

Feature reduction

Violaine Antoine

ISIMA / LIMOS

January, 2020

Outline

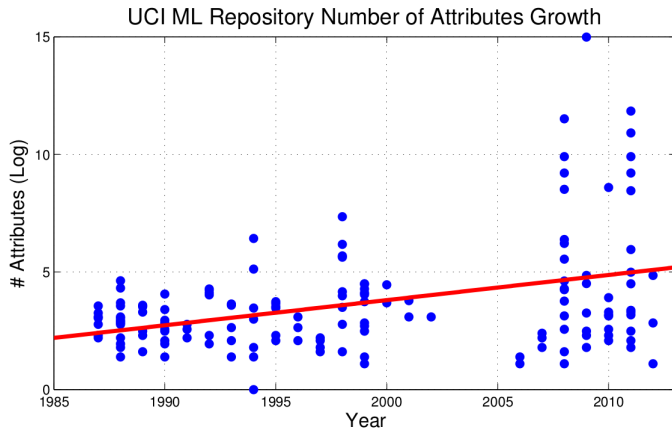
- 1 Introduction
- 2 Feature selection

Outline

- 1 Introduction
- 2 Feature selection

Dimensionality reduction necessity [1]

- improve learning quality
- reduce computational time
- reduce required storage



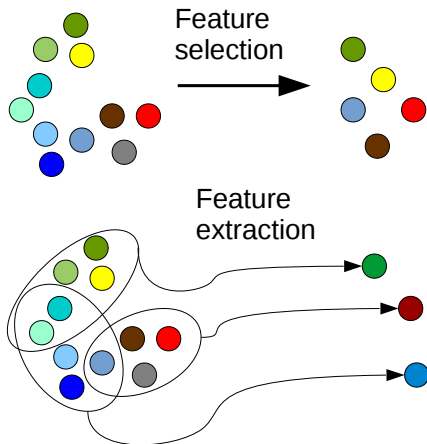
Feature reduction methods

- Feature selection

- many irrelevant and very redundant attributes
- attribute interpretation necessity

- Feature extraction

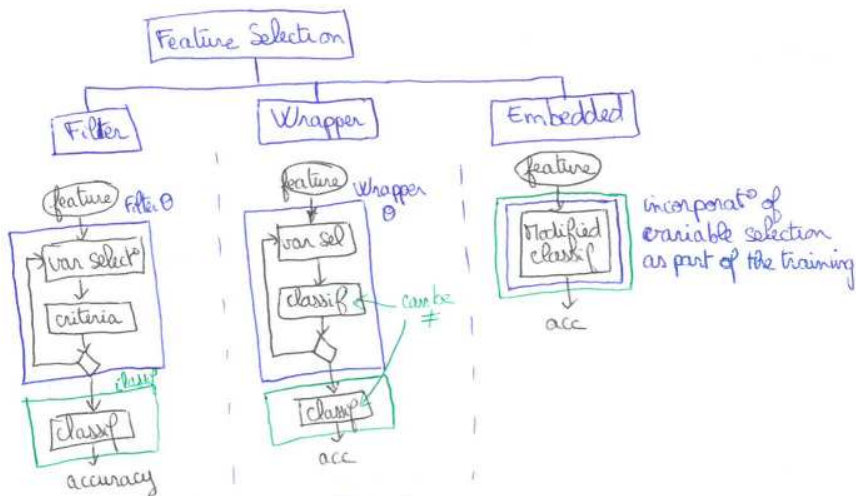
- existence of correlations between variables
- no understandable format needed



Outline

- 1 Introduction
- 2 Feature selection

Attributes selection



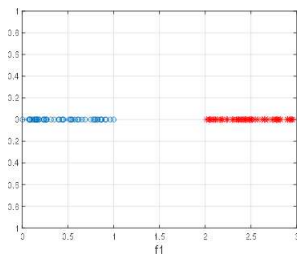
Attributes selection

- Filter : not relied on a particular method
 - + generally faster
 - + more universal
 - choice of criteria difficult, may fail to select usefull features
- Wrapper : tied to a particular method
 - + adapted to a given algorithm
 - overfitting possible
- Embedded : classification/clustering including feature selection
 - + specialized method for a problem
 - may be computational expensive and not already implemented

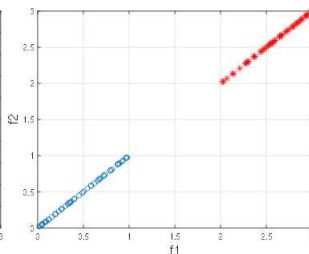
Attributes selection

Focus on feature selection for clustering [3]

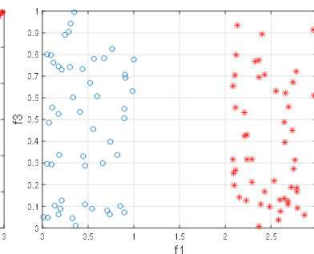
- ignore redundant features
- remove irrelevant features



