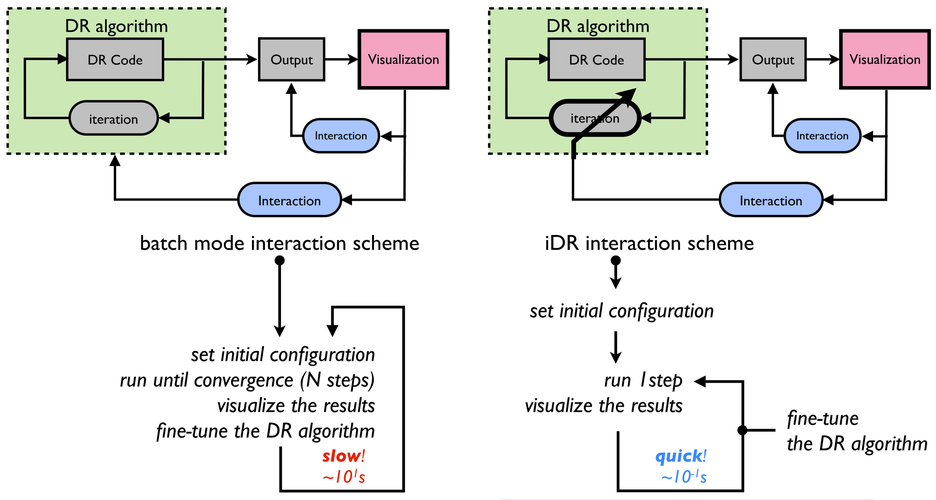
Interactive Dimensionality Reduction for Visual Analytics

Ignacio Díaz[1], Abel A. Cuadrado[1], Daniel Pérez[1],
Francisco J. García[1] and Michel Verleysen[2] *

1- Electrical Engineering Dept. University of Oviedo
Edif. Dept. 2, campus de Viesques s/n 33204, Gijón, SPAIN
2- Univ. Catholique de Louvain - Machine Learning Group
ICTEAM/ELEN - Place du Levant, 3 1348 Louvain-la-Neuve, Belgium

Abstract. In this work, we present a novel approach for data visualization based on interactive dimensionality reduction (iDR). The main idea of the paper relies on considering for visualization the intermediate results of non-convex DR algorithms under changes on the metric of the input data space driven by the user. With an appropriate visualization interface, our approach allows the user to focus on the relationships among dynamically selected groups of variables, as well as to assess the impact of a single variable or groups of variables in the structure of the data.

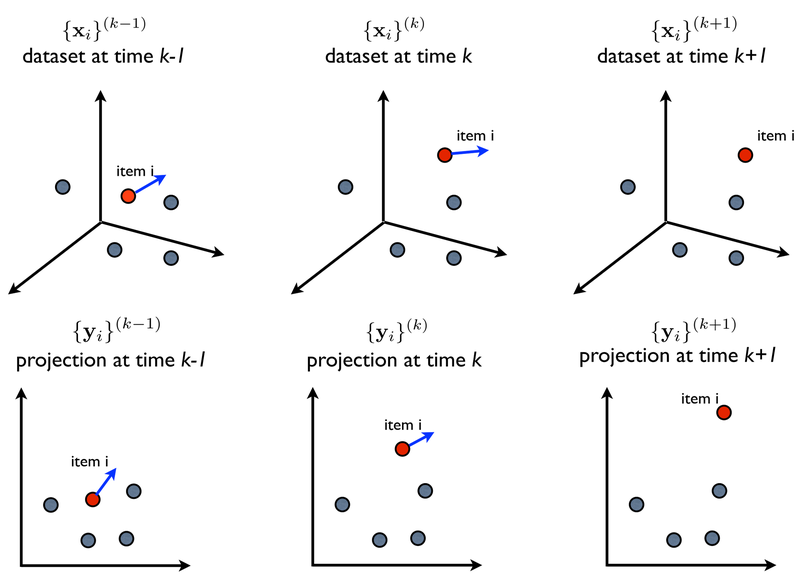
The iDR approach



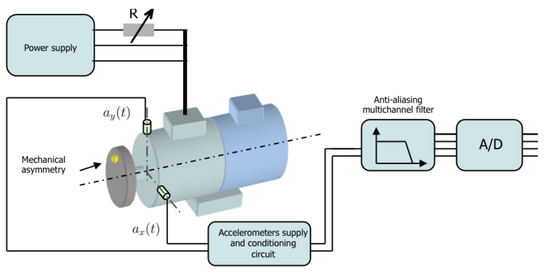
enables a quick feedback to the user and hence a much better user integration in the exploration process

