Clustering models for high dimensional, temporal, and dissimilarity data

Maurizio Vichi
Dip. Statistics Probability & Applied Statistics
University "La Sapienza" Rome
maurizio.vichi@uniroma1.it

1. Introduction

Often data present a multiway structure, and they can be arranged into a Three-way Data Set X, i.e.,a set **X** of $n \times K \times T$ values related to: K variables measured (observed, estimated) on n objects (individuals, products) at T occasions (assessors, times, locations, etc.). Let X_1, X_2, ..., X_K be K quantitative variables observed on n units (objects) at T consecutive time points (Figure 1).

The observed data can be arranged into a three-way longitudinal data set

$$\mathbf{Y} \equiv \left[\mathbf{y}_{i\cdot t} = (x_{i1t}, x_{i2t}, ..., x_{ikt}, t) : i \in I, t = 1, ..., T\right]$$

where x_{ijt} is the value of the j-th variable collected on the i-th object at time t; $I=\{1,...,n\}$ $J=\{1,...,k\}$ and $U=\{1,...,T\}$ are the set of indices pertaining to objects,

variables and time points, respectively.

For each object i, $Y(i) =\{y_{i.t}$ t=1,...,$T\}$ describes a *time trajectory* of the i-th object

according to the k examined variables. The trajectory $Y(i)$ is geometrically represented by T-1 segments connecting T points $y_{i.t}$ of M^{k+1}. Two time trajectories in M^3