

3o. Trabalho Computacional (TIP8311 – Reconhecimento de Padrões)

Programa de Pós-Graduação em Engenharia de Teleinformática (PPGETI)
Departamento de Engenharia de Teleinformática (DETI)
Universidade Federal do Ceará (UFC)

Responsável: Prof. Guilherme de Alencar Barreto

Data: 21/11/2019

*** Questão 1 (clusterização via algoritmo *K*-médias):** Implementar as versões *batch* e *sequencial* algoritmo *K*-médias discutidas em sala de aula usando o conjunto de dados disponibilizado no SIGAA (datasetTC3.dat).

Pede-se:

1.1. Estimar o número de agrupamentos usando os índices de validação descritos em sala de aula (Dunn, Davies-Bouldin e Calinski-Harabasz). Houve divergência entre os resultados sugeridos pelos índices? Em caso afirmativo, o que justifica tal divergência?

1.2. Usando o menor dos valores para o número de agrupamentos do item anterior, realizar a análise estatística de cada agrupamento encontrado. Fornecer a posição de cada protótipo, além dos valores mínimo, máximo, mediana e desvio-padrão de cada atributo por agrupamento.

OBS: Usar a metodologia apresentada na página 29 dos slides, 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.

*** Questão 2 (opcional):** Demonstrar matematicamente que as duas versões do algoritmo *K*-médias são equivalentes.

Boa Sorte!