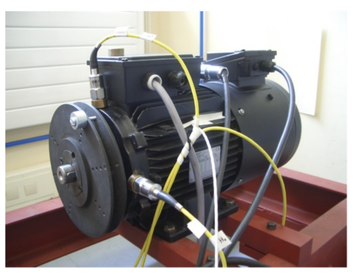
1. Analysis of time-varying input datasets

Analysis of a fixed set of samples, each one characterized by a set of measurements that evolve with time (e.g., analysis of a batch of fruits, analysis of the evolution of a set of patients on an epidemics, time evolution of social networks, etc.)



Application demo: fault analysis of AC motor





Application interface with iDR user-driven modification of the input metric space
Javascript application using processingjs (http://processingjs.org)
Analysis of three vibration signals $a_x(t)$ $a_y(t)$ $a_z(t)$ and two phase currents $i_R(t)$, $i_S(t)$
of a 4kW 2 pole-pair asynchronous motor http://isa.uniovi.es/~idiaz/demos/iDR-vibracionesMotor/

2. Changes in the metric of the input space

User-driven modification of the distance metrics allows for detection of correlations in groups of variables

$$\|\mathbf{x}\|^2_{\mathbf{\Omega}} := \sum_r \sum_s x_r \omega_{rs} x_s$$

$$\mathbf{\Omega} = \mathrm{diag}(\omega_1, \omega_2, \ldots, \omega_n)$$

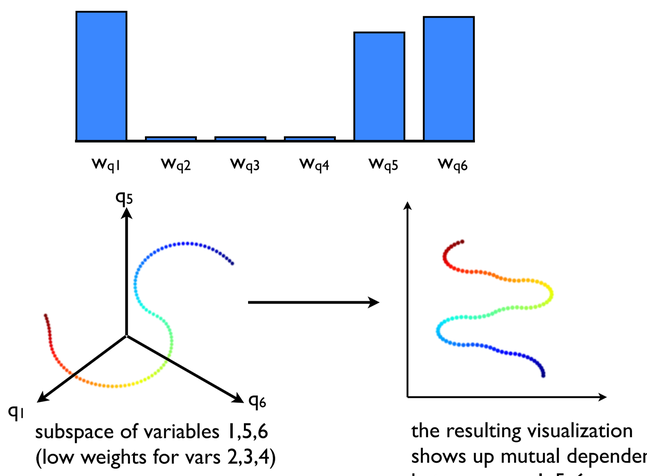
Interactive exploration of correlations
In the case of parametric dependency of this type

$$x_{q_1} = f_{q_1}(t),\ x_{q_2} = f_{q_2}(t),\ \ldots,\ x_{q_K} = f_{q_K}(t)$$

DR methods yield an easily recognizable "snake shape" figure

Weighting variables can be exploited to select subsets of variables to "test" such type of dependency, more general than linear correlation

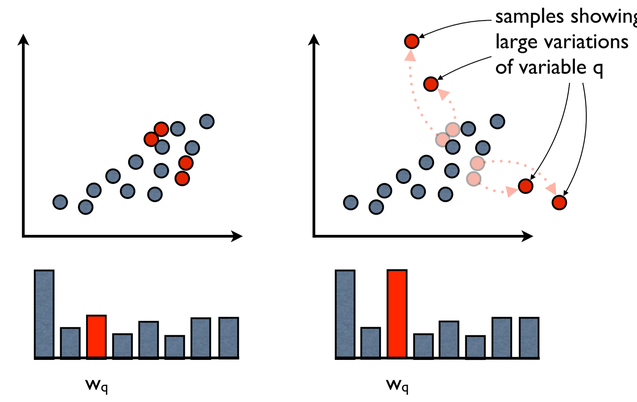
Example
sample user-driven weight variation



subspace of variables 1,5,6 (low weights for vars 2,3,4)

the resulting visualization shows up mutual dependency between vars 1, 5, 6

Sensitivity analysis
Whenever the user increases a single weight "w_q" the input distance pattern d_{ij} becomes more sensitive to variable q. Points (samples) showing significantly large deviations in variable q, will move apart upon changes in w_q, thus revealing their dependency on variable q.



