

Sistemas Inteligentes

Clustering: K-means

Profa: Deborah Magalhães



Clusterização (Agrupamento)

ESPORTES



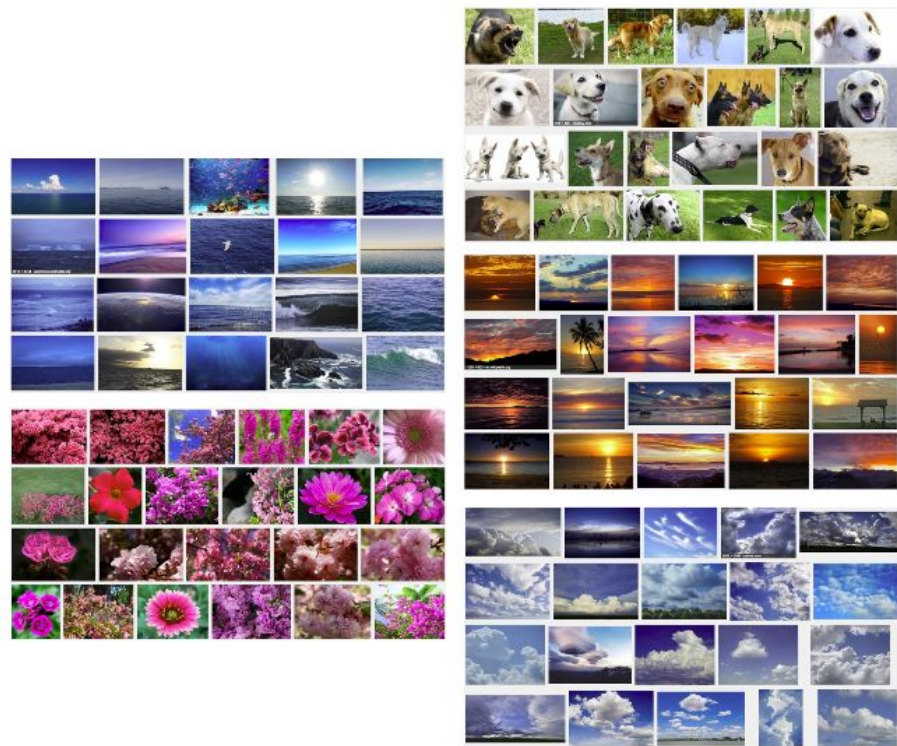
ENTRETENIMENTO



