

4o. Trabalho Computacional
(TIP8311 – Reconhecimento de Padrões)
(CCP9014 - Tópicos Especiais em Inteligência Computacional e Otimização)

Programa de Pós-Graduação em Engenharia de Teleinformática (PPGETI)
Programa de Pós-Graduação em Modelagem e Métodos Qualitativos (PPGMMQ)
Universidade Federal do Ceará (UFC)

Responsável: Prof. Guilherme de Alencar Barreto

Data: 20/12/2023

* **Questão Única:** Implementar os algoritmos K -médias e K -medianas usando o conjunto de dados disponibilizado no SIGAA (datasetTC4.dat).

Pede-se:

1.1. Usando o algoritmo K -médias, estimar o número de agrupamentos através dos índices de validação Dunn, Davies-Bouldin e Calinski-Harabasz. Responder: (i) Qual valor para o número de agrupamentos foi sugerido por cada técnica de validação? (ii) Se houve divergência entre os resultados sugeridos pelos índices, o que justifica tal divergência?

OBS: Usar a metodologia apresentada na página 40 dos slides sobre clusterização de dados, repetindo-a por um certo número de rodadas (e.g. 100 rodadas). Escolher como o número ótimo de clusters o número mais frequente (moda) ao longo das 100 rodadas.

1.2. Rodar o algoritmo K -medianas usando o número de agrupamentos escolhidos no Subitem 1. Responder: (i) As posições finais dos protótipos diferem daquelas obtidas pelo algoritmo K -médias? (ii) Projete os dados em 2 dimensões via PCA para visualizar as posições dos protótipos obtidos pelos algoritmos K -médias e K -medianas. (iii) Que conclusões podem ser tiradas desse experimento em relação à presença ou não de *outliers* nos dados?

1.3 Usando o número de agrupamentos escolhidos no Subitem 1 e o algoritmo K -médias, realizar a análise estatística de cada agrupamento encontrado. Forneça a posição de cada protótipo, além dos valores mínimo, máximo, mediana e desvio-padrão de cada atributo por agrupamento (ver slide 44).

Boa Sorte!