

## 1. Especificação do Problema

Implemente por meio de *storage procedures* o algoritmo de agrupamento k-medoid para realizar a geração de agrupamentos de acordo com respectiva(s) tabela(s) de entrada. O objetivo de tal algoritmo é minimizar o erro quadrático médio. Como critério de convergência para tal algoritmo pode-se utilizar uma quantidade pré-definida de iterações ou um limiar (e.g. erro) entre iterações subsequentes. O fluxograma do algoritmo k-medoid é explicitado na Figura 1.

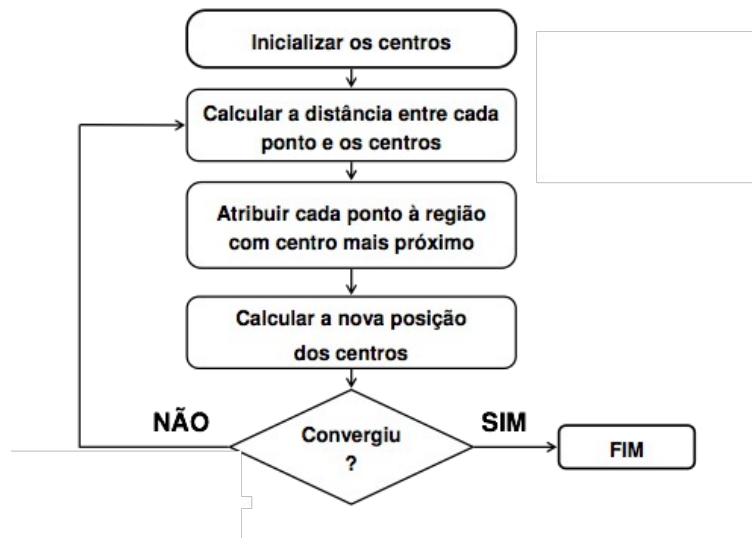


Figura 1 – Fluxograma genérico do algoritmo de agrupamento

O escopo do respetivo trabalho é delimitado pela utilização de valores numéricos (e.g. int, float) para a realização do processo de agrupamento. Vale ressaltar que tais valores devem ser normalizados para posteriormente serem utilizados. Como método de normalização dos dados deve-se utilizar a normalização *min-max*, sendo os dados normalizados em  $[0, 1]$ . Para o cálculo da dissimilaridade entre diferentes amostras da base de dados empregue a distância euclidiana.

### 1.4. Especificação do(s) Storage Procedure(s)

Sugestão de Parâmetros de Entrada:

- Tabela(s) com dados a serem agrupados
- Quantidade de agrupamentos a serem definidos

Sugestão de Saída:

- Centróides (Medoids)
- Amostras rotuladas de acordo com os respectivos agrupamentos

## 1.5 Experimentos

Como entrada para tal experimento utilize as bases disponibilizadas via moodle. Caso determinado grupo desejar utilizar um base de dados distinta especificar a mesma, bem como justificar sua respectiva relevância e adequação às premissas definidas no presente trabalho.

Após a implementação do respectivo algoritmo realize experimentos variando a a base de dados, bem como a quantidade de agrupamentos (valor de  $k$ ), gerando as respectivas saídas e análise de tais variações.

## 1.6. Especificação da Entrega

Utilizar o Moodle para submissão do trabalho. A entrega do respectivo trabalho deve conter um arquivo compactado no formatos *zip* ou *tgz*. Incluir em tal arquivo os seguintes itens:

- script de criação e população dos dados; scripts de criação dos storage procedures; scripts de execução do trabalho gerando as devidas saídas. Além disso, é necessário comentar os trechos importantes do código. Não se esqueça de inserir em cada script seu nome e RA, bem como inclua um arquivo “README.txt” com especificações de como executar o respectivo trabalho;
- Relatório com as análises realizadas a partir dos resultados dos experimentos.

**Observação:** É salutar o compartilhamento de soluções e formas de resolução do problema. No entanto não o código. Soluções e ideias desenvolvidas em colaboração com outro(s) grupo(s), devem ser devidamente documentadas em comentários no referido trecho. A detecção de cópia de parte ou de todo código-fonte, de qualquer origem, implicará reprovação direta no trabalho.