NOTÍCIAS



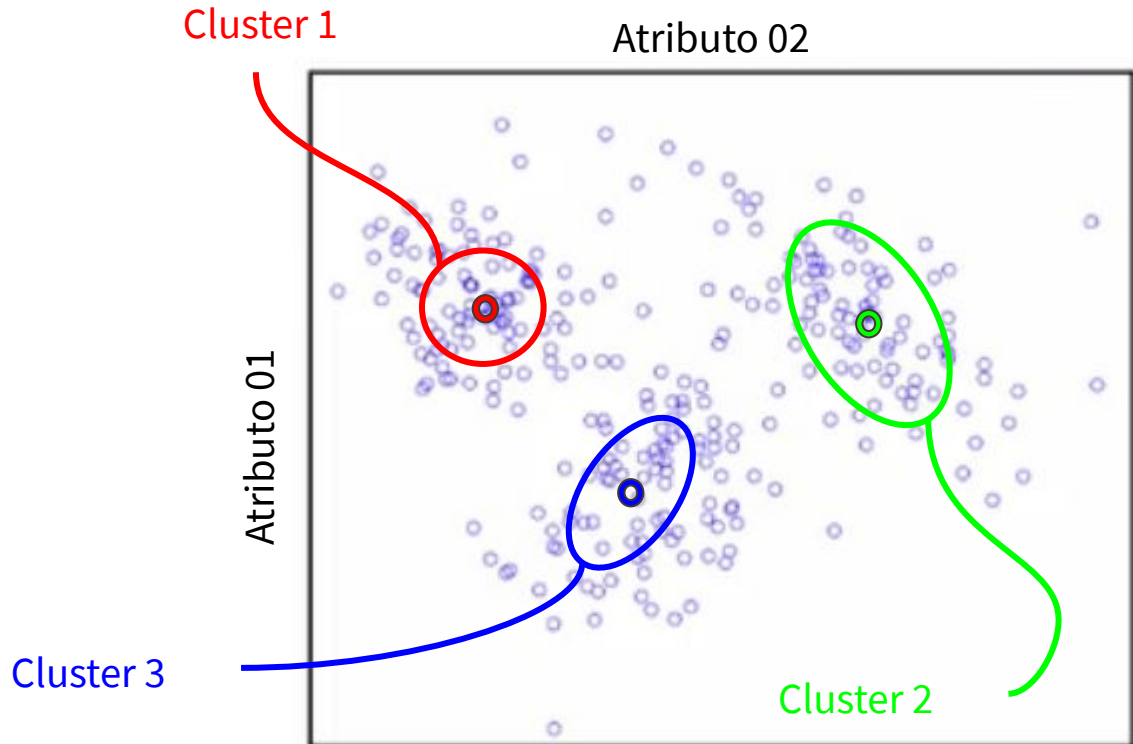
CIÊNCIA



Clustering

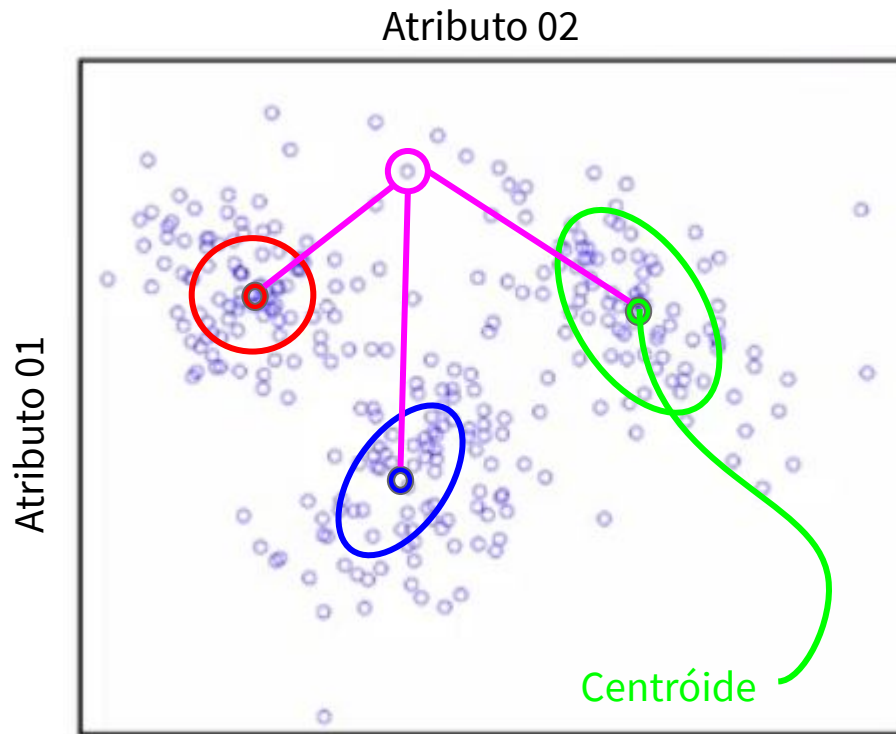
- Entrada: vetores de atributos
- Saída: rótulos dos clusters

