Multidimensional Scaling Using Majorization

A Anish   ES20BTECH11001
Thupten Dukpa ES20BTECH11029

Introduction

- Multidimensional Scaling (MDS) is a powerful statistical technique for analyzing the structure of pairwise dissimilarity or similarity data.
- We will look into how MDS, coupled with Majorization, enhances our ability to extract meaningful insights from complex datasets.

Majorization

- The principle of majorization is to construct a surrogate function which majorizes a particular objective function.
- Let us assume we have a function f(x) to be minimized. Finding an analytical solution for complicated f(x) can be rather cumbersome/hard.
- The majorization principle suggests to find a simpler, more manageable surrogate function g(x, y) which majorizes f(x), i.e., for all x

$$g(x, y) \geq f(x)$$

y = supporting point i.e, g(y, y) = f(y)

- Let x* be minimizer of g(x,y), we get

$$f(x^*) \leq g(x^*, y) \leq g(y, y) = f(y)$$

This is called sandwich inequality.

Majorization Algorithm

Majorization is an iterative procedure

- ○ Choose $y = y^{(0)}$
- ○ Find $x^{(t)}$ such that $g(x^{(t)}, y) \leq g(y^{(0)}, y)$
- ○ Stop if $f(y) - f(x^{(t)}) < \epsilon$,

        else $y = x^{(t)}$ and proceed with the above step.

SMACOF (Scaling by Majorizing a Complicated Function)

- Let's take the input data as **Δ** (n x n matrix). **Δ** is called dissimilarities. This is based on observed data.
- Assume **Δ** is symmetric, non-negative and hollow (zero diagonal).
- Example: **Δ** can be the distances between the points (explained in class).
- We have to find a matrix **X** in p-dimensional space such that the distances between the points approximate to the given dissimilarities δ_{ij}.

$$d_{ij}(X) = \sqrt{\sum_{s=1}^{p}(x_{is} - x_{js})^2}$$

- We want to find X such that

$$d_{ij}(X) \approx \delta_{ij}$$

- Elements of **X** are called *configurations* of the objects.
- Each object is scaled in a p-dimensional space such that the distances between the points in the space match as well as possible to the observed dissimilarities.

Simple SMACOF

- Define Stress $\sigma(X)$

$$\sigma(X) = \sum_{i<j} w_{ij}(\delta_{ij} - d_{ij}(X))^2$$

- W is a known n x n matrix of weights, also assumed to be symmetric, non-negative, and hollow.
- WLOG

$$\sum_{i<j} w_{ij}\delta_{ij}^2 = \frac{n(n-1)}{2}$$

- Expanding the stress function we get the following

$$\sigma(X) = \sum_{i<j} w_{ij}\delta_{ij}^2 + \sum_{i<j} w_{ij}d_{ij}^2(X) - 2\sum_{i<j} w_{ij}\delta_{ij}d_{ij}(X)$$

Simple SMACOF

- Let us define a few matrices

$$A_{ij} = (e_i - e_j)(e_i - e_j)^T$$

$$V = \sum_{i<j} w_{ij} A_{ij}$$

$$B(X) = \sum_{i<j} w_{ij} s_{ij}(X) A_{ij}$$

$$s_{ij}(X) = \begin{cases} \delta_{ij}/d_{ij}(X) & d_{ij}(X) > 0 \\ 0 & d_{ij}(X) = 0 \end{cases}$$

Simple SMACOF

Using these matrices we can write the terms of $\sigma(X)$ as

$$\sum_{i<j} w_{ij} d_{ij}^2(X) = \mathbf{tr}\left(X^T V X\right)$$

$$\sum_{i<j} w_{ij} \delta_{ij} d_{ij}(X) = \mathbf{tr}\left(X^T B(X) X\right)$$

Simple SMACOF

- After using these matrices, we get the final equation as

$$\sigma(X) = n(n-1)/2 + \mathbf{tr}X^T V X - 2\mathbf{tr}X^T B(X) X$$

- Now majorizing σ(X).

