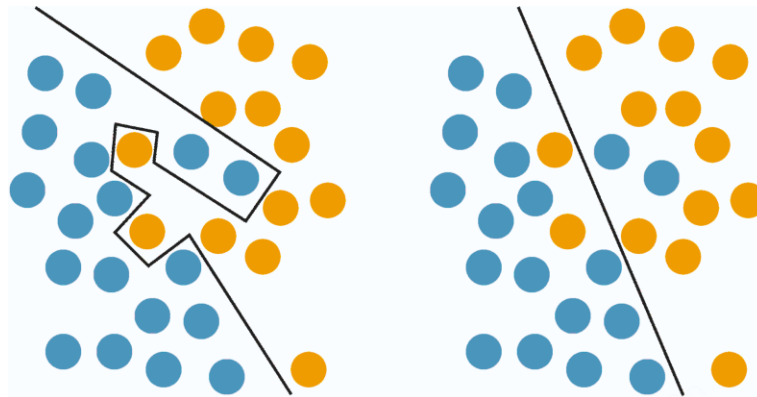


Algoritmo K-means



Prof. Dr. Diego Bruno

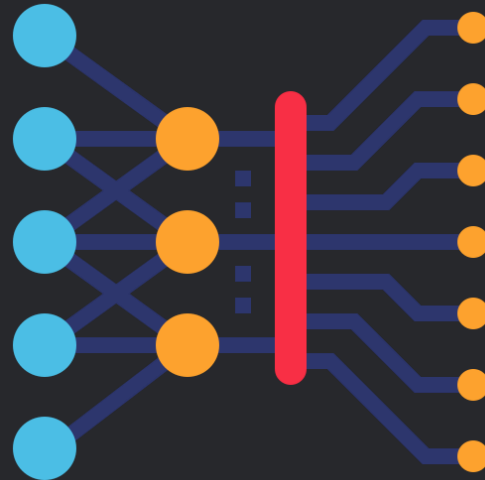
Education Tech Lead na DIO

Doutor em Robótica e *Machine Learning* pelo ICMC-USP

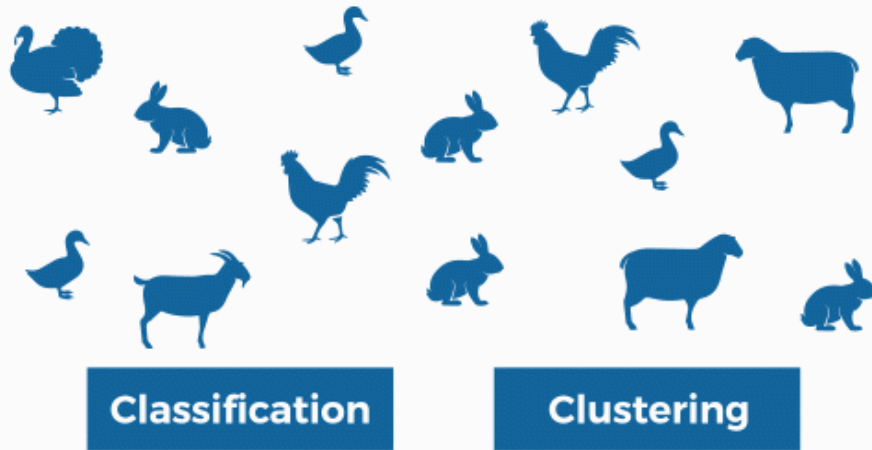
K-means

Prof. Dr. Diego Bruno

Machine Learning