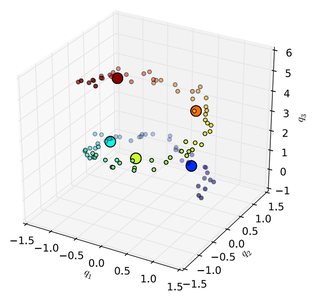
3. Interactive incorporation of class knowledge

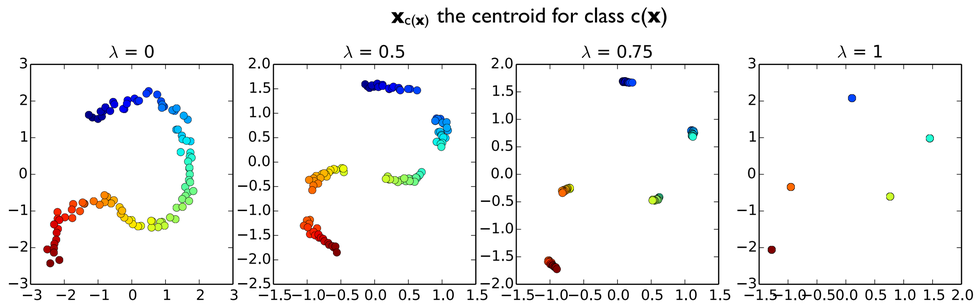
Extended data point using class info

original space	class space
$\mathbf{x}_e(\lambda) = [(1-\lambda)\mathbf{x},\ \lambda\bar{\mathbf{x}}_{c(\mathbf{x})}]$	

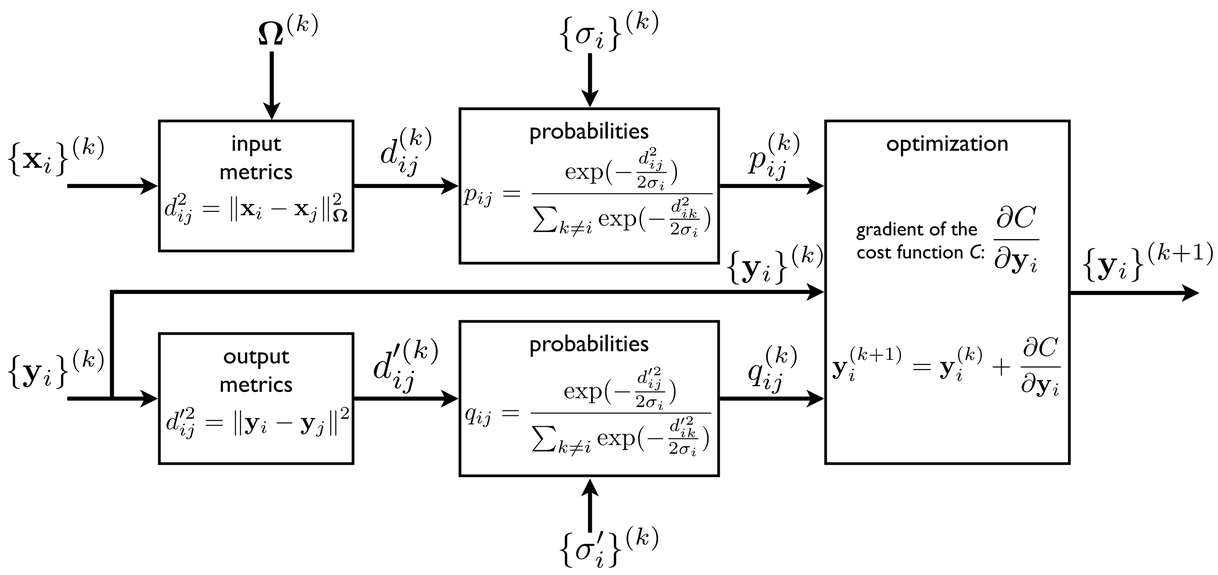
- $\mathbf{x}$ original data point in the input space
- $c(\mathbf{x})$ class to which the point $\mathbf{x}$ belongs
- λ the lambda factor allows to balance between pure class info (a set of a few centroids) and the original data

$\mathbf{x}_{c(\mathbf{x})}$ the centroid for class $c(\mathbf{x})$





iDR approach using the SNE algorithm



*The authors would like to thank financial support from the Spanish Ministry of Economy (MINECO) and FEDER funds from the EU.