

Agrupamento de Dados de Expressão Gênica utilizando WEKA

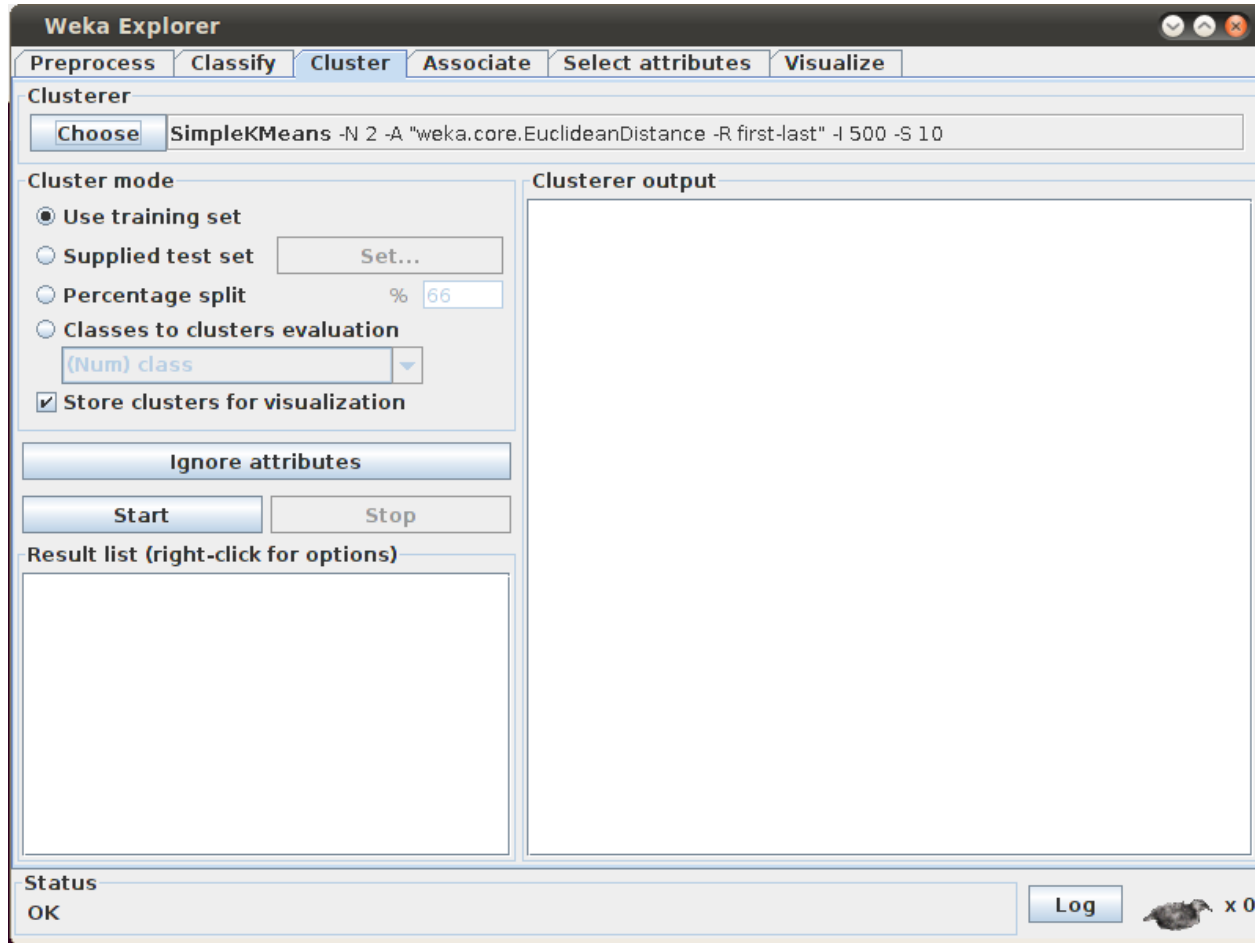
Estagiário PAE: Pablo Andretta Jaskowiak

Professor: Ricardo J. G. B. Campello

Sumário

- Aplicação de Algoritmos
- Aplicação em Dados de Expressão Gênica
- Outras Ferramentas

Aplicação de Algoritmos



Aplicação de Algoritmos

- Algoritmo *k-means* (*SimpleKMeans*)
 - ▣ Presente em praticamente todas as versões
 - ▣ Escolha da medida de proximidade
 - ▣ Número de grupos
 - ▣ Número máximo de iterações
 - ▣ Semente

Aplicação de Algoritmos

□ Hierárquicos

▣ Presentes somente nas versões mais novas

□ Versão 3.7.2

▣ Escolha da medida de proximidade

□ Entre objetos

□ Entre grupos

□ *Single-Linkage*

□ *Average-Linkage*

□ *Complete-Linkage*

▣ Número de grupos*

Aplicação de Algoritmos

- Utilizando os algoritmos passo a passo
 - ▣ Abrir o arquivo *base_1.arff*
 - ▣ Aplicar o algoritmo *k-means*
 - ▣ Aplicar os algoritmos hierárquicos

Aplicação de Algoritmos

1. Abrir cada uma das bases

base_2.arff

base_3.arff

2. Visualizar as bases

3. Avaliar os resultados

Gerados como sumário

Avaliação visual

Bases de Dados

- Agrupamento de Dados de Expressão Gênica
 - Amostras
 - Provenientes de diferentes tecidos
 - Genes
 - Experimentos de séries temporais
 - Expressão medida ao longo do tempo

Bases de Dados

□ Agrupamento de Amostras

■ Golub *et al.*, 1999

□ Descoberta de tipos de leucemia

□ AML (25 amostras)

□ ALL (47 amostras)

■ Alizadeh *et al.*, 2000

□ Identificação de diferentes tipos de linfomas

□ DLBCL1 (21 amostras)

□ DLBCL2 (21 amostras)

Golub, T. R. *et al.* *Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring.* Science, 286(5439):531-537, 1999.