Aprendizagem **NÃO**
Supervisionada



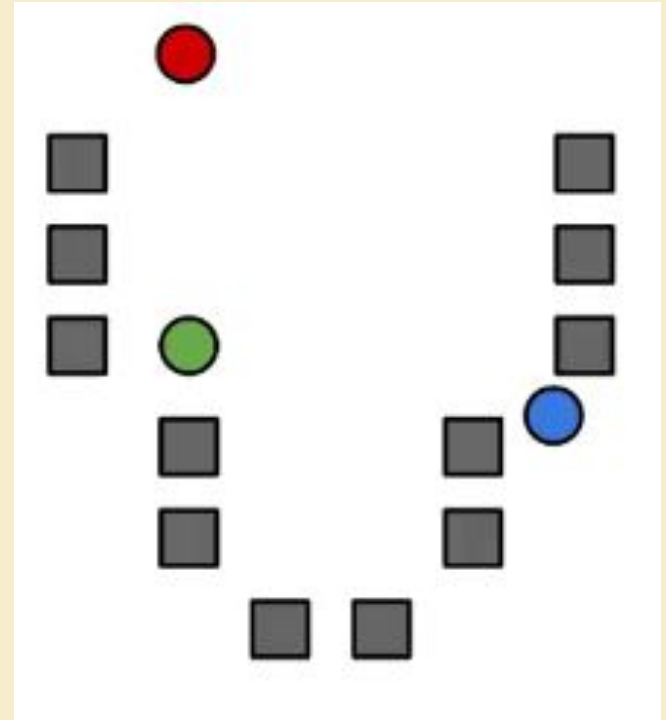
Clustering

- Quantos clusters devem ser formados?
- Como é definido que uma amostra pertence a um ponto?



