Visualization
CH10: Multidimensional Scaling (MDS)

Jong-June Jeon

University of Seoul, Department of Statistics

Contents

Introduction

Motivational Example: Perceptual Distances

- Suppose we conduct a psychological experiment where participants are asked to rate the similarity between pairs of objects, such as:
  - Musical genres
  - Flavors of soft drinks
  - Animal species
- The result is a distance matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$, where D_{ij} represents how dissimilar object i is from object j.
- However, the objects themselves do not have explicit coordinates in Euclidean space.
- **Question:** Can we find a low-dimensional representation that reflects these perceptual distances?

Motivational Example: Perceptual Distances

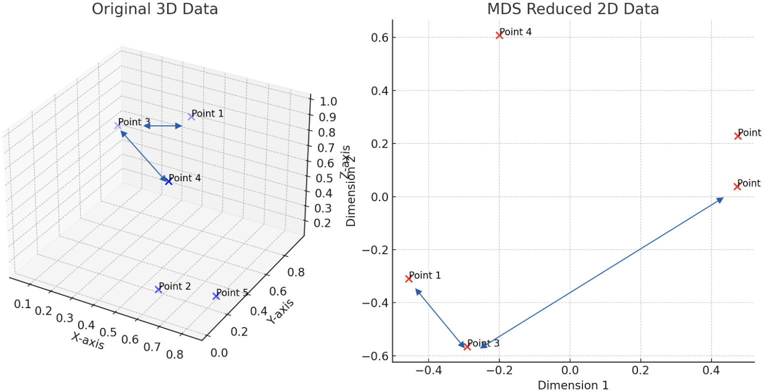

Figure: Illustration of MDS

Motivating Example 1: Musical Genres

- Participants rate the similarity between pairs of musical genres (e.g., jazz, rock, hip-hop, classical).
- These ratings produce a distance matrix $\mathbf{D}$, without any explicit feature vectors for the genres.
- **Goal:** Recover a low-dimensional spatial map reflecting perceived similarities.
- **Classical MDS:** Embeds genres in $\mathbb{R}^2$ or $\mathbb{R}^3$, preserving pairwise distances.
- **Insight:** Genres like rock and metal may appear close together, while classical and hip-hop may be distant.

Motivating Example 2: Flavors of Soft Drinks

- A sensory test asks consumers to compare flavors of soft drinks.
- The results form a dissimilarity matrix based on perceived taste differences.
- **Challenge:** Taste profiles are not easily quantifiable in feature space.
- **Classical MDS:** Provides a perceptual flavor map based on consumer similarity judgments.
- **Use Case:** Market positioning of products; e.g., Sprite and 7-Up are mapped close, while Coke and Dr. Pepper are further apart.

Motivating Example 3: Animal Species

- Participants are asked how similar different animal species are (e.g., dog, wolf, penguin, elephant).
- This results in a subjective distance matrix with no predefined coordinates.
- **Goal:** Understand how humans perceive biological similarity.
- **Classical MDS:** Embeds species into a spatial layout that reflects perceptual clustering.
- **Example:** Dog and wolf are nearby; penguin appears distant from terrestrial mammals.

Summary: What Problem Does MDS Solve?

Example	Problem	MDS Solution
Musical Genres	No coordinates, only pairwise similarity	Find spatial layout of genres
Soft Drink Flavors	Sensory data without numeric features	Visualize perceptual flavor space
Animal Species	Subjective similarity only	Map perceived biological proximity

- In all cases: **Distances are known, but coordinates are not.**
- **Classical MDS** reconstructs coordinates that best preserve the distance structure.

Classical MDS

Problem Setting

- Let $\mathbf{x}_i = (x_{i1}, \cdots, x_{id})^\top \in \mathbb{R}^d$ be the observed data point for object i.
- The data matrix is $\mathbf{X} \in \mathbb{R}^{n \times d}$, where each row represents an object.
- The pairwise Euclidean distance matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ is defined as:

$$(\mathbf{D})_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2 = (\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j)$$

- Goal: Find a low-dimensional representation $\mathbf{z}_i \in \mathbb{R}^k$ with $k \ll d$ such that distances are preserved:

$$\|\mathbf{z}_i - \mathbf{z}_j\|^2 \approx \|\mathbf{x}_i - \mathbf{x}_j\|^2$$

Distance and Inner Product

- Let $\mathbf{x}_i \in \mathbb{R}^d$ and $\mathbf{x}_j \in \mathbb{R}^d$. The Euclidean distance between them is:

$$d_{ij} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j)}$$

- The squared distance can be written as:

$$d_{ij}^2 = \|\mathbf{x}_i\|^2 + \|\mathbf{x}_j\|^2 - 2\mathbf{x}_i^\top \mathbf{x}_j$$

- $\sum_{j=1}^n d_{ij}^2$ and $\sum_{i=1}^n d_{ij}^2$ are quantities representing **scaling factors** of the distance matrix.
- $\mathbf{x}_i^\top \mathbf{x}_j$ for i and j are corresponds to **structural factor** of the distance matrix.

