Data Mining

Notes on Clustering Strategies

24/11/2021

NOVA-IMS

Fernando Lucas Bação

bacao@isegi.unl.pt

http://www.isegi.unl.pt/fbacao

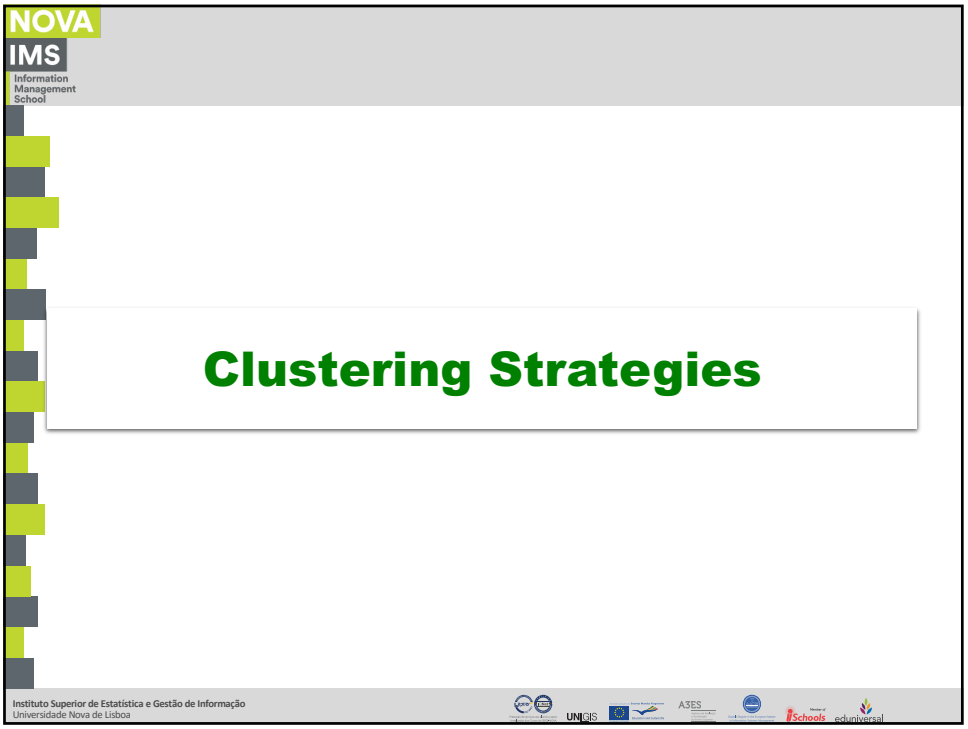
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

1

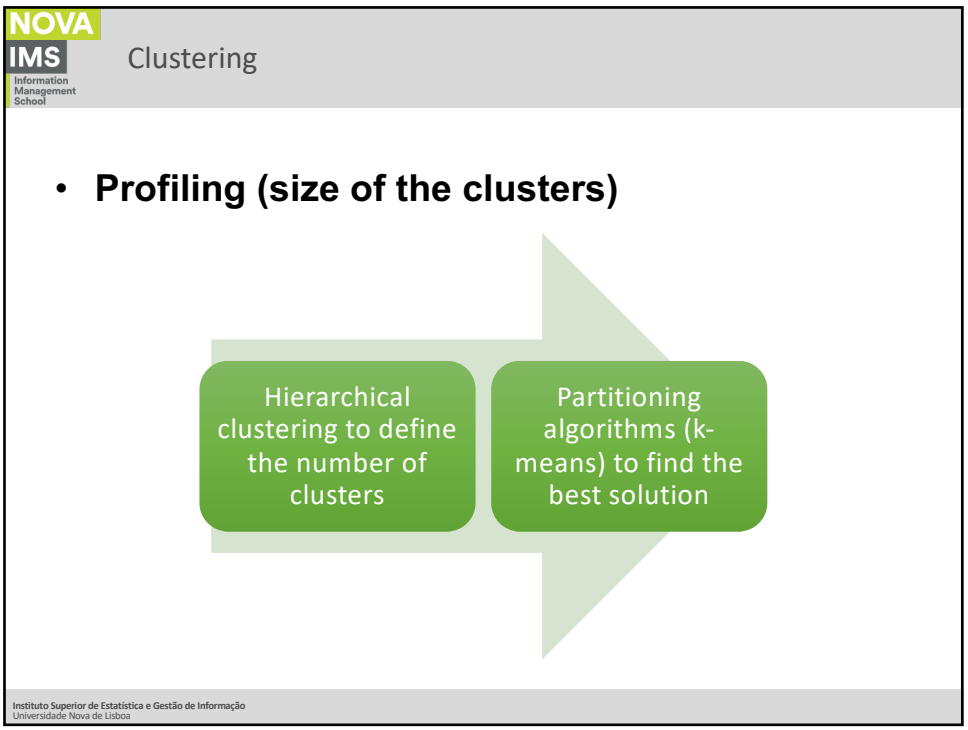
AGENDA

- Cluster analysis
  - Clustering Strategies
  - Reintroducing the Outliers or New Observations

2


NOVA
IMS
Information Management School
Clustering Strategies
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

3


NOVA
IMS
Information Management School
Clustering
• Profiling (size of the clusters)
Hierarchical clustering to define the number of clusters
Partitioning algorithms (k-means) to find the best solution
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

