

# K-Means

Felipe Eduardo Lopes

E-mail: felipe\_lopes@outlook.com

## 1 Introdução

O objetivo do algoritmo K-Means é fornecer uma classificação de informações de acordo com os dados fornecidos. Esta classificação, é baseada em análises e comparações entre os valores. O algoritmo vai, automaticamente fornecer uma classificação sem a necessidade de nenhuma supervisão humana, ou seja, sem precisar de pré-classificação. Por causa desta característica, o K-Means é considerado como um algoritmo de mineração de dados não supervisionado. [1]

## 2 Descrição do Algoritmo

O algoritmo tem como objetivo fazer o agrupamento de elementos em classes, o número de classes é o mesmo número de centroides. Podem existir diversos centroides, que deverão ser gerados aleatoriamente. O Número de Centroides é definido pela constante K.

## 3 Métodos

Para mostrar o funcionamento do K-means, vamos separar seu funcionamento em alguns passos:

### 3.0.1 Centroides

O número de centroides no algoritmo é definido pela constante K, ou seja se o  $K = 2$  então teremos dois centroides. As posições dos centroides devem ser geradas aleatoriamente em uma posição do plano.

### 3.0.2 Matriz de Distâncias

Após os centroides serem gerados, devemos calcular a distância de cada ponto do plano com os

centroides. Para o cálculo, normalmente é utilizado o algoritmo euclidiano.

### 3.0.3 Classificação dos prontos

Cada ponto do plano deve ser classificado, é atribuído ao centroide o ponto mais próximo a ele. Caso tenhamos N centroides, teremos N classes com os pontos mais próximos a ele.

### 3.0.4 Mover os centroides

Após a primeira classificação dos pontos para seu centroide, é necessário recalculá-lo com o objetivo de deixar a classificação mais refinada possível. O novo valor dos centroides é calculado fazendo-se a média de cada atributo de todos os pontos que pertencem a esta classe. Caso exista alteração em sua posição deve-se voltar para o passo 3.0.2. Quando o centroide não se mover mais, teremos a classificação final dos pontos.

## 4 Exemplos de aplicações

As aplicações para o algoritmo são diversas, pois com ele podemos buscar padrões difíceis de serem detectados pelo seres humanos. Como por exemplo, classificação de clientes em classes, cliente ouro, bronze e prata. Como demonstrado na Figura 1, o algoritmo foi capaz de classificar elementos em um plano 3D em duas classes.

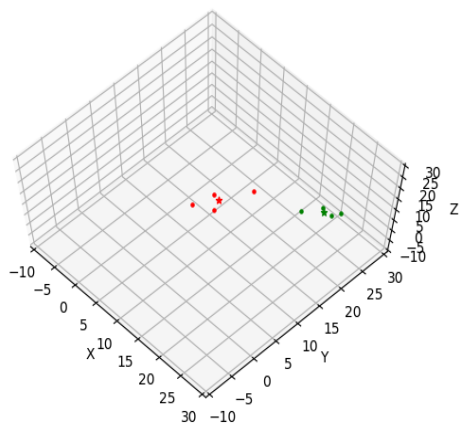


Figura 1: Agrupamento de elementos em um espaço 3D. Fonte: Próprio Autor

## Referências

- [1] Mauro Pichiliani. Data mining na prática: Algoritmo k-means. [https://imasters.com.br/artigo/4709/sql\\_server/data\\_mining\\_na\\_pratica\\_algoritmo\\_k-means/](https://imasters.com.br/artigo/4709/sql_server/data_mining_na_pratica_algoritmo_k-means/), 2006.