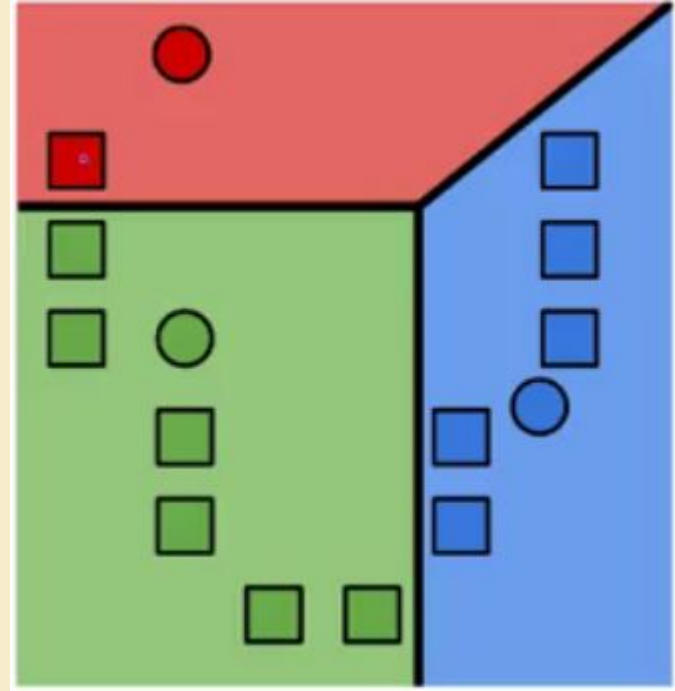
Algoritmo K-means

1. Definir o número de clusters (K)
2. Inicializar os centróides dos cluster (randomicamente)



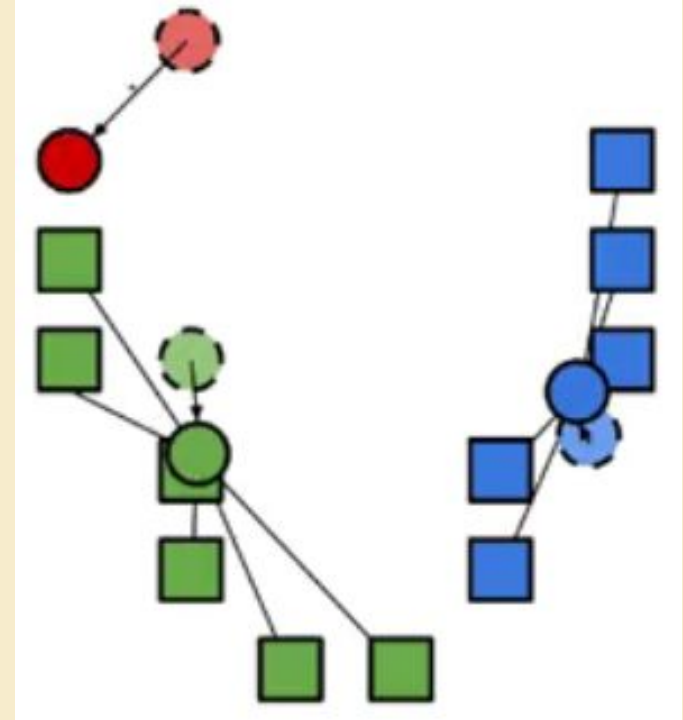
Algoritmo K-means

1. Definir o número de clusters (K)
2. Inicializar os centróides dos cluster (randomicamente)
3. Associar cada amostra ao centróide mais próximo



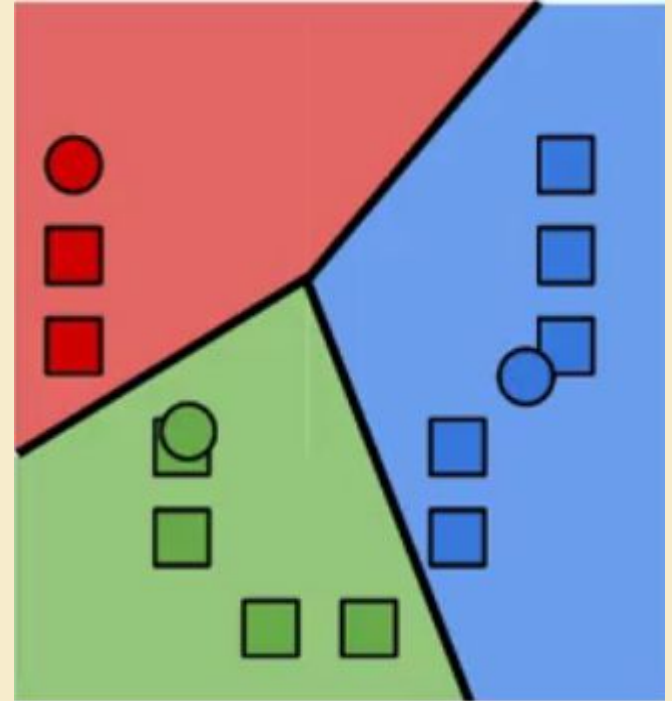
Algoritmo K-means

1. Definir o número de clusters (K)
2. Inicializar os centróides dos cluster (randomicamente)
3. Associar cada amostra ao centróide mais próximo
4. Calcular o novo centróide para cada cluster



