escola britânica de artes criativas & tecnologia

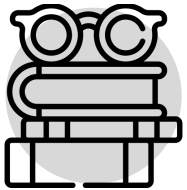
Profissão Cientista de Dados M29

escola
britânica de
artes criativas
& tecnologia

GLOSSÁRIO

K-means

- **Compreenda a Distância**
- **Entenda Como Funciona o Algoritmo**
- **Determine o Número de Clusters**



Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

Compreenda a Distância

Compreenda a Distância

Algoritmo K-means

É um algoritmo de aprendizado de máquina não supervisionado que agrupa dados em K grupos distintos com base em suas características. O agrupamento é feito minimizando a soma das distâncias entre os pontos de dados e o centro do grupo (centróide) ao qual pertencem.

Distância Euclidiana

É uma medida de distância entre dois pontos em um espaço de múltiplas dimensões. É calculada como a raiz quadrada da soma dos quadrados das diferenças entre as coordenadas dos pontos.

Compreenda a Distância

- **Função “standard”**

  É uma função em Python usada para padronizar variáveis. Ela subtrai a média e divide pelo desvio padrão de cada variável.

Entenda Como Funciona o Algoritmo

Entenda Como Funciona o Algoritmo

Algoritmo Cluster

Um tipo de algoritmo de aprendizado de máquina não supervisionado que agrupa dados semelhantes.

Convergência

No contexto do algoritmo K-means, a convergência ocorre quando os centróides não se movem mais entre as iterações. Isso indica que o algoritmo encontrou a melhor divisão dos pontos de dados em clusters, de acordo com a medida de distância usada.

Centróides

São os pontos centrais de cada cluster em um algoritmo K-means. Inicialmente, eles são definidos aleatoriamente, mas são recalculados a cada iteração do algoritmo para refletir a média das coordenadas dos pontos de dados em cada cluster.

Entenda Como Funciona o Algoritmo

Distância

É a medida usada para determinar quão próximos ou distantes dois pontos estão um do outro. No algoritmo K-means, a distância entre um ponto de dados e os centróides dos clusters é usada para determinar a qual cluster o ponto de dados pertence.

Parâmetros iniciais

São as configurações definidas antes de executar o algoritmo K-means. Eles incluem o número de grupos que o algoritmo deve formar e os centróides iniciais.

Grupos

São os clusters que o algoritmo K-means forma. Cada grupo é definido por seu centróide e contém os pontos de dados que estão mais próximos desse centróide do que de qualquer outro.

Pontos de dados

São as observações individuais que o algoritmo K-means tenta agrupar em clusters. Cada ponto de dados tem um conjunto de características que o algoritmo usa para determinar a qual cluster ele pertence.

Determine o Número de Clusters

Determine o Número de Clusters

- **Método da Silhueta**
  É uma técnica usada para determinar a qualidade de um agrupamento. Calcula uma pontuação para cada ponto em um conjunto de dados que mede quão semelhante é a outros pontos em seu cluster em comparação com pontos em outros clusters.

- **Soma dos Quadrados da Distância**
  É uma medida de quão bem os pontos se encaixam em seus clusters atribuídos. É calculada somando as distâncias quadradas de cada ponto ao centro do cluster ao qual foi atribuído.

escola britânica de artes criativas & tecnologia

Bons estudos!