Distance and Inner Product

Goal: Find a low-dimensional representation $\mathbf{z}_i \in \mathbb{R}^k$ with $k \ll d$ such that distances are preserved:

$$\|\mathbf{z}_i - \mathbf{z}_j\|^2 \approx \|\mathbf{x}_i - \mathbf{x}_j\|^2$$

in the sense of preserving the structural properties of the distance matrix.

$$\mathbf{z}_i^\top \mathbf{z}_j \approx \mathbf{x}_i^\top \mathbf{x}_j$$

on average. This is the motivation of introducing the doubly centered distance matrix. Next we'll see how the scaling factor is excluded in the distance approximation procedure.

Matrix Form of Squared Distances

- Let $\mathbf{D} \in \mathbb{R}^{n \times n}$ be the matrix where $(\mathbf{D})_{ij} = d_{ij}^2$.
- Then, $\mathbf{D}$ can be expressed as:

$$\mathbf{D} = M_1 + M_2 - 2\mathbf{X}\mathbf{X}^\top$$

where:

$$(M_1)_{ij} = \|\mathbf{x}_i\|^2 \quad \text{for all } j \quad \text{(row-wise constant)}$$
$$(M_2)_{ij} = \|\mathbf{x}_j\|^2 \quad \text{for all } i \quad \text{(column-wise constant)}$$

- The term $(\mathbf{X}\mathbf{X}^\top)_{ij} = \mathbf{x}_i^\top \mathbf{x}_j$ is the inner product of data vectors $\mathbf{x}_i$ and $\mathbf{x}_j$.
- For notational convenience, we may denote d_{ij}^2 simply as d_{ij} in the following discussion.

Projection Matrix for Centering

- Define the projection matrix onto the space spanned by the constant vector $\mathbf{1}$:

$$\mathbf{\Pi}_{\mathcal{C}(1)} = \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$$

- The centering matrix is:

$$\mathbf{H} = \mathbf{I} - \mathbf{\Pi}_{\mathcal{C}(1)}$$

  which projects vectors onto the orthogonal complement of $\mathrm{span}(\mathbf{1})$.

For convenience, denote $\mathbf{\Pi}_{\mathcal{C}(1)}$ by $\mathbf{\Pi}_1$.

Double-Centering a Distance Matrix D

Assume $\mathbf{D}$ is a symmetric matrix of squared Euclidean distances.

1. $\mathbf{D}(\mathbf{I} - \mathbf{\Pi}_1)$: subtracts the column means of $\mathbf{D}$

$$d_{ij} \to d_{ij} - \bar{d}_{\cdot j}$$

2. Since $\mathbf{D}$ is symmetric, we have:

$$(\mathbf{I} - \mathbf{\Pi}_1)\mathbf{D} = \mathbf{D}(\mathbf{I} - \mathbf{\Pi}_1)$$

3. Applying centering from both sides:

$$(\mathbf{I} - \mathbf{\Pi}_1)\mathbf{D}(\mathbf{I} - \mathbf{\Pi}_1)$$

   This removes both row and column means from $\mathbf{D}$, resulting in:

$$d_{ij} - \bar{d}_{i\cdot} - \bar{d}_{\cdot j} + \bar{d}_{\cdot\cdot}$$

   where:
   - $\bar{d}_{i\cdot}$: row mean
   - $\bar{d}_{\cdot j}$: column mean
   - $\bar{d}_{\cdot\cdot}$: grand mean

Double-Centering Formula

The double-centered distance matrix is:

$$(\mathbf{D}')_{ij} = d_{ij} - \bar{d}_{i\cdot} - \bar{d}_{\cdot j} + \bar{d}_{\cdot\cdot}$$

This transformation is equivalent to:

$$\mathbf{D}' = (\mathbf{I} - \mathbf{\Pi}_1)\mathbf{D}(\mathbf{I} - \mathbf{\Pi}_1)$$

