



Preditiva.ai

Aprendizado Não Supervisionado

Algoritmos de Cluster: K-Means

Aprendizado Não Supervisionado

Qual a diferença?



A principal diferença entre **modelos supervisionados** e **não supervisionados** é que os não supervisionados **não possuem uma variável target**.

Ou seja, a identificação de padrões se restringe as variáveis explicativas ou *features*. Essa ausência do target sem dúvida torna o processo de identificação dos padrões um pouco **mais desafiador**, mas veremos que existem diversas **técnicas desenvolvidas** especificamente para essa **abordagem**.

Entre essas técnicas temos:

- **Algoritmos de Cluster**: são utilizados para agrupar e segmentar as observações a partir das suas características presentes nas *features*.
- **Redução de Dimensionalidade**: como o próprio nome diz, reduzem a dimensão do espaço das features (ou seja, reduz a quantidade de variáveis do dataset), muitas vezes concentrando a informação em um espaço dimensional menor do que o original.
- **Detecção de Anomalias**: são utilizados para capturar padrões que permitam identificar eventos raros.

Algoritmos de Cluster

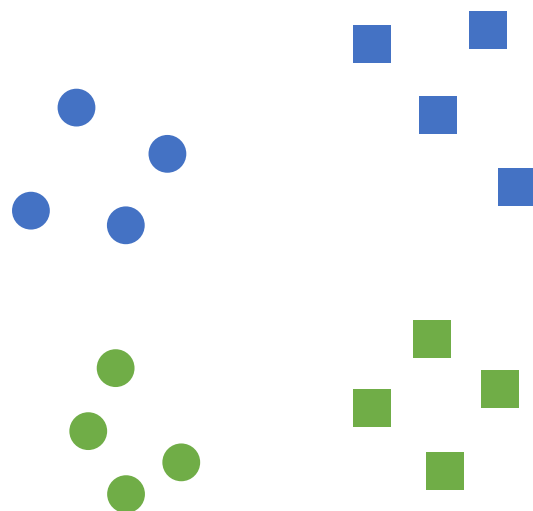
K-Means



Preditiva.ai

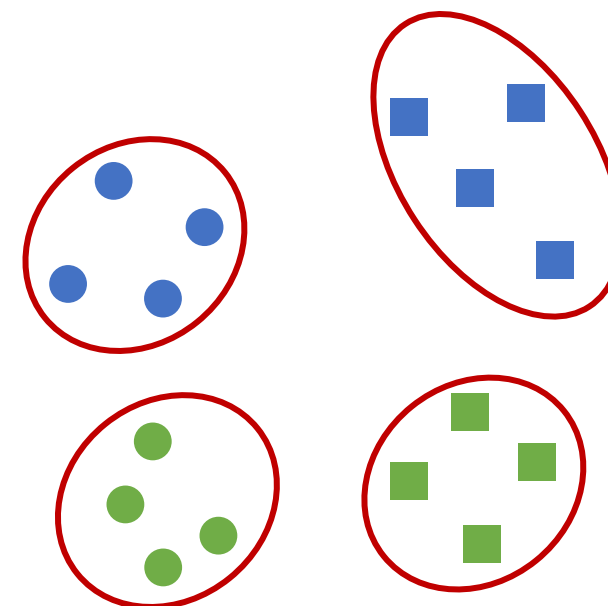
O algoritmo **K-Means** tem como principal objetivo **agrupar observações semelhantes** considerando suas características.

Conjunto de Dados sem Target



Clusterização

Agrupamento por semelhança das características

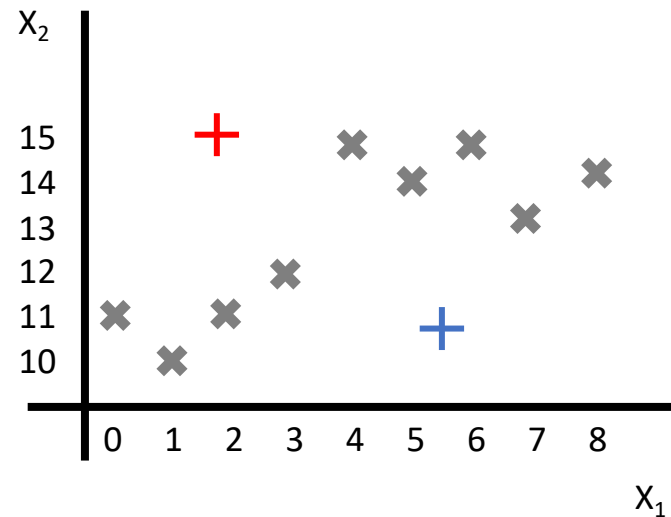


Algoritmos de Cluster

K-Means: Algoritmo



Vejamos agora como o algoritmo do K-Means funciona. Suponha o seguinte conjunto de variáveis X_1 e X_2 :



Passo 1) Definimos o número **K** de clusters que desejamos encontrar.