$$\sigma(X) = n(n-1)/2 + \mathbf{tr}X^T V X - 2\mathbf{tr}X^T B(X)X$$
$$\leq n(n-1)/2 + \mathbf{tr}X^T V X - 2\mathbf{tr}X^T B(Y)Y = \tau(X,Y)$$

- Finding minimizer of τ(X,Y)

$$\frac{\partial \tau(X,Y)}{\partial X} = 2VX - 2B(Y)Y = 0$$

Simple SMACOF

- We get ,

$$VX = B(Y)Y$$

- Since majorization is a iterative process, in step t = 0 we set Y := X$^{(0)}$ where X$^{(0)}$ is a start configuration. Within each iteration t we compute X$^{(t)}$ which, for simple SMACOF, gives us the update X$^{(t)}$. Now the stress σ(X$^{(t)}$) can be calculated and we stop iterating if

$$\sigma(X^{(t)}) - \sigma(X^{(t-1)}) < \epsilon$$

Extension to SMACOF

- SMACOF version with restrictions on the configuration matrix X, this is called confirmatory MDS with external constraints.
- Simplest restriction is a linear combination.
- The linear restriction is of the form

$$X = ZC$$

Z = predictor Matrix (n x q) (q>p)

C is a q x p matrix of regression weights to be estimated

- The optimization problem is basically the same as in simple SMACOF.
- $\boldsymbol{\tau}$(X,Y) can be written as

$$\tau(X,Y) = n(n-1)/2 + \mathbf{tr}\, X^T V X - 2\mathbf{tr}\, X^T V \overline{X}$$
$$= n(n-1)/2 + \mathbf{tr}(X - \overline{X})^T V (X - \overline{X}) - \mathbf{tr}\, \overline{X}^T V \overline{X}$$