Classificação x Clustering



Classificação x Clustering

Clustering



C1

C2

dataaspirant.com

dataaspirant.com



Class 1



Class 3



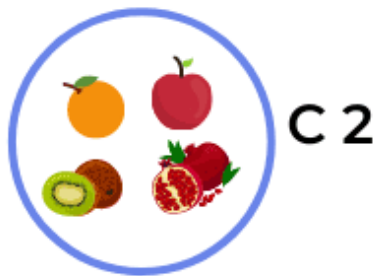
Class 2



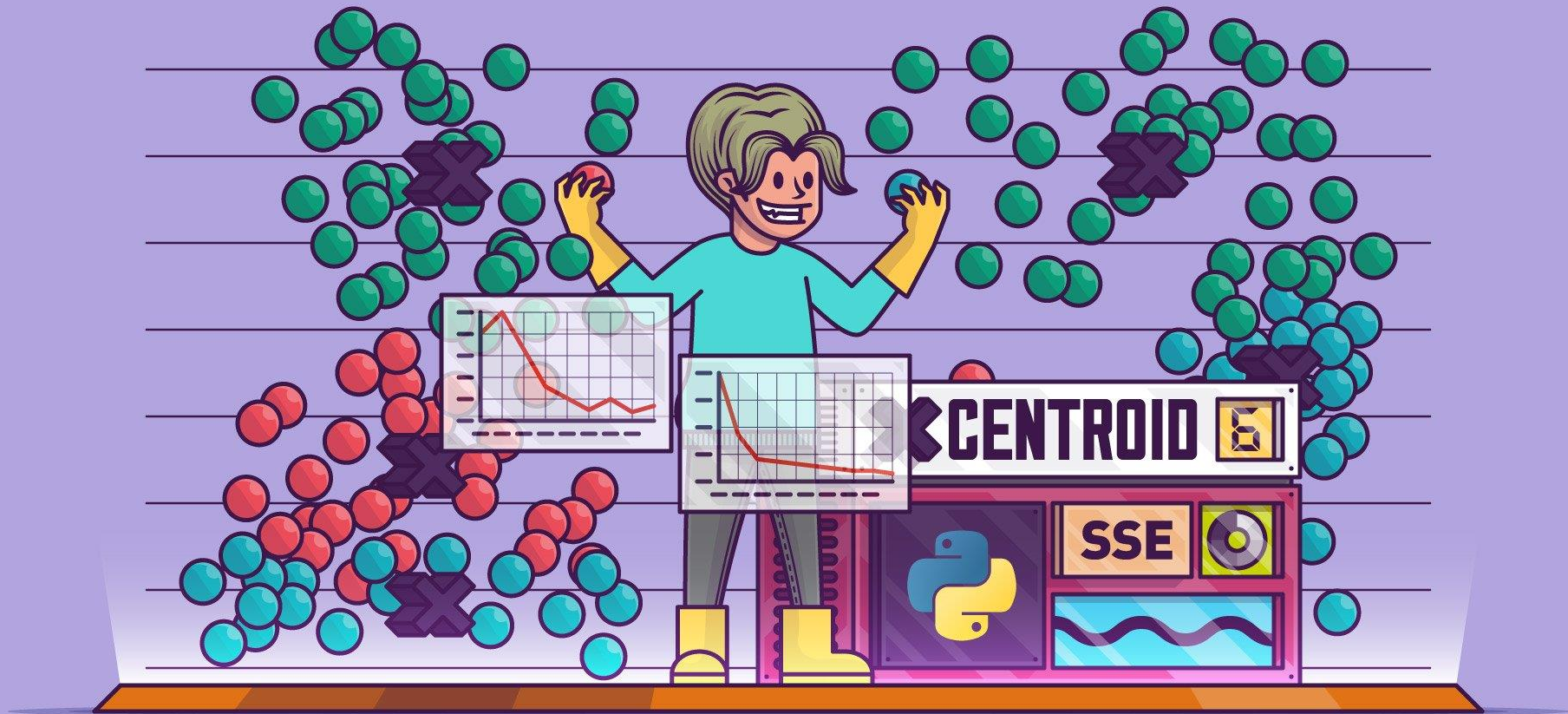
Class 4

Classification

Classificação x Clustering



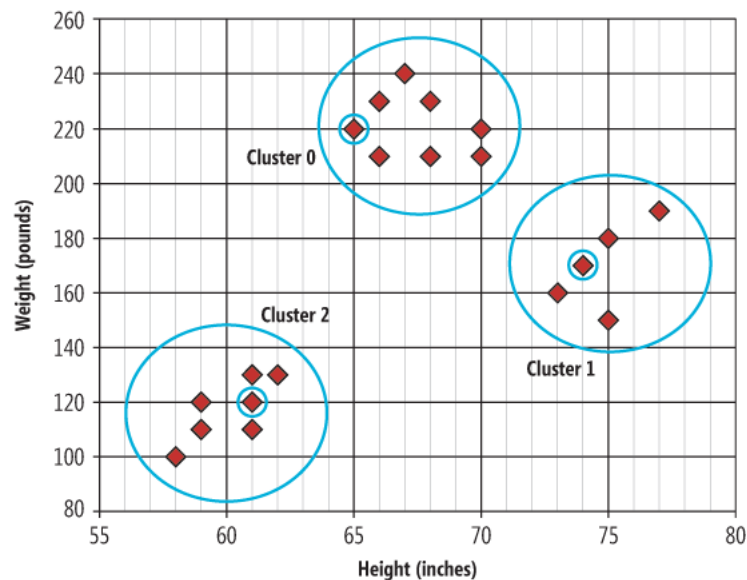
