



K-Means - Teoria

☰ Ciclo	Ciclo 05: Aprendizado não-supervisionado
# Aula	34
🕒 Created	@February 24, 2023 8:50 AM
☑ Done	☑
☑ Ready	☑

Objetivo da Aula:

- ☐ O que é K-Means?
- ☐ O processo de treinamento
- ☐ As 5 premissas assumidas
- ☐ Resumo
- ☐ Próxima aula

Conteúdo:

▼ 1. O que é K-Means?

O K-Means é um algoritmo de clustering (agrupamento) que divide um conjunto de dados em k grupos (clusters) baseado nas suas similaridades. Cada grupo é representado por um centroide, que é o ponto médio de todos os pontos do grupo.

O algoritmo funciona iterativamente, alocando cada ponto ao cluster mais próximo e recalculando os centróides até que a convergência seja alcançada.

▼ 2. O processo de treinamento

▼ 2.1 Exemplo visual do funcionamento

1.

▼ 2.2 Os 6 passos para treinar o K-Means

Passo 01: Carregue os dados

Passo 02: Inicialize os K centróides aleatoriamente no espaço de dados

Passo 03: Calcule a distância entre todos os pontos e os K centróides

Passo 04: Atribua os pontos aos seus centróides mais próximos.

Passo 05: Calcule o ponto médio dentro de cada K cluster e reposicione o centróide para o centro.

Passo 06: Refaça os passos 3 até 5 até não haver mais mudanças dos pontos entre clusters.

▼ 3. As 5 premissas assumidas

As 5 premissas do K-Means são as seguintes:

1. O **número de clusters K** é conhecido.
2. **Os clusters têm formas esféricas e são igualmente densos:** o algoritmo K-means assume que todos os clusters têm formas esféricas e têm a mesma densidade.
3. **Todos os pontos de dados pertencem a pelo menos um cluster:** o algoritmo assume que cada ponto de dados pertence a exatamente um cluster.
4. **A variância dos dados é a mesma em todos os clusters:** o algoritmo pressupõe que os clusters têm tamanhos semelhantes e distribuições semelhantes.
5. **Os pontos de dados são independentes e idênticos:** Isso significa que cada ponto de dados é tratado como uma observação aleatória e não há relação entre os pontos de dados.

▼ 4. Resumo

1. O algoritmo K-Means agrupa os pontos através da similaridade entre eles, dados por alguma métrica de distância.
2. O processo de treinamento é iterativo e possui 6 passos

3. O algoritmo possui 5 premissas para o seu funcionamento

▼ 5. Próxima aula

K-Means - Prática