N-DIMENSIONAL ARCHETYPAL ANALYSIS FOR MULTIVARIATE DATA

A PREPRINT

Aleix Alcacer
Department of Mathematics
Jaume I University
Castelló de la Plana
aalcacer@uji.es

Irene Epifanio
Department of Mathematics
Jaume I University
Castelló de la Plana
epifanio@uji.es

December 3, 2023

ABSTRACT

TODO

Keywords archetypal analysis • multivariate analysis • tensor decomposition • convex optimization • non-negative tensor factorization

1 Introduction

TODO

2 Methods

TODO

2.1 Background

TODO

2.2 N-dimensional Archetypal Analysis

Let $\mathcal{X} \in \mathbb{R}^{M^{I_1 \times \cdots \times I_N}}$ be a data tensor where $\mathbf{x}_{i_1,\ldots,i_N} \in \mathbb{R}^M$ be its $(i_1, \ldots, i_N)$-th entry.

In this scenario, the archetypal tensor $\mathcal{Z} \in \mathbb{R}^{M^{K_1 \times \cdots \times K_N}}$ will be defined as convex combinations of the tensor $\mathcal{X}$, i.e. $\mathcal{Z} = \mathcal{X} \times_1 \mathcal{B}^{(1)} \times_2 \cdots \times_N \mathcal{B}^{(N)}$ where $\mathcal{B}^{(n)} \in \mathbb{R}^{K_n \times I_n}$ for all $n = 1, \ldots, N$ are stochastic matrices. Simultaneously, the tensor $\mathcal{X}$ will be approximated by convex combinations of the archetypal tensor $\mathcal{Z}$, i.e. $\mathcal{X} \simeq \mathcal{Z} \times_1 \mathcal{A}^{(1)} \times_2 \cdots \times_N \mathcal{A}^{(N)}$ where $\mathcal{A}^{(n)} \in \mathbb{R}^{I_n \times K_n}$ for all $n = 1, \ldots, N$ are also stochastic matrices.

What has been described above can be expressed as an optimization problem:

$$\underset{\mathcal{B}^{(1)},\ldots,\mathcal{B}^{(N)},\mathcal{A}^{(1)},\ldots,\mathcal{B}^{(N)}}{\arg\min} \quad \ell\left(\mathcal{X} | (\mathcal{X} \times_1 \mathcal{B}^{(1)} \times_2 \cdots \times_N \mathcal{B}^{(N)}) \times_1 \mathcal{A}^{(1)} \times_2 \cdots \times_N \mathcal{A}^{(N)}\right) \tag{1}$$

such that:

- $\ell\left(\mathcal{X}|\tilde{\mathcal{X}}\right)$ should be a loss function.
- $\sum_{k=1}^{K_n} \mathcal{A}_{ki}^{(n)} = 1$ with $\mathcal{A}_{ki}^{(n)} \in [0,1]$ for each $i = 1, \ldots, I_n$ and $n = 1, \ldots, N$.
- $\sum_{i=1}^{I_n} \mathcal{B}_{ik}^{(n)} = 1$ with $\mathcal{B}_{ik}^{(n)} \in [0,1]$ for each $k = 1, \ldots, K_n$ and $n = 1, \ldots, N$.

2.3 Related work

TODO

3 Results and discussion

TODO

3.1 Example 1: synthetic data

TODO

3.2 Example 2: real data 1

TODO

3.3 Example 3: real data 2

TODO

4 Conclusion and future work

TODO

Acknowledgements

This research was partially supported by the Spanish Ministry of Universities (FPU grant FPU20/01825).

References