Double-Centering: Matrix Decomposition

The relationship between $\mathbf{D}'$ and $\mathbf{X}\mathbf{X}^\top$: Recall that $\mathbf{D}$ is a squared Euclidean distance matrix,

$$\mathbf{D} = M_1 + M_2 - 2\mathbf{X}\mathbf{X}^\top$$

Apply double-centering:

$$\mathbf{D}' = (\mathbf{I} - \mathbf{\Pi}_1)\mathbf{D}(\mathbf{I} - \mathbf{\Pi}_1) = (\mathbf{I} - \mathbf{\Pi}_1)(M_1 + M_2 - 2\mathbf{X}\mathbf{X}^\top)(\mathbf{I} - \mathbf{\Pi}_1)$$

Consider the following simplifications:

1. $M_1(\mathbf{I} - \Pi_1) = M_1 - M_1\Pi_1 = 0$ (M_1 is row-wise constant)
2. $(\mathbf{I} - \Pi_1)M_2 = M_2 - \Pi_1 M_2 = 0$ (M_2 is column-wise constant)
3. $(\mathbf{I} - \Pi_1)\mathbf{X}\mathbf{X}^\top(\mathbf{I} - \Pi_1) = (\mathbf{I} - \Pi_1)\mathbf{X}((\mathbf{I} - \Pi_1)\mathbf{X})^\top = \mathbf{X}\mathbf{X}^\top$ uf $\mathbf{X}$ is column-wisely centered.
   (Note that distance matrix is invariant to the locational-shift. Thus, we can assume the centered $\mathbf{X}$ without loss of generality.)

Centered Inner Product

- From the previous steps:

$$\mathbf{D}' = (\mathbf{I} - \Pi_1)\mathbf{D}(\mathbf{I} - \Pi_1) = -2\mathbf{X}\mathbf{X}^\top$$

- Thus, the double-centered distance matrix yields the (scaled) Gram matrix:

$$\mathbf{B} := -\frac{1}{2}(\mathbf{I} - \Pi_1)D(\mathbf{I} - \Pi_1) = \mathbf{X}\mathbf{X}^\top$$

- $\mathbf{B}$ is the inner product matrix of the centered data, and is the foundation of classical MDS.

Note: In MDS, it is not necessary to observe $\mathbf{X}$. Only the distance matrix $\mathbf{D}$ is required.

Classical MDS: Double-Centering Perspective

- Let $\mathbf{Z} \in \mathbb{R}^{n \times k}$ be the low-dimensional latent configuration such that:

$$(\mathbf{I} - \Pi_1)\mathbf{Z} = \mathbf{Z} \quad \text{(i.e., } \mathbf{Z} \text{ is centered)}$$

- Let $\widetilde{\mathbf{D}}$ be the distance matrix computed from $\mathbf{Z}$:

$$(\widetilde{\mathbf{D}})_{ij} = \|\mathbf{z}_i - \mathbf{z}_j\|^2$$

- Then, applying double-centering to $\widetilde{\mathbf{D}}$, we obtain:

$$\widetilde{\mathbf{D}}' = (\mathbf{I} - \Pi_1)\widetilde{\mathbf{D}}(\mathbf{I} - \Pi_1) = -2\mathbf{Z}\mathbf{Z}^\top$$

Classical MDS: Optimization Formulation

Basic objective function of MDS

$$\min_{\mathbf{Z} \in \mathbb{R}^{n \times k}} \left\| \mathbf{D}' - \widetilde{\mathbf{D}}' \right\|_F^2 ,$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix. The objective function denotes a distance betweeem the double centered distance matrices on the original space and the reduced dimensional space.

Interpretation of MDS obj ftn: Frobenius Norm of a Matrix

Let $A \in \mathbb{R}^{n \times m}$ be a matrix.

Frobenius norm

$$\|A\|_F = \sqrt{\sum_{i,j} (A)_{ij}^2} = \sqrt{\mathrm{tr}(A^\top A)} = \sqrt{\sum_{i=1}^{\min(n,m)} \sigma_i^2(A)}$$

where $\sigma_i(A)$ are the singular values of A (See the SVD).

Interpretation of MDS obj ftn: Frobenius Norm Decomposition via Projection

Let $\Pi_1 = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ and $\mathbf{I} - \Pi_1$ be the centering projection. Then:

$$\|(\mathbf{I} - \Pi_1)A + \Pi_1 A\|_F^2 = \operatorname{tr}\left[((\mathbf{I} - \Pi_1)A + \Pi_1 A)^\top ((\mathbf{I} - \Pi_1)A + \Pi_1 A)\right]$$

$$= \operatorname{tr}\left[((\mathbf{I} - \Pi_1)A)^\top (\mathbf{I} - \Pi_1)A + (\Pi_1 A)^\top (\Pi_1 A)\right]$$

$$= \|(\mathbf{I} - \Pi_1)A\|_F^2 + \|\Pi_1 A\|_F^2$$

- Interpretation: The Frobenius norm is preserved under orthogonal decomposition.
- Application: Useful for analyzing total variance in centered vs. mean components.