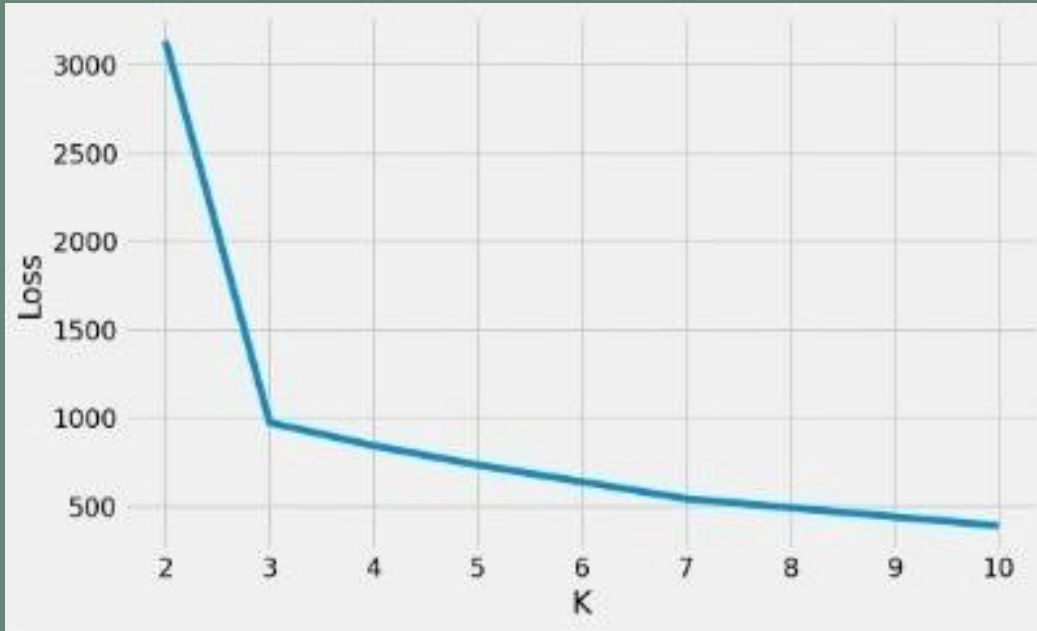
Algoritmo K-means

1. Definir o número de clusters (K)
2. Inicializar os centróides dos cluster (randomicamente)
3. Associar cada amostra ao centróide mais próximo
4. Calcular o novo centróide para cada cluster
5. Repetir os passos 4-3 até convergir



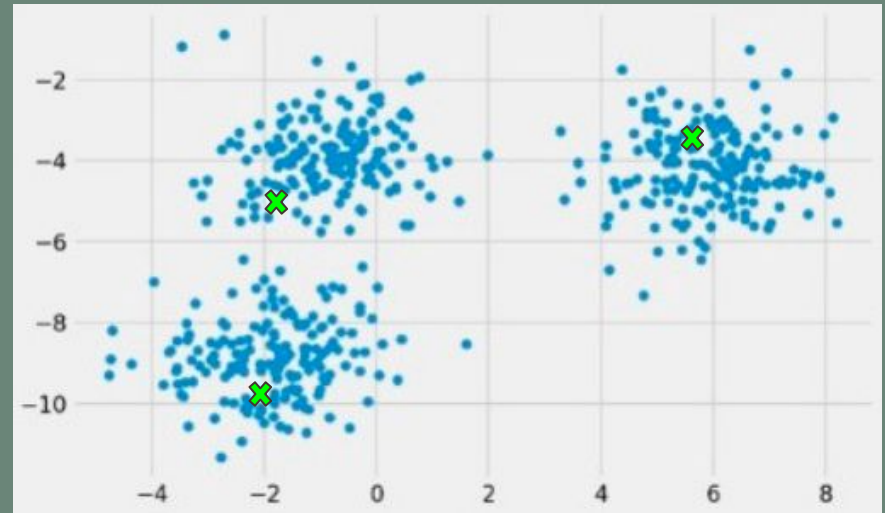
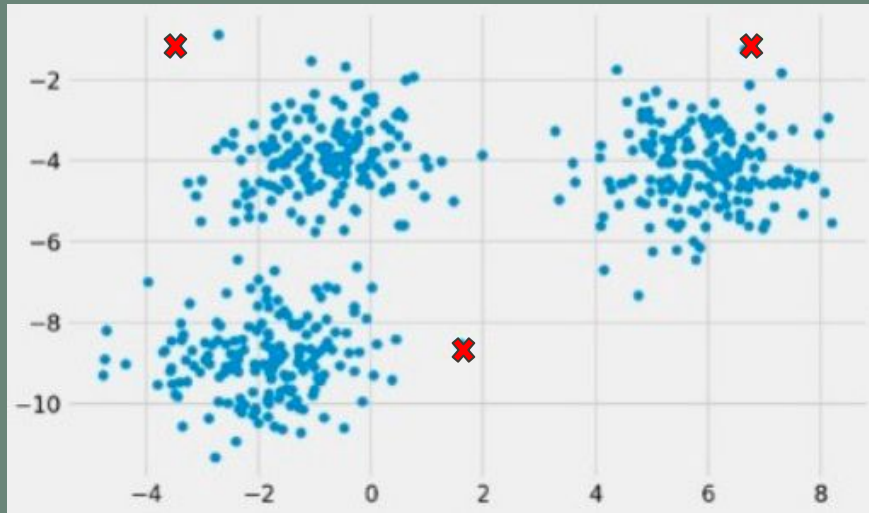
Desafios da parametrização do K-means