K Means Clustering



Real Python

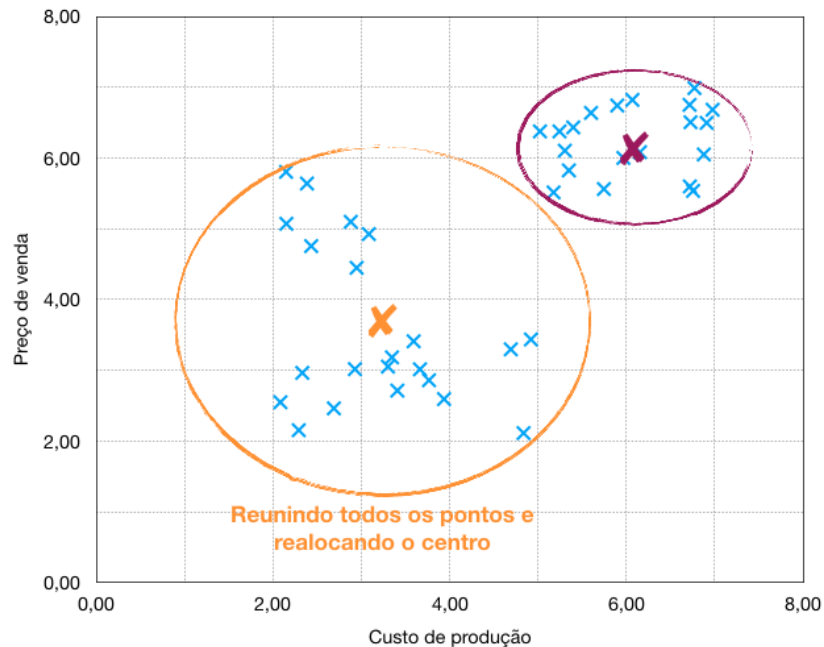
Algoritmo K-means

1. Primeiro, preciso definir um 'K', ou seja, um número de clusters.
2. Depois, preciso definir, aleatoriamente, um centroide para cada cluster.
3. O próximo passo é calcular, para cada ponto, o centroide de menor distância.

