

Disciplina: Inteligência Artificial

Professora: Cristiane Neri Nobre

Data de entrega: 04/11

Esta lista poderá ser feita em grupo

Para fazer as questões abaixo, sugiro que estude o material sobre **Agrupamento** que está no CANVAS. Assista também os vídeos disponibilizados sobre este assunto. Está junto com os slides.

Além disso, acesso o notebook “**Kmeans.ipynb**”, disponibilizado no CANVAS.

Questão 01

Considere o seguinte conjunto de 8 exemplos de pares de pontos no plano real. Dados os exemplos indicados na tabela abaixo, caracterizados por dois atributos (pontos em R^2), aplique o algoritmo *Kmeans* para determinar uma partição adequada desses dados em 2 grupos.

Realize a primeira execução do algoritmo, partindo dos exemplos 3 (**Ex3 -> X=0 e Y=2**) e 4 (**Ex4 -> X=0 e Y = -2**). Utilize a **distância euclidiana** para o cálculo de distância.

Execute apenas duas iterações do algoritmo e mostre o resultado do agrupamento.

Exemplo	X	Y
Ex1	-5	2
Ex2	5	-2
Ex3	0	2
Ex4	0	-2
Ex5	-5	1
Ex6	-5	-2
Ex7	5	2
Ex8	5	-1

Questão 02

Uma das grandes dificuldades em agrupamento é saber identificar o número k grupos a serem descobertos. Para isso, existem várias métricas que nos ajudam nesta difícil tarefa (Coeficiente de Silhouette, Sum of Squared Error (SSE), Davies-Bouldin, Elbow, dentre outras.). Explique o funcionamento das métricas **Silhouette index** e **Elbow**.

Mostre como estas métricas são calculadas e como avaliar os seus resultados.

Questão 03

Assista o vídeo: https://www.youtube.com/watch?v=E2M_yTulcmU

E analise as limitações deste algoritmo.

Questão 04

Faça um resumo do artigo “A Comprehensive Survey Of Clustering” que está no CANVAS.

Questão 05

Para as questões a seguir, você deverá utilizar o notebook “**Kmeans.ipynb**”, que está no CANVAS. Considere a base de dados da Iris.

5.1. Adicione ao notebook as etapas de identificação de outlier e seleção de atributos. Isso pode não ser necessário para a base da Iris, mas são muito necessárias em outras bases.

5.2 Rode a base da Iris com o atributo de classe. Ou seja, deixe ele na base (convertido para numérico, claro) e experimente o Kmeans neste agrupamento. O que você observa?

5.3 Rode a base da Iris novamente sem o atributo de classe e veja o que acontece. Discuta estes dois experimentos (com e sem o atributo de classe).

5.4 Verifique e liste quais instâncias estão sendo agrupadas incorretamente.

5.5 Refaça os experimentos, separando 20% para testar o agrupamento. Ou seja, faça os agrupamentos e descubra em quais grupos os seus 20% estarão agrupados.

Questão 06 (EXTRA) – 05 pontos

Rode o algoritmo Kmeans na base de dados a seguir:

<https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python?resource=download>

Encontre os agrupamentos, discuta a qualidade destes agrupamentos e caracterize os agrupamentos obtidos.

Faça um pequeno relatório explicando todas as etapas de pré-processamento realizadas e explicando todos os achados.

Observação

Coloque os links para os códigos produzidos ao final de cada questão