Ex: Vamos definir $K = 2$.

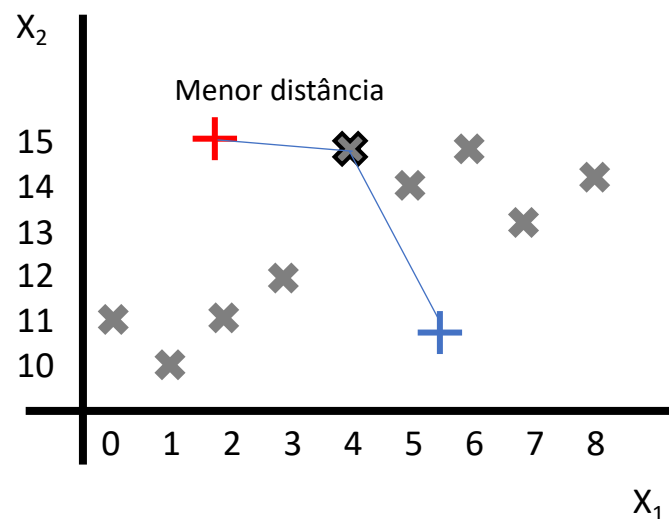
Passo 2) Sorteamos aleatoriamente a posição dos K clusters. Essa posição dos clusters é chamada de **Centróide**.

Algoritmos de Cluster

K-Means: Algoritmo



Vejamos agora como o algoritmo do K-Means funciona. Suponha o seguinte conjunto de variáveis X_1 e X_2 :



Passo 3) Para cada observação da base, calculamos a distância entre essa observação e os K centróides do clusters. Desta forma, cada observação terá K distâncias calculadas. No nosso exemplo, 2 distâncias.

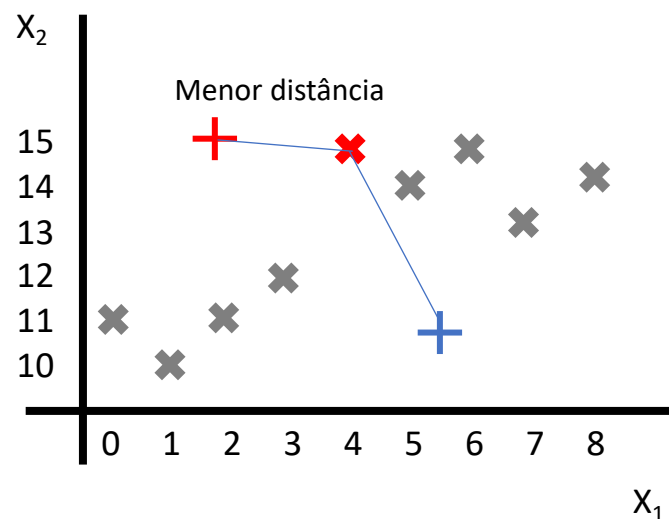
Escolhemos a menor distância entre elas e **associamos a observação ao cluster** do centroide que originou a menor distância.

Algoritmos de Cluster

K-Means: Algoritmo



Vejamos agora como o algoritmo do K-Means funciona. Suponha o seguinte conjunto de variáveis X_1 e X_2 :



Passo 3) Para cada observação da base, calculamos a distância entre essa observação e os K centróides do clusters. Desta forma, cada observação terá K distâncias calculadas. No nosso exemplo, 2 distâncias.

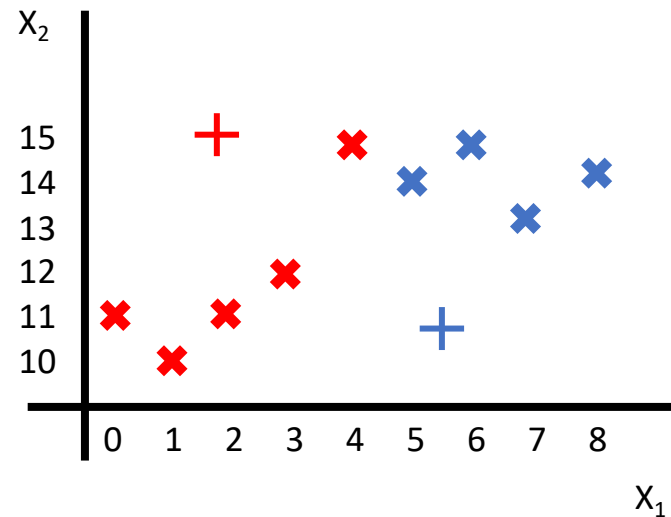
Escolhemos a menor distância entre elas e **associamos a observação ao cluster** do centroide que originou a menor distância.