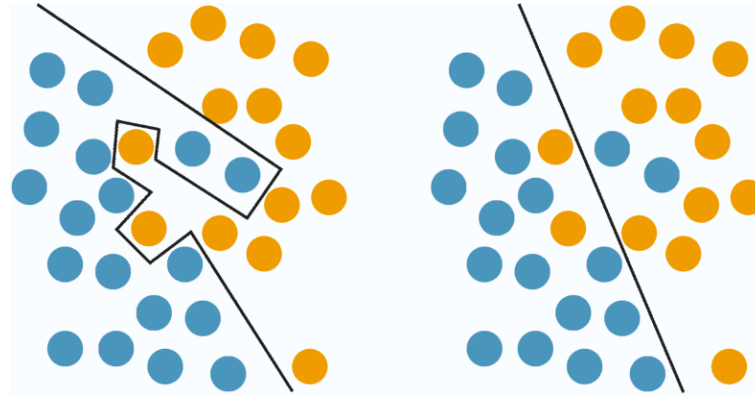
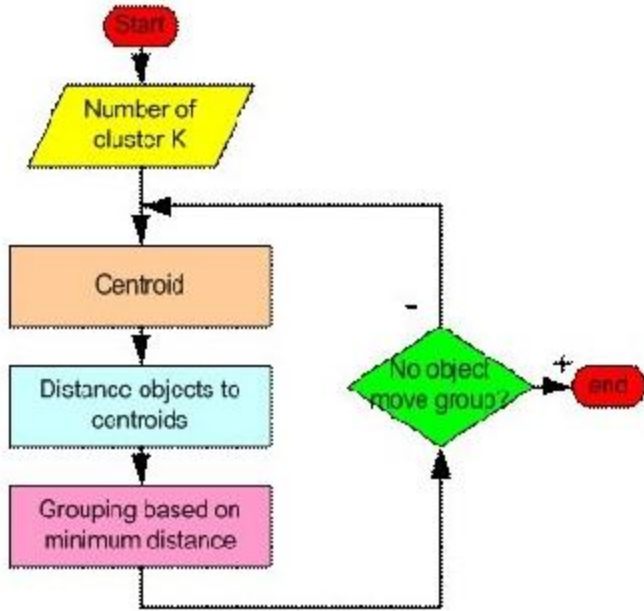


Algoritmo K-means

- 4. Agora, devo reposicionar o centroide. A nova posição do centroide deve ser a média da posição de todos os pontos do cluster.
- 5. Os dois últimos passos são repetidos, iterativamente, até obtermos a posição ideal dos centroides.