4

NOVA IMS Information Management School

Clustering

- **Profiling (size of the clusters)**

Partitioning algorithm (k-means) with a large number of clusters

Hierarchical algorithm to find the appropriate number of clusters

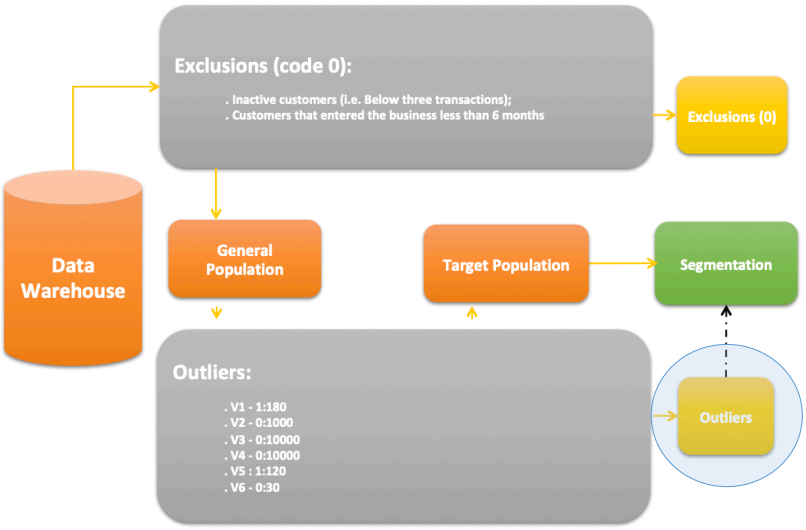
5

NOVA IMS Information Management School

Reintroducing the Outliers or New Observations

6

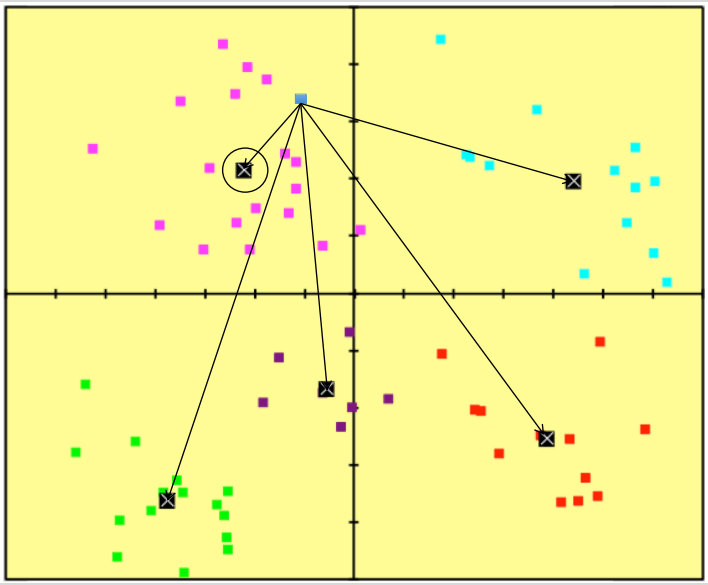
Clustering



Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

7

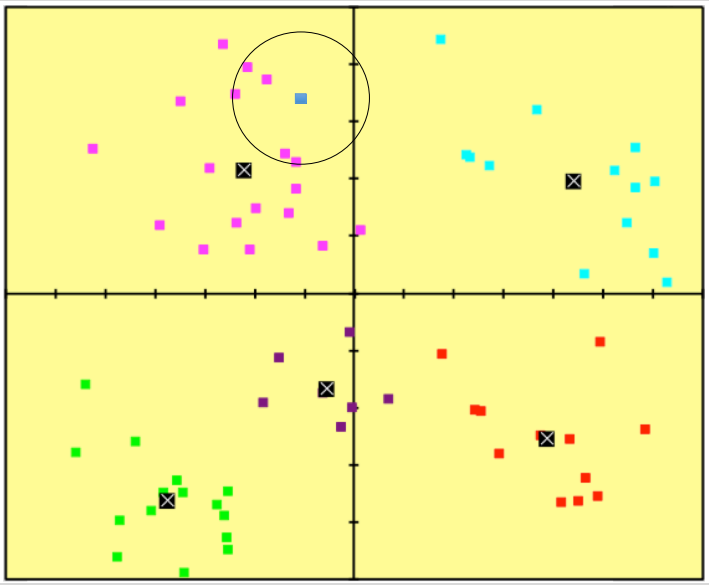
Going from clustering into classification



Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

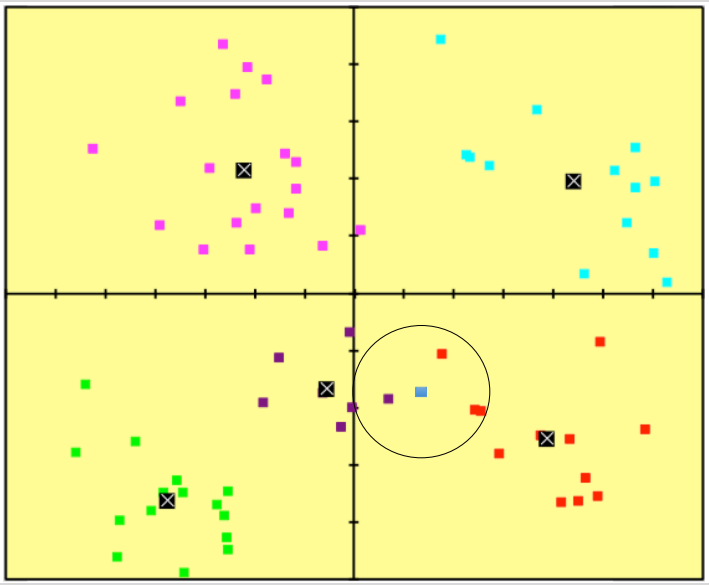
8

Going from clustering into classification



9

Going from clustering into classification



10