SMACOF with Restrictions

- The second term denotes the lack of confirmation fit and becomes zero if no restrictions are imposed.
- In each iteration t of the majorization algorithm we first compute the Guttman transform X_bar$^{(t)}$ of our current best configuration, and then solve the configuration projection problem of the form
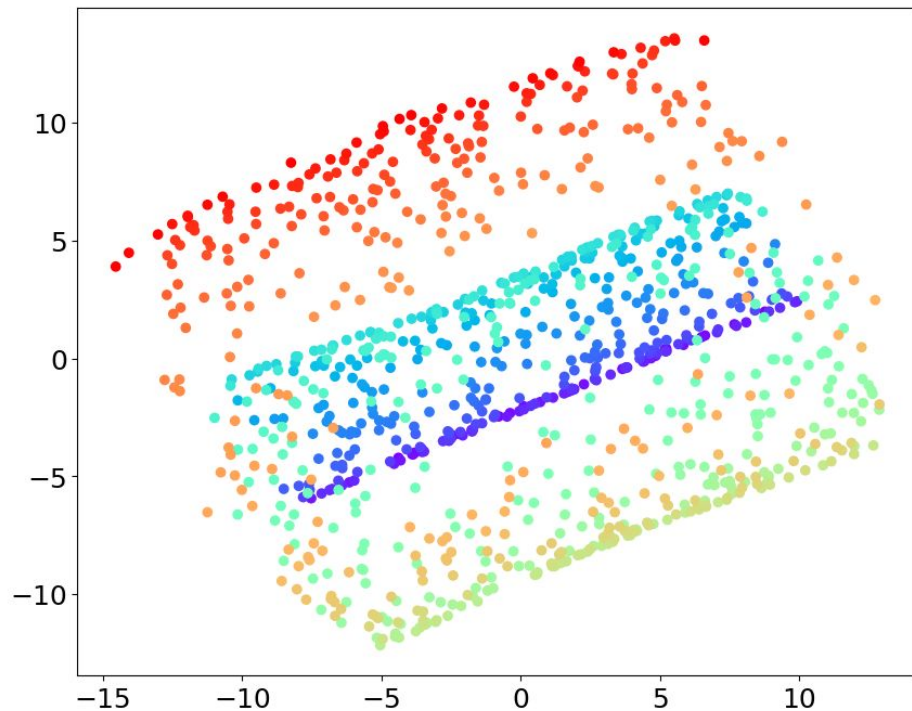
$$min \ \mathbf{tr}(X - \overline{X}^{(t)})^T V (X - \overline{X}^{(t)})$$
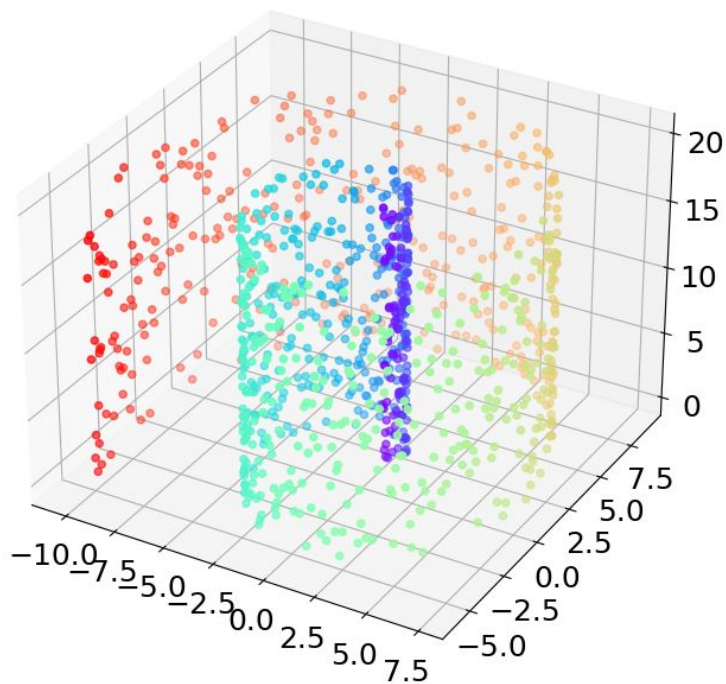
- We project X_bar$^{(t)}$ on the manifold of constrained configurations.