Algoritmos de Cluster

K-Means: Algoritmo



Vejamos agora como o algoritmo do K-Means funciona. Suponha o seguinte conjunto de variáveis X_1 e X_2 :



Passo 3) Para cada observação da base, calculamos a distância entre essa observação e os K centróides do clusters. Desta forma, cada observação terá K distâncias calculadas. No nosso exemplo, 2 distâncias.

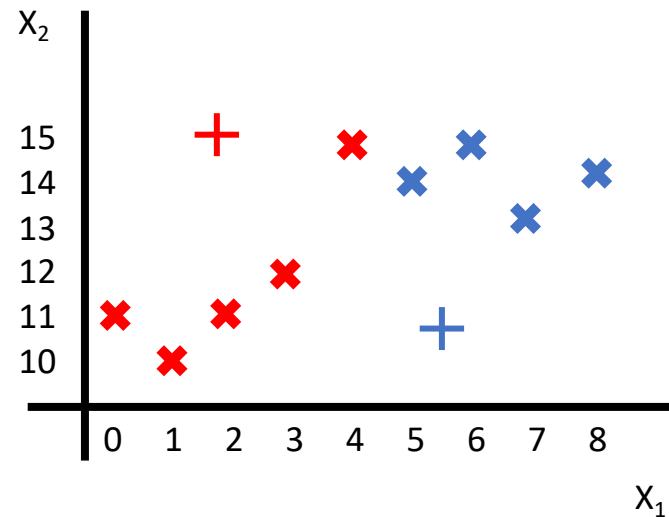
Escolhemos a menor distância entre elas e **associamos a observação ao cluster** do centroide que originou a menor distância.

Algoritmos de Cluster

K-Means: Algoritmo



Vejamos agora como o algoritmo do K-Means funciona. Suponha o seguinte conjunto de variáveis X_1 e X_2 :



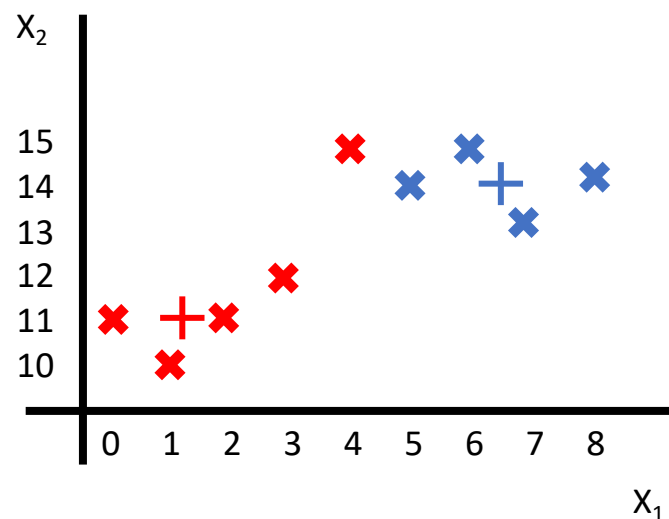
Passo 4) Neste passo, após associar todas as observações aos clusters dos centroides, calculamos a média de todas as observações dos clusters. Desta forma, **movemos o centróide para essa média**.

Algoritmos de Cluster

K-Means: Algoritmo



Vejamos agora como o algoritmo do K-Means funciona. Suponha o seguinte conjunto de variáveis X_1 e X_2 :



Passo 4) Neste passo, após associar todas as observações aos clusters dos centroides, calculamos a média de todas as observações dos clusters. Desta forma, **movemos o centróide para essa média**.