(a) relevant feature f_1

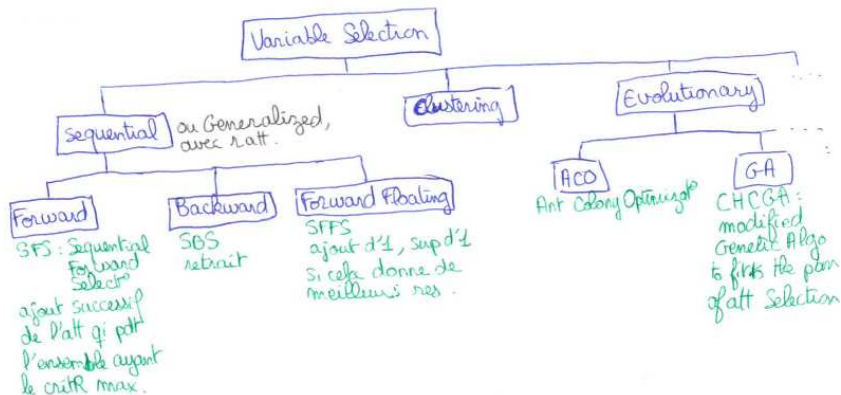


(b) redundant feature f_2



(c) irrelevant feature f_3

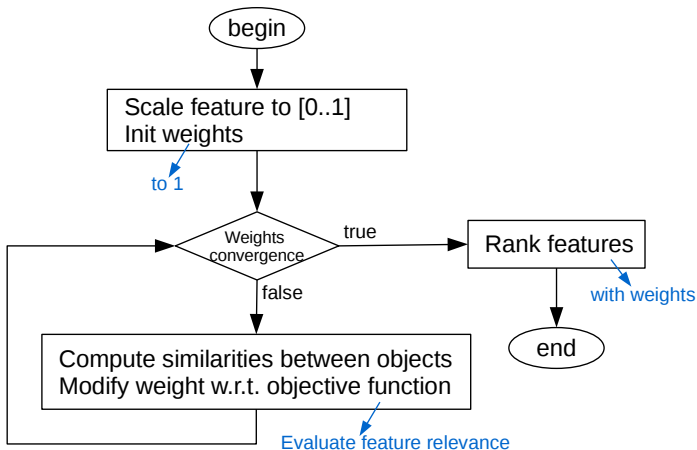
Variable selection



criteria measures

- separability measures
- cluster density measures
- correlation/redundancy between features

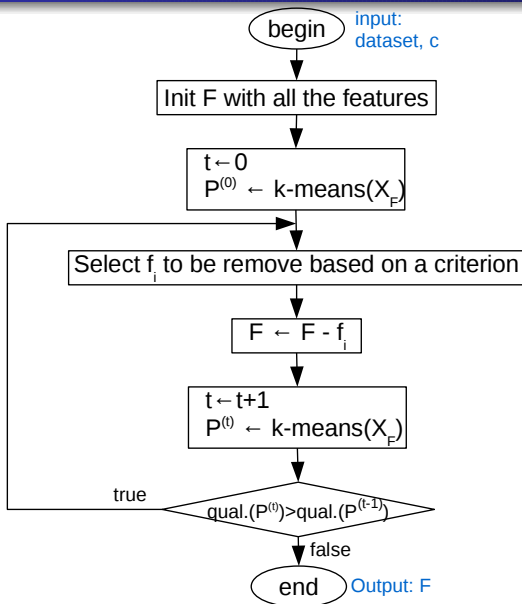
Example of Filter method : weighting features



Examples of algorithms :

- Spectral Feature Selection
- Laplacian Score

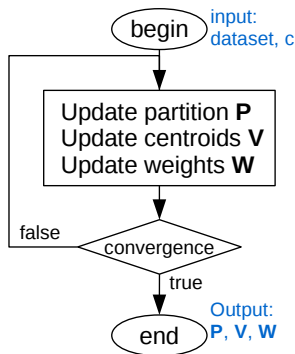
Example of Wrapper method



Example of Embedded method

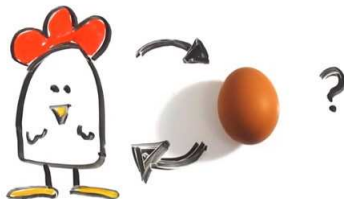
Weighting k-means

$$J_{WKM}(\mathbf{P}, \mathbf{V}, \mathbf{W}) = \sum_{i=1}^n \sum_{k=1, \mathbf{x}_i \in P_k}^c \sum_{l=1}^p w_l^\beta d(x_{il}, v_{jl})$$



Attributes selection criticisms

- Wrapper methods : the chicken and egg dilemma
 - Which one comes first : clustering or feature selection ?
- Determining optimal p selected feature is an ill-problem
- Some methods are not stable
 - e.g. to new samples
- Some methods are not scalable
 - although methods are used to reduce the dimensionality !



<http://iterated-reality.com/en/2015/03/17/the-chicken-or-the-egg-causality-dilemma-solved-by-unity-consciousness>

References I

- [1] S. Alelyani, J. Tang, and H. Liu. Feature selection for clustering : A review. *Data Clustering : Algorithms and Applications*, 29 :110–121, 2013.
- [2] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9) :1464–1480, 1990.
- [3] J. Li, K. Cheng, S. Wang, F. Morstatter, R. Trevino, J. Tang, and H. Liu. Feature selection : A data perspective. *ACM Computing Surveys (CSUR)*, 50(6) :94, 2017.
- [4] E. Lumer and B. Faieta. Diversity and adaptation in populations of clustering ants. In *Proceedings of the third international conference on Simulation of adaptive behavior*, pages 501–508. MIT Press, 1994.