Alizadeh, A. A. *et al.* *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.* Nature, 403(6769):503-511, 2000.

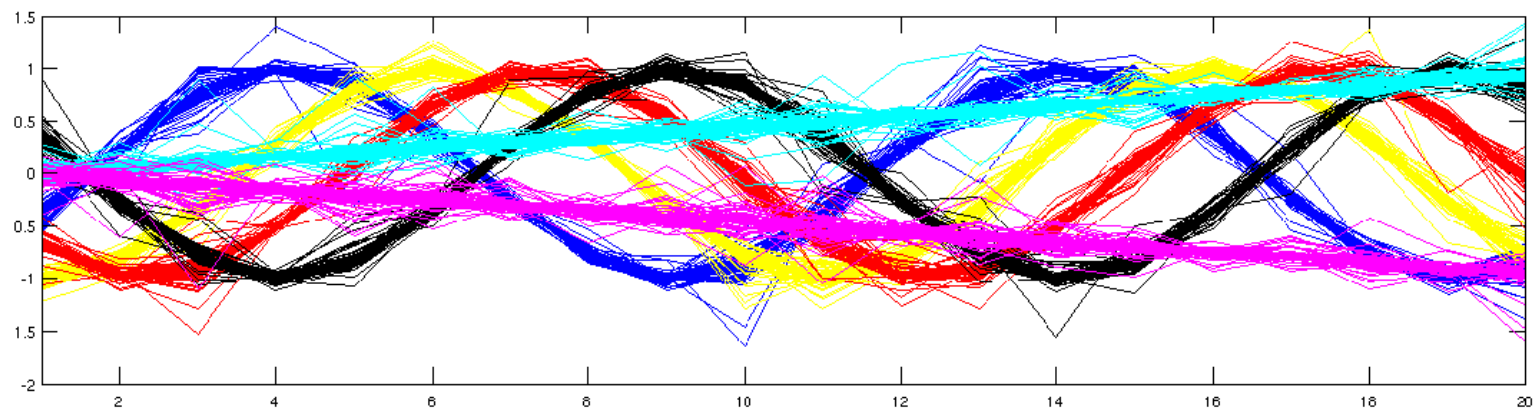
Bases de Dados

□ Agrupamento de Genes

■ Bases sintéticas (Yeung *et al.*, 2003), com 6 grupos

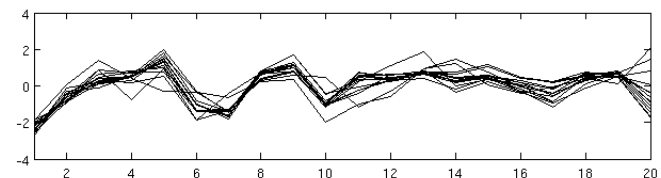
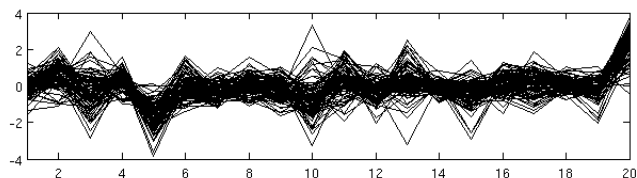
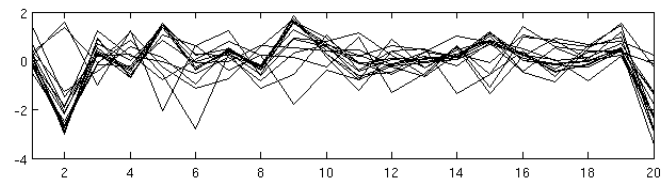
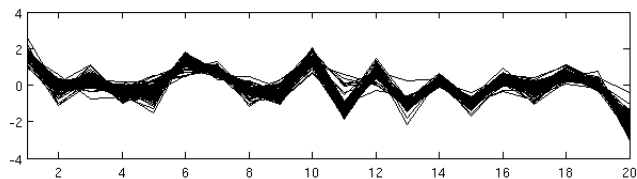
□ Sin2

□ Sin4



□ Agrupamento de Genes

- Base de dados yeast (Cho *et al.*, 1998), com 4 grupos



Outras Ferramentas

- Mayday – Microarray Data Analysis
 - ▣ <http://www-ps.informatik.uni-tuebingen.de/mayday/wp/>
- RapidMiner
 - ▣ <http://rapid-i.com/content/view/181/190/>
- R Language
 - ▣ <http://www.r-project.org/>