Desafio: definir o número de clusters (K)

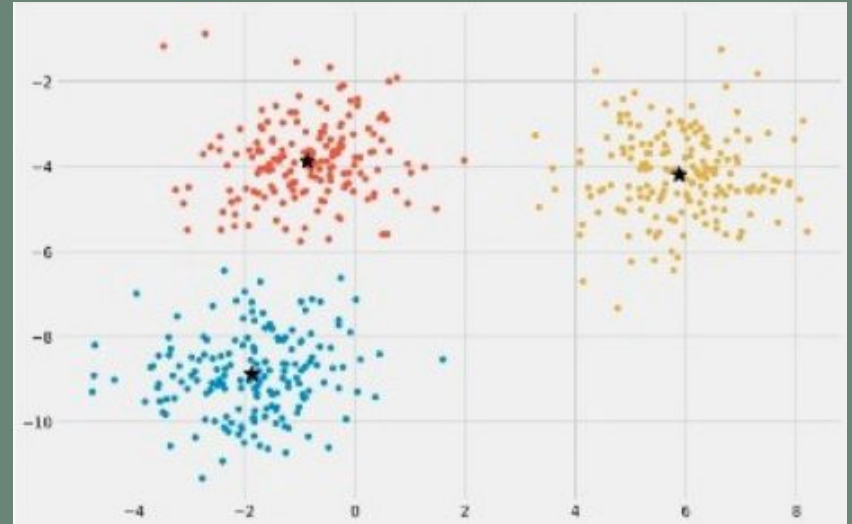
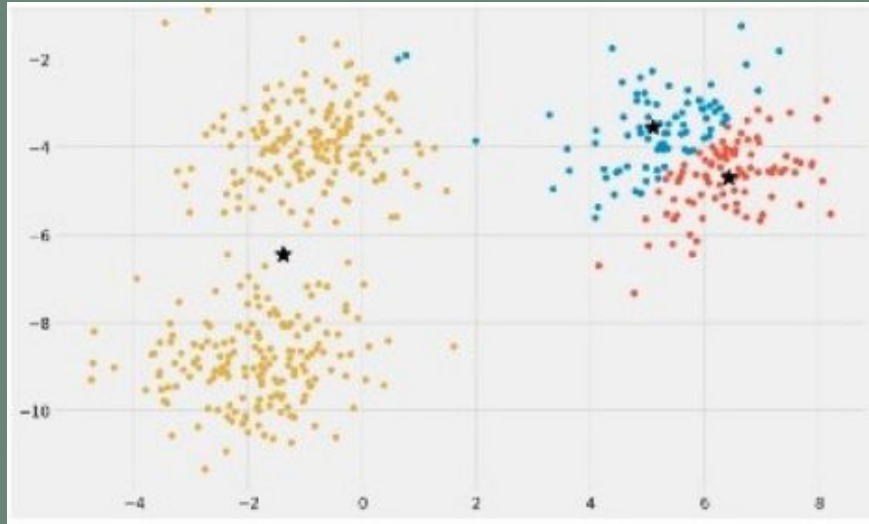


$$loss = \underset{K}{\operatorname{argmin}} \sum_{i=1}^K \sum_{x \in S_i} \|x - C_i\|^2$$