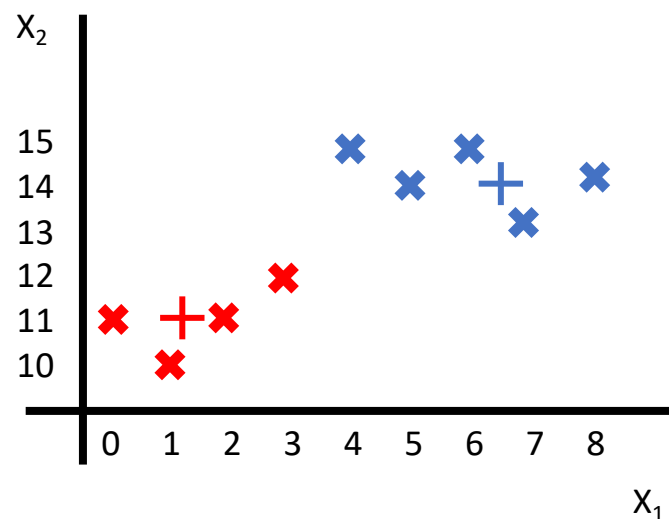
Passo 5) Neste passo, repetimos o passo 3 e calculamos todas as distâncias entre as observações e a posição do novo centróide. Após isso, reassociamos cada observação ao cluster cujo centróide forneceu a menor distância.

Algoritmos de Cluster

K-Means: Algoritmo



Vejamos agora como o algoritmo do K-Means funciona. Suponha o seguinte conjunto de variáveis X_1 e X_2 :



Passo 4) Neste passo, após associar todas as observações aos clusters dos centroides, calculamos a média de todas as observações dos clusters. Desta forma, **movemos o centróide para essa média**.

Passo 5) Neste passo, repetimos o passo 3 e calculamos todas as distâncias entre as observações e a posição do novo centróide. Após isso, reassociamos cada observação ao cluster cujo centróide forneceu a menor distância.

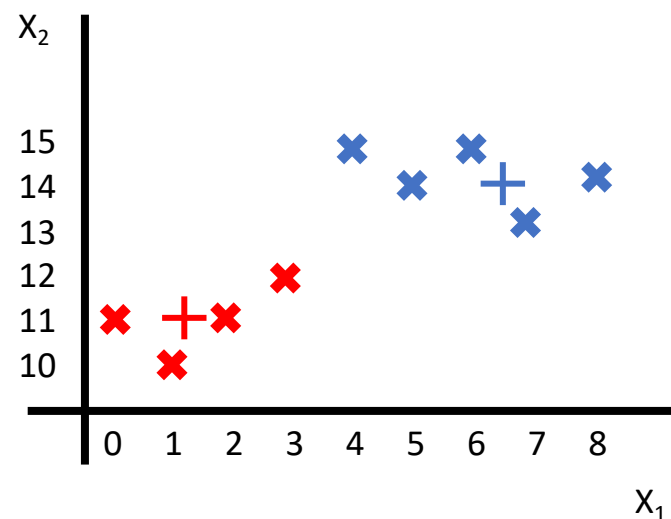
Algoritmos de Cluster

K-Means: Algoritmo



Preditiva.ai

Vejamos agora como o algoritmo do K-Means funciona. Suponha o seguinte conjunto de variáveis X_1 e X_2 :



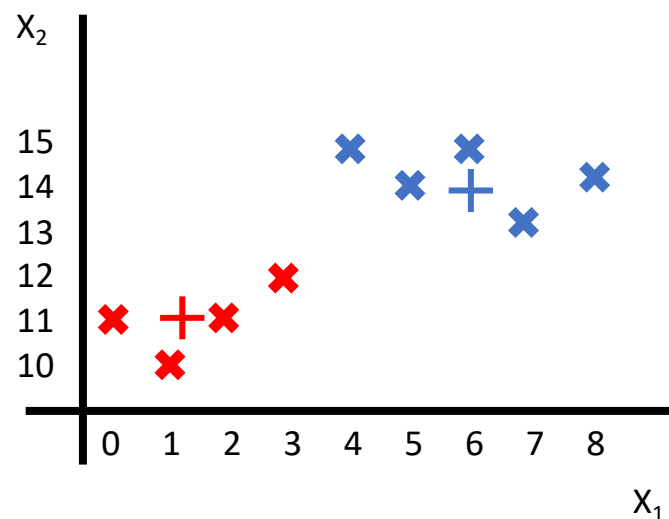
Passo 6) Neste passo, repetimos o passo 4 e posicionamos o centróide na média das observações dos respectivos clusters.

Algoritmos de Cluster

K-Means: Algoritmo



Vejamos agora como o algoritmo do K-Means funciona. Suponha o seguinte conjunto de variáveis X_1 e X_2 :



Passo 6) Neste passo, repetimos o passo 4 e posicionamos o centróide na média das observações dos respectivos clusters.

A partir daí os passos 3 e 4 são repetidos até que os centroides não tenham mais mudanças de posição.

Quando chegamos neste ponto, dizemos que o **algoritmo K-Means convergiu** para a solução de descoberta dos clusters.



Preditiva.ai