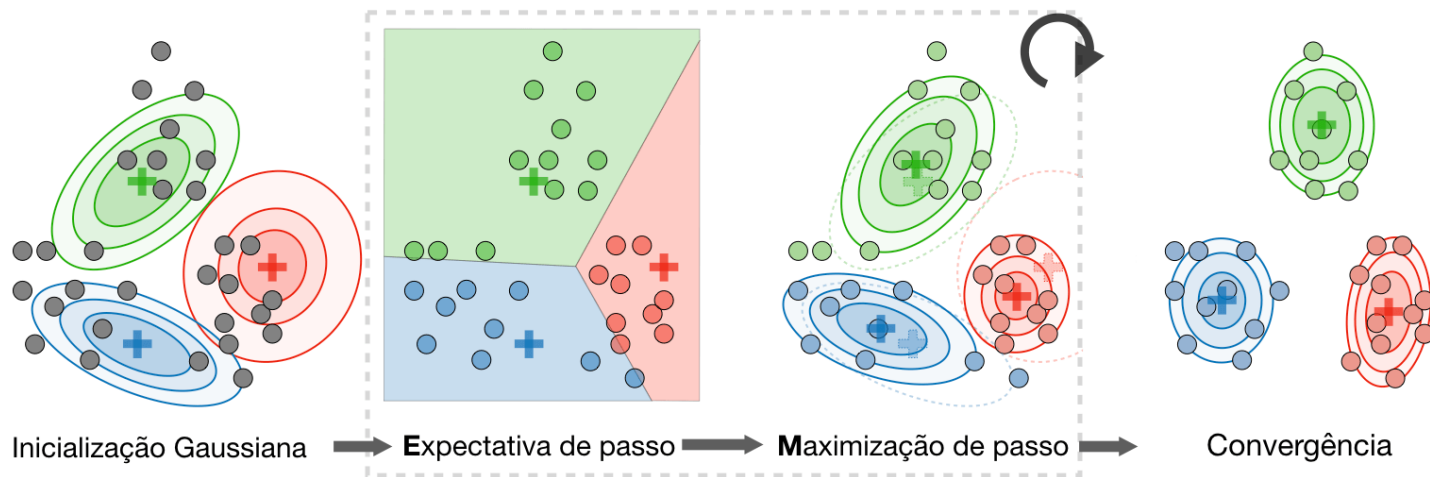


Algoritmo K-means



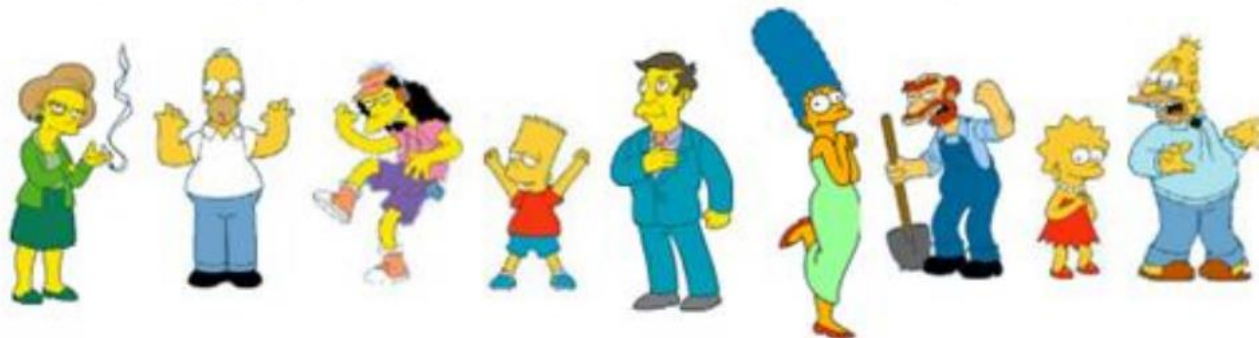
Comportamento

Funcionamento do algoritmo.



Agrupamentos

O que é agrupamento entre os seguintes objetos?



Grupo é um conceito subjetivo.



