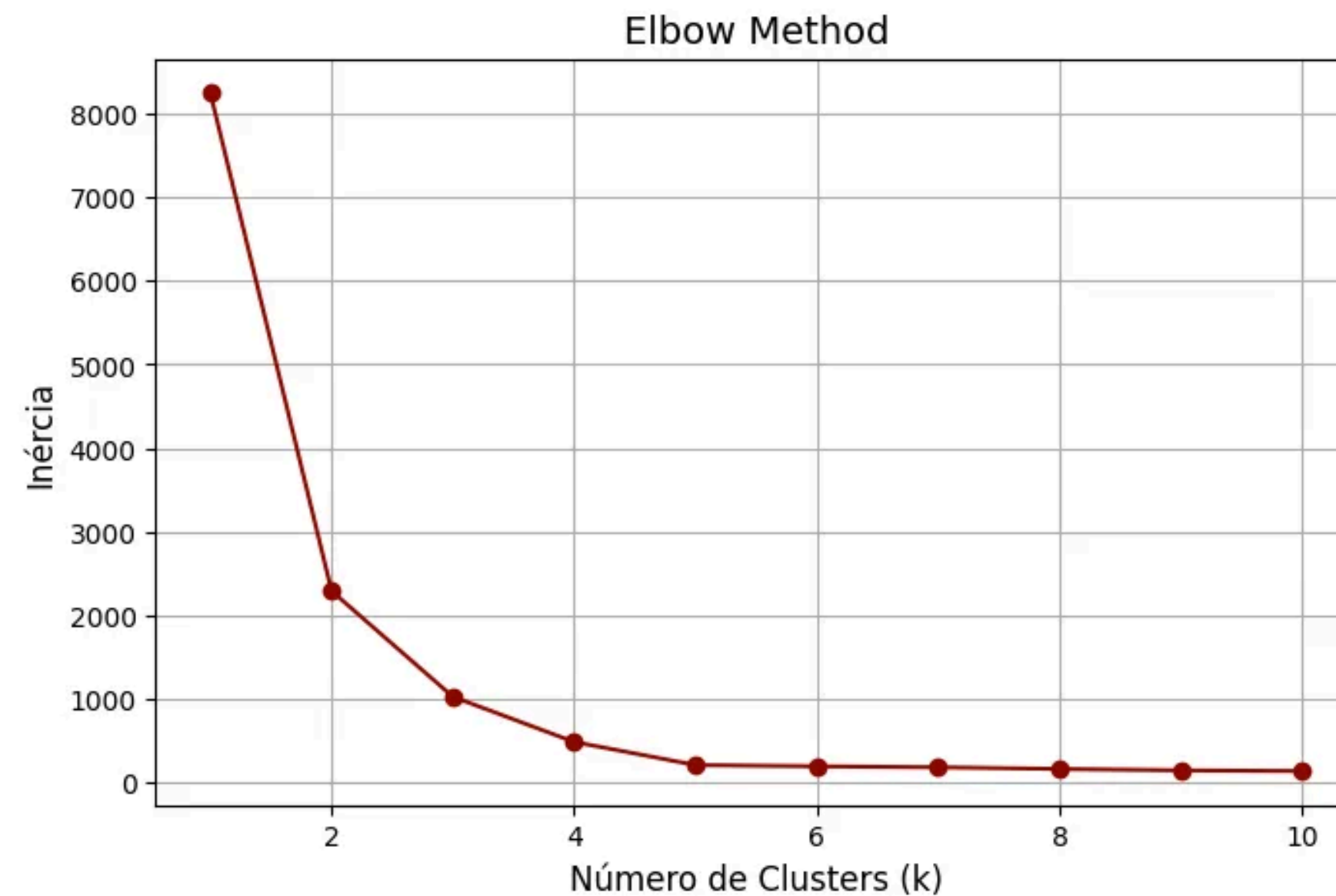
Inércia

$$\text{Inertia} = \sum_{i=1}^n (x_i - C_k)^2$$

