## Trabalho 2 – Análise de Agrupamentos

Desenvolver um programa do método de agrupamento *k*-means (algoritmo abaixo) e testá-lo para a base de dados Wine.

Os resultados devem ser apresentados através de um relatório impresso contendo:

1. Listagem do programa fonte, onde o cálculo de distância deve ser feito em procedimento específico. Desenvolver dois procedimentos para o cálculo de distância: um usando a distância *Manhattan* e o outro a distância Euclidiana.
2. Listagem das amostras da base de dados, acompanhadas da respectiva classe real e da classe calculada pelo *k*-means, somente para a execução que apresentar o maior percentual de acertos.
3. Breve análise das execuções realizadas (sugestões: estatística sucinta de cada execução, tentar obter uma visão bidimensional conforme exemplo abaixo, comparar desempenho com o WEKA etc.).

Observações:

1. O fornecimento da classe de cada amostra serve somente para o cálculo do percentual de acertos e, evidentemente, não pode ser considerada para o agrupamento.
2. Devem ser realizadas pelo menos seis execuções, sendo três para cada tipo de distância (*Manhattan* e Euclidiana) e usando:
   2.1. Base de dados originalmente fornecida.
   2.2. Base de dados original, padronizando os valores de todas as variáveis (escore-z).
   2.3. Base de dados original, normalizando linearmente os valores de todas as variáveis.
3. O trabalho pode ser desenvolvido por equipe de, no máximo, 2 alunos.
4. Correção Comparativa.

O algoritmo ***k*-means**:

**Entrada:** *D*: um conjunto de dados com *n* objetos.
*k*: o número de grupos.

**Saída:** Um conjunto de *k* grupos.

**Algoritmo:**

(1) "escolha arbitrariamente *k* objetos de *D* como os centros iniciais dos grupos";
(2) **repita**
(3) "atribua cada objeto para o grupo que tenha o centro mais próximo do objeto";
(4) "calcule os novos centros dos grupos (novo centro de um grupo = valor médio dos objetos do grupo)";
(5) **até** "não ocorrer mais mudanças nos centros";

Exemplo de visão bidimensional gráfica: **Percentual de Acertos = 100.00%**