Família



Empregados da Escola



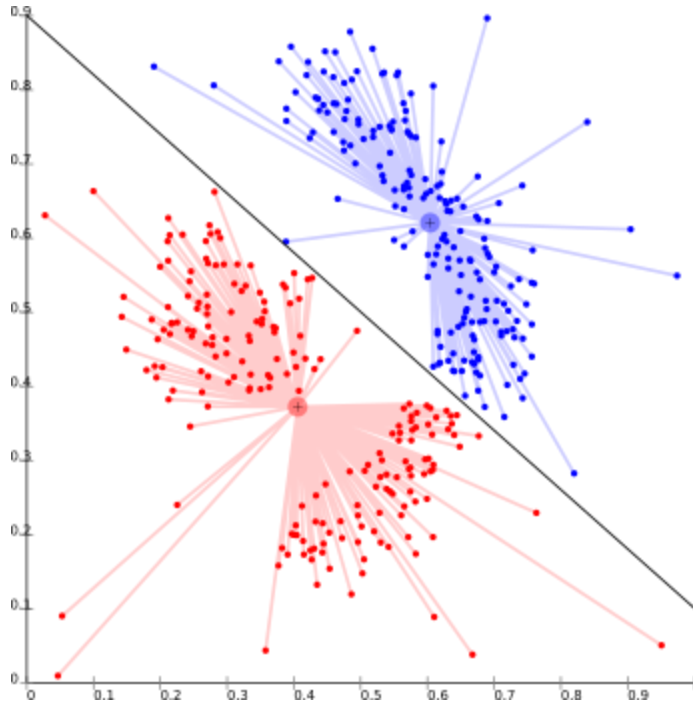
Mulheres



Homens

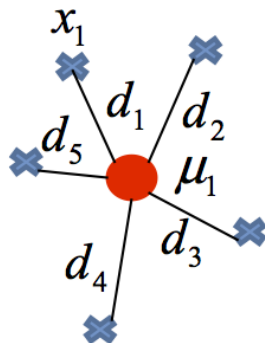
Comportamento

Cálculo da característica de cada amostra.

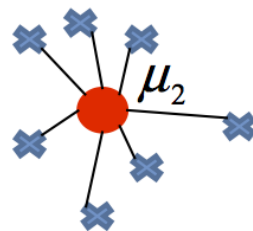


Comportamento na prática

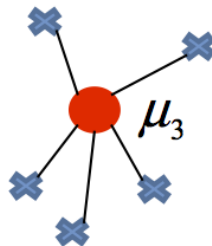
● Centroid
✕ Sample



$$\sum_{x_j \in S_2} d_j^2$$



$$\sum_{x_j \in S_1} d_j^2 = d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$$