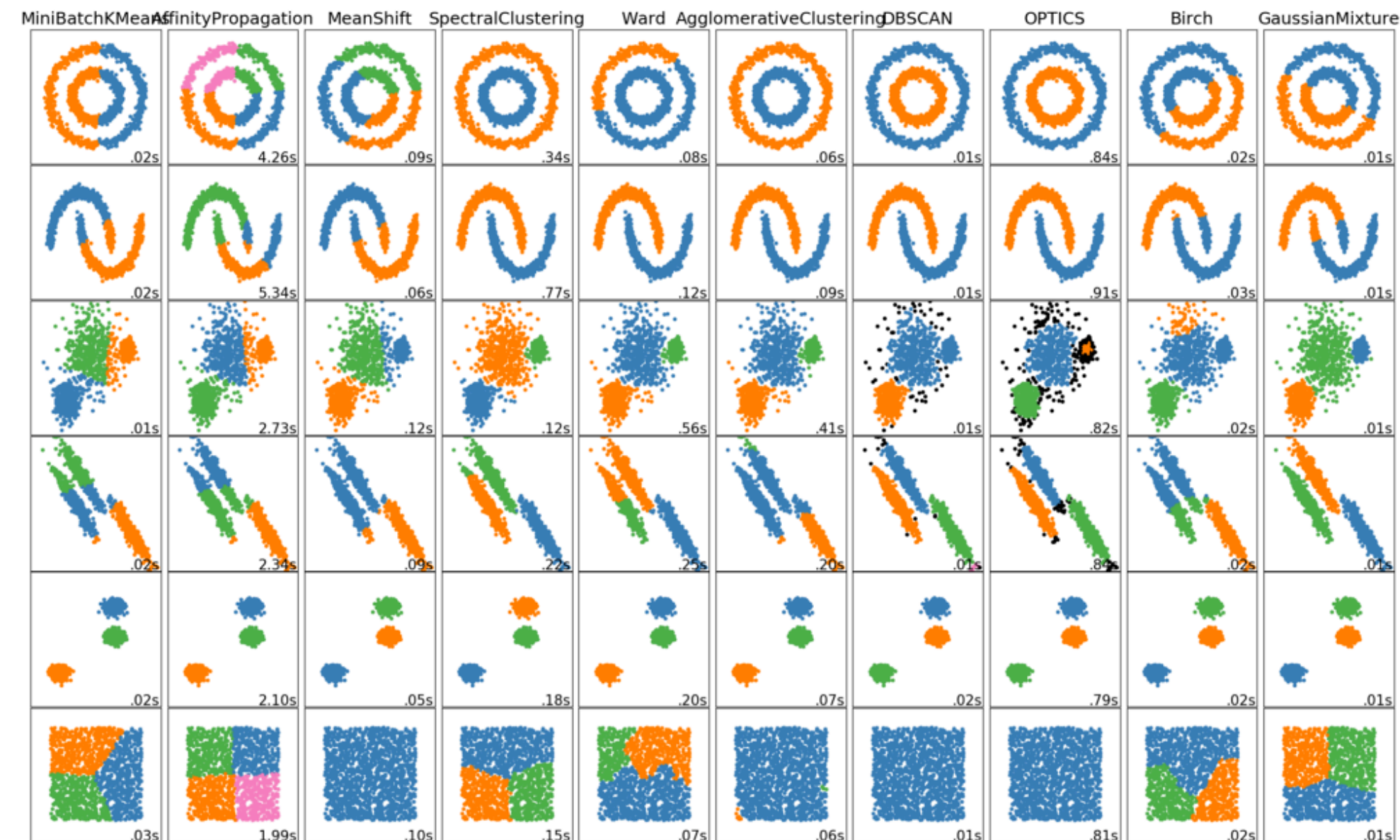
Inércia



Descobrir o Número de Clusters



Diferentes Modelos



Obrigado!