

GCC 128 - Inteligência Artificial

Ahmed Ali Abdalla Esmin - Anna Paula Figueiredo

Trabalho Prático 02 - Clusterização *K-means*

Atividade Proposta

A ideia dessa atividade é fixar os conhecimentos adquiridos em sala de aula, desenvolvendo uma aplicação que simule o algoritmo *K-means*.

Você deve implementar o algoritmo, isto é, desenvolvê-lo do zero (*hardcore*).

O trabalho poderá ser feito em dupla.

Atente-se à data de entrega: **16 / 04 / 2024.**

Total de pontos da atividade: **15 pontos.**

Instruções

- Utilize a [base de dados Iris](#) como referência para o *cluster*. **obs: você deve desconsiderar a classe alvo, isto é, a variável target para realizar a clusterização.**
- Você encontra essa base no [keaggle](#), ou pode importá-la através de bibliotecas que a disponibilizam, como por exemplo o [Sklearn](#).
- Desenvolva uma clusterização utilizando a linguagem de programação [Python](#) ([google colab](#), [pycharm](#), [vscode](#), [jupyter notebook](#), entre outros), **sem utilizar bibliotecas que contêm o algoritmo *K-means* implementado.** Você pode utilizar bibliotecas para leitura e tratamento dos dados, como por exemplo o [Pandas](#), quando necessário, além de utilizar qualquer tipo de estrutura de dados para compor sua solução.
- Quantidade de clusters: um experimento para **3** e outro experimento para **5**.
- Avaliação: após desenvolvê-lo, calcule o [silhouette score](#) utilizando a biblioteca [Sklearn](#) para a quantidade de centróides indicadas anteriormente. Faça uma análise do resultado.
- Feito isso, desenvolva um cluster, utilizando a biblioteca [Sklearn](#), ou outra similar, desde que possua o algoritmo já implementado. Em seguida, imprima também os valores referentes à métrica de avaliação.
- Faça um experimento, com o melhor **k** encontrado, de modo a reduzir a dimensionalidade da base de dados utilizando a técnica de [PCA](#). Varie entre 1 e 2 componentes. Plote o cluster e os centróides para as versões com 1 e 2 componentes.
- Faça uma análise de desempenho dos dois algoritmos de cluster, sendo a implementação *hardcore* e a implementação que você utilizou a biblioteca, com o *kmeans* já implementado por terceiros.

Você deve: Comparar as métricas de avaliação de ambos os clusters.

Você pode: Comparar o tempo de execução, utilizar as classes originais, que foram removidas, para comparar com os cluster encontrados, memória utilizada e afins.

Entrega: único arquivo nome1_nome2.ZIP

Código das aplicações (`.py` ou `.ipynb`)

Relatório de até 1 página (`.pdf`): contendo o resultado da comparação entre as aplicações, e uma breve análise do desempenho das aplicações (com sua conclusão).

BOM TRABALHO!