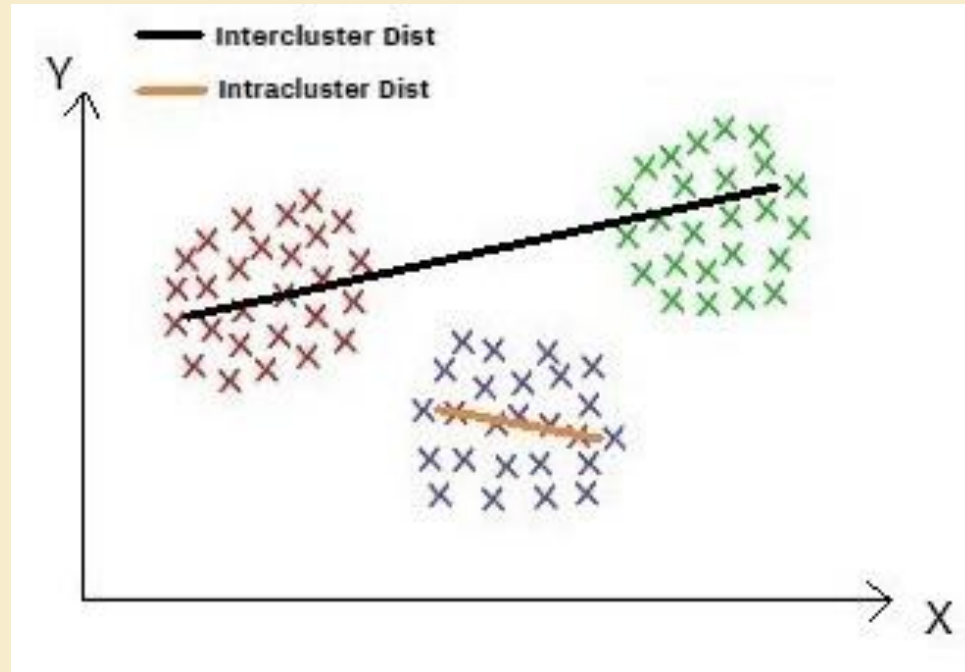
Desafio: inicialização dos centróides



Desafio: inicialização dos centróides

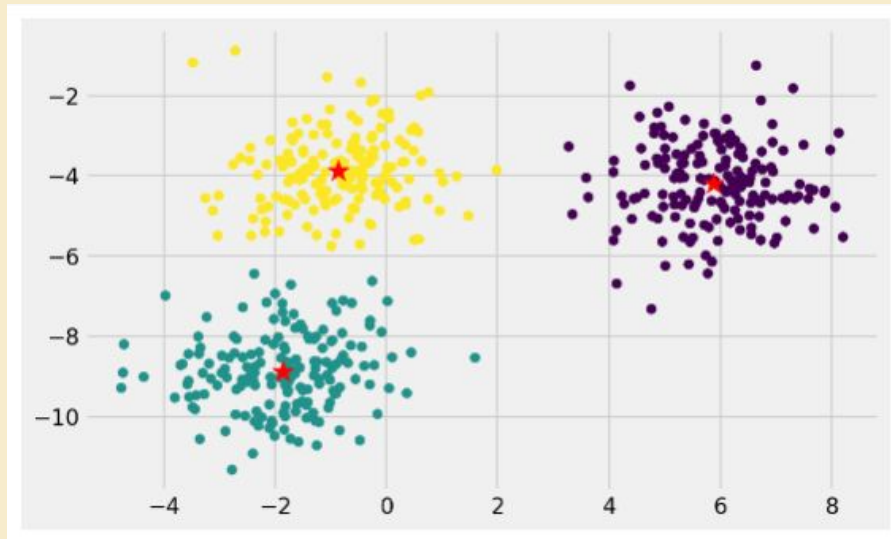


Abordagem I: repetição da seleção randômica



Abordagem II: k-means++

1. Definir o primeiro centróide aleatoriamente
2. Calcular a distância de todos as amostras para o centróide selecionado
3. Definir o ponto mais distante como o novo centróide
4. Repetir 2-3 até que todos os k centróides sejam definidos





Muito Obrigada!

Se você tiver qualquer dúvida ou sugestão:

- deborah.vm@ufpi.edu.br