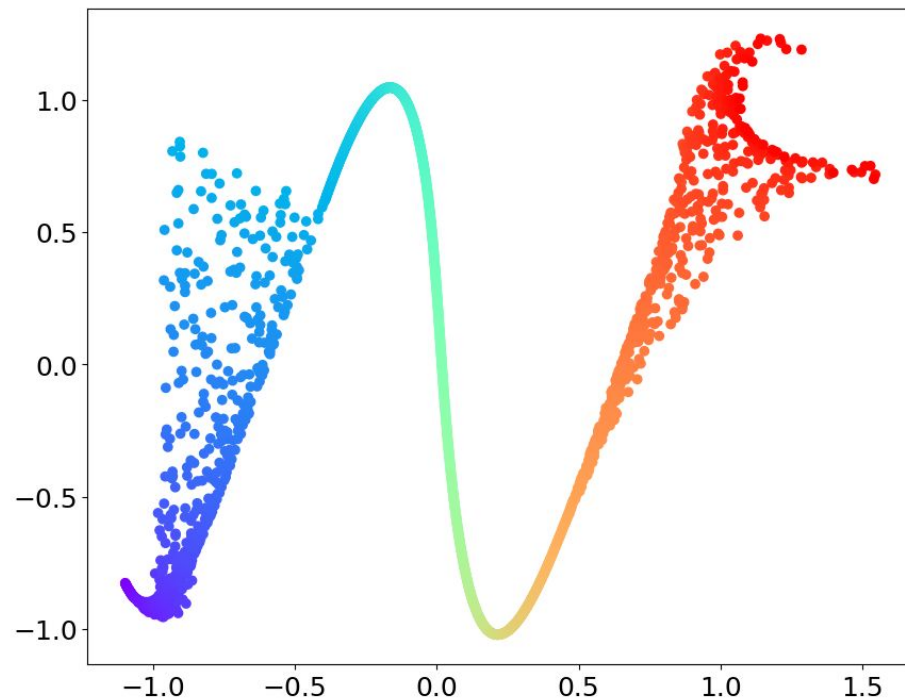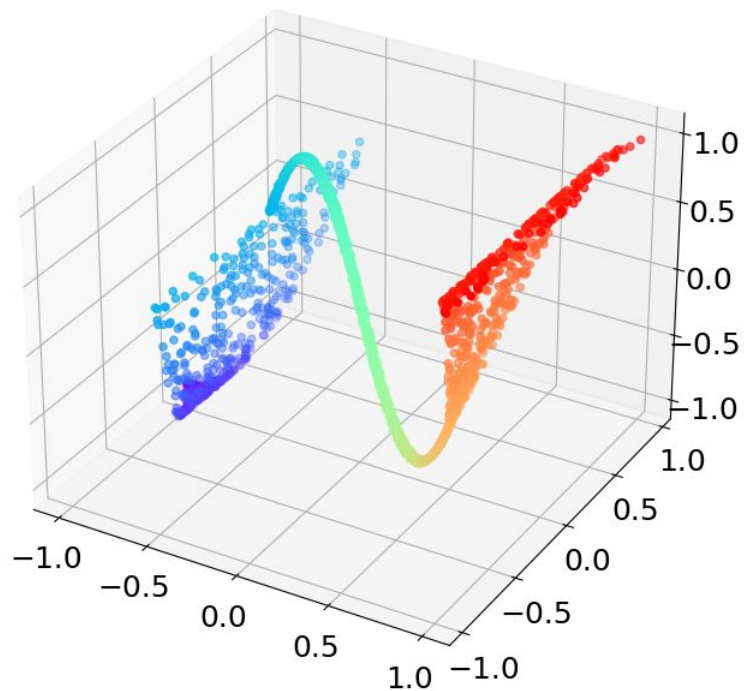$$X^{(t)} = ZC^{(t)} = Z(Z^T V Z)^{-1} Z^T V \overline{X}^{(t)}$$

- With linear restrictions this projection gives us the update, also checking if $\quad \sigma(X^{(t+1)})$ -$\sigma(X^{(t)})$ < ϵ

Code Results

Swiss Roll Dataset

Arrow Dataset

Advantages of MDS with Majorization

- *Improved Convergence:*
  - Faster and more stable convergence of the optimization algorithm.
  - Beneficial when dealing with complex dissimilarity matrices.
- *Global Optimality:*
  - More likely to converge to the global minimum of the stress function, providing more accurate representation of underlying structure in the data.
- *Handling Missing Data:*
  - The initial weights w_{ij} can be taken to be 0 if δ_{ij} is missing.

Applications

- Geographic Mapping:
  - Two-dimensional graphical representation of cities using airline distances between them as symmetric input matrix.
- Genome Analysis:
  - MDS is applied to high-throughput chromosome conformation capture(Hi-C) data on genomic interactions to visualize DNA loops.
- Information Retrieval and Document Analysis:
  - It can assist in clustering and visualizing documents based on semantic similarity.
- Healthcare and Medicine:
  - To visualize and analyze the structural relationships between chemical compounds, aiding in drug discovery and design.

Limitations

- Limited Robustness to Outliers:
  - MDS algorithms, including majorization, can be sensitive to outliers in the data.
- Sensitivity to Initialization.
  - In some cases, different initializations may lead to different final solutions.
- Lack of scalability:
  - Majorization might be computationally expensive for very large datasets.

Limitations

- Jan de Leeuw and Patrick Mair, "Multidimensional Scaling Using Majorization: SMACOF in R",August 2009.
- Patrick J. F. Groenen, Jan de Leeuw and Patrick Mair, "Multidimensional Scaling in R: SMACOF".

Thank You