

Análise Multivariada de Dados - Aula 04

Análise de Conglomerados III: Implementações

Kaique Matias de Andrade Roberto

Ciências Atuariais

HECSA - Escola de Negócios

FIAM-FAAM-FMU

1. Conceitos que aprendemos em Aulas anteriores
2. Implementações
3. Comentários Finais
4. Referências

Conceitos que aprendemos em Aulas anteriores

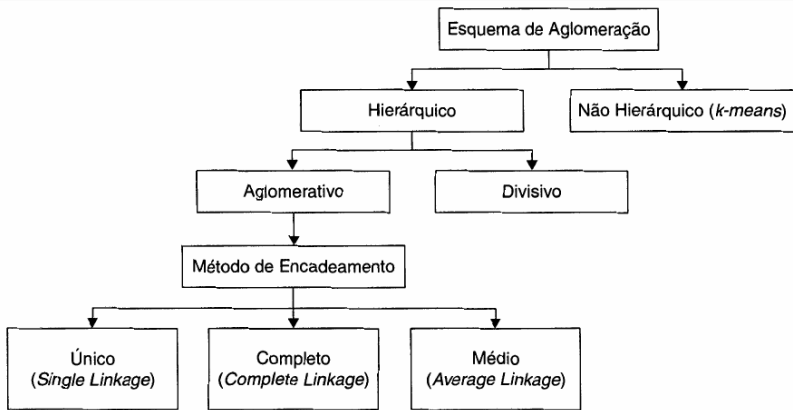
Nas últimas Aulas focamos em aspectos teóricos dos Esquemas de Aglomeração, no caso:

- medidas de distância;
- medidas de similaridade;
- descrevemos os esquemas de aglomeração;
- descrevemos os algoritmos para esquemas hierárquicos;
- realizamos alguns exemplos para esquemas hierárquicos
- descrevemos o algoritmo para o esquema não-hierárquico k-means.

Implementações

Relembramos a seguir o nosso quadro de técnicas para Análise de Agrupamentos

Implementações



Ao usar algum método hierárquico, podemos considerar vários tipos de distância e vários métodos de encadeamento. Para as distâncias reveja a Aula-02. Dentre os métodos de encadeamento mais comuns temos:

Implementações

Método de Encadeamento	Ilustração	Distância (Dissimilaridade)
Único <i>(Nearest Neighbor ou Single Linkage)</i>		d_{23}
Completo <i>(Furthest Neighbor ou Complete Linkage)</i>		d_{15}
Médio <i>(Between Groups ou Average Linkage)</i>		$\frac{d_{13} + d_{14} + d_{15} + d_{23} + d_{24} + d_{25}}{6}$

Na aula de hoje vamos implementar alguns destes métodos.

Vimos nas Aulas 02 e 03 algumas dificuldades na implementação destes métodos em Excel[©].

Neste curso, vamos adotar uma estratégia híbrida: tratar uma parte dos dados com Excel[©] e implementar alguns métodos em R.

Lembrando que **a implementação não é conteúdo de prova nem conteúdo obrigatório para o curso.**

Apenas considero importante que vocês tenham uma visão prática de como estes métodos funcionam em ambientes mais "robustos" que o Excel[©].

Ademais, vocês verão em seguida que de fato, o R é um ambiente **altamente adequado** para lidar com métodos estatísticos.

Vocês podem baixar as implementações no AvA.

Comentários Finais

Em resumo, na aula de hoje nós fizemos algumas implementações iniciais com foco em Análise de Conglomerados.

Na próxima aula vamos lidar com teoria de Matrizes. Precisamos de alguns resultados para desenvolver os próximos métodos multivariados.

Referências

Não usamos referências específicas para esta Aula.

Mas recomendo alguns links caso vocês queiram se aprofundar em R ou Python:

Site Oficial do Projeto R

- <https://www.r-project.org/>

Statology

- Site legal para exemplos básicos de implementações em R:
<https://www.statology.org/>

Apostila de R em Português

- <https://www.ime.usp.br/~rvicente/Rembrapa.pdf>

Documentação Oficial Python

- <https://www.python.org/doc/>

Documentação Oficial scikit-learn

- https://scikit-learn.org/stable/getting_started.html

Documentação Oficial SciPy

- <https://docs.scipy.org/doc/scipy/>