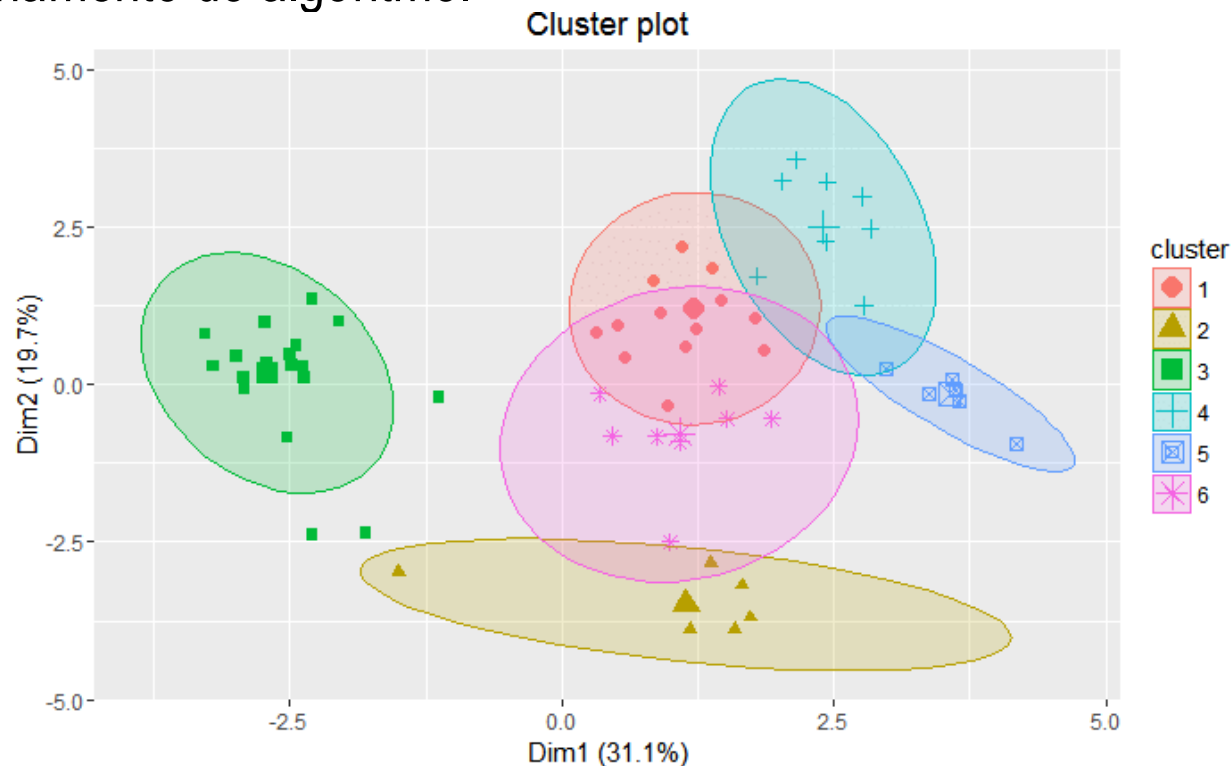


$$\sum_{x_j \in S_3} d_j^2$$

$$\min_S E(\mu_i) = \sum_{x_j \in S_1} d_j^2 + \sum_{x_j \in S_2} d_j^2 + \sum_{x_j \in S_3} d_j^2$$

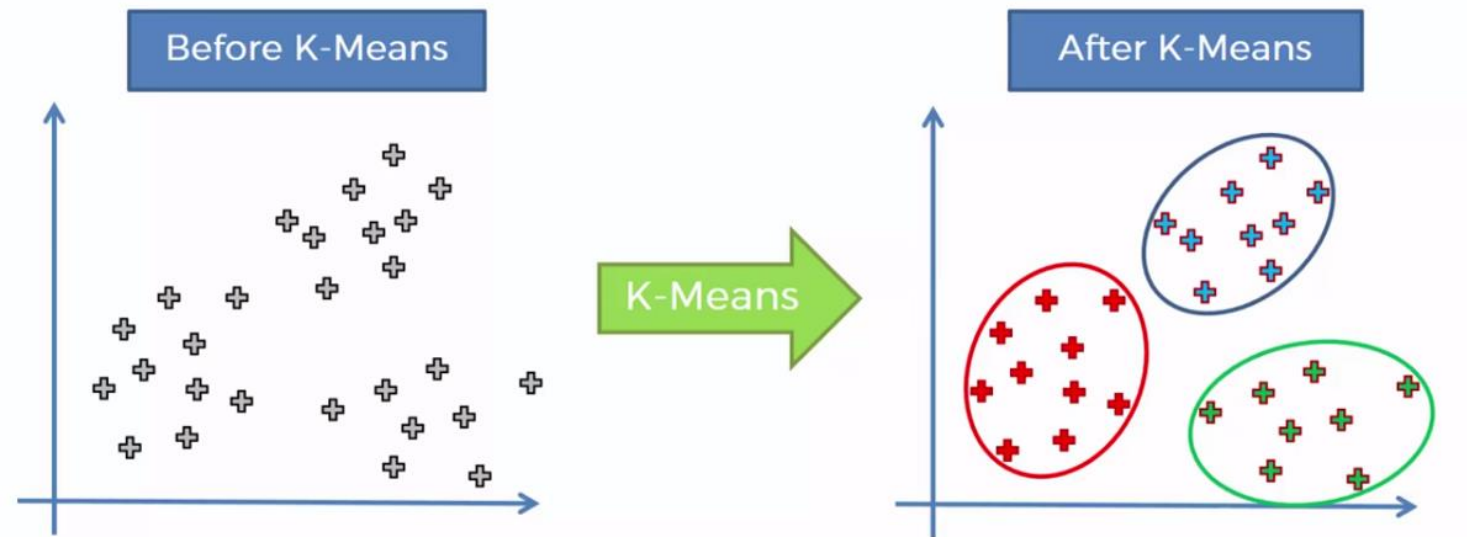
Comportamento na prática

Funcionamento do algoritmo.



Algoritmo K-means

Funcionamento do algoritmo.



Quando aplicar?

O melhor momento para usar a aprendizagem **não supervisionada** é quando **não** há dados pré-existentes sobre os resultados preferidos.

O aprendizado **não supervisionado** pode identificar recursos que podem ser úteis na categorização de conjuntos de dados desconhecidos.

Obrigado!

Machine Learning

Prof. Dr. Diego Bruno