Interpretation of MDS obj ftn: Difference Between Distance Matrices

- Let D and $\widetilde{D}$ be two squared distance matrices. Then:

$$\|D - \widetilde{D}\|_F^2 = \|(\mathbf{I} - \Pi_1)(D - \widetilde{D})\|_F^2 + \|\Pi_1(D - \widetilde{D})\|_F^2$$

- The first term is the projection onto the orthogonal complement of $\mathbf{1}$, and is directly related to the MDS objective:

$$\left\|\mathbf{D}' - \widetilde{\mathbf{D}}'\right\|_F^2 = \|(\mathbf{I} - \Pi_1)(\mathbf{D} - \widetilde{\mathbf{D}})\|_F^2 = 4\|\mathbf{X}\mathbf{X}^\top - \mathbf{Z}\mathbf{Z}^\top\|_F^2$$

Interpretation of MDS obj ftn: Scale Invariance of MDS

- The second term, corresponding to the mean of distances, is the scaling factor of distance matrix:

$$\|\Pi_1(D - \widetilde{D})\|_F^2 = \sum_i n \left(\bar{d}_i^2 - \bar{d}_i'^{\,2}\right)^2$$

  where:

$$\bar{d}_i^2 = \frac{1}{n}\sum_{j=1}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|^2, \quad \bar{d}_i'^{\,2} = \frac{1}{n}\sum_{j=1}^{n} \|\mathbf{z}_i - \mathbf{z}_j\|^2$$

- These terms depend on the scale (magnitude) of the data but are not part of the MDS objective.
- Therefore, MDS captures only the structural geometry (e.g., relative positions), not absolute distances.

Interpretation of MDS obj ftn: Optimization Formulation

- Therefore, the classical MDS objective $\min_{\mathbf{Z}\in\mathbb{R}^{n\times k}} \left\|\mathbf{D}' - \tilde{\mathbf{D}}'\right\|_F^2$ is to match the centered inner product matrices:

$$\min_{\mathbf{Z}\in\mathbb{R}^{n\times k}} \left\|(\mathbf{I}-\Pi_1)\mathbf{D}(\mathbf{I}-\Pi_1) - (\mathbf{I}-\Pi_1)\widetilde{\mathbf{D}}(\mathbf{I}-\Pi_1)\right\|_F^2$$

- Since:

$$(\mathbf{I}-\Pi_1)\mathbf{D}(\mathbf{I}-\Pi_1) = -2\mathbf{X}\mathbf{X}^\top, \quad (\mathbf{I}-\Pi_1)\widetilde{\mathbf{D}}(\mathbf{I}-\Pi_1) = -2\mathbf{Z}\mathbf{Z}^\top,$$

  the problem becomes:

$$\min_{\mathbf{Z}\in\mathbb{R}^{n\times k}} 4\left\|\mathbf{X}\mathbf{X}^\top - \mathbf{Z}\mathbf{Z}^\top\right\|_F^2$$

- This is a low-rank approximation problem with orthogonality implied for the columns of $\mathbf{Z}$.

Training MDS: Distance Matrix and Double Centering

- Define the centering matrix:
$$\mathbf{H} = \mathbf{I}_n - \mathbf{\Pi}_{\mathcal{C}(1)}$$
- From squared distance matrix $\mathbf{D}$, define the inner product matrix $\mathbf{B}$:
$$\mathbf{B} = -\frac{1}{2}\mathbf{HDH}$$

Connection to Low-dimensional Embedding

- We compute eigen-decomposition of $\mathbf{B}$:

$$\mathbf{B} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$$

- Low-dimensional embedding is given by:

$$\mathbf{Z} = \mathbf{V}_k\boldsymbol{\Lambda}_k^{1/2}$$

- Then, the squared distances in the new space are:

$$\|\mathbf{z}_i - \mathbf{z}_j\|^2 = B_{ii} + B_{jj} - 2B_{ij}$$

  which approximates the original D_{ij}

- Thus, double centering enables recovery of dot products from pairwise distances, which makes distance-preserving embedding possible.

Low-dimensional Embedding

- Perform eigen-decomposition of $\mathbf{B}$:

$$\mathbf{B} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$$

  where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$
- Define the low-dimensional coordinates:

$$\mathbf{Z} = \mathbf{V}_k\boldsymbol{\Lambda}_k^{1/2} \in \mathbb{R}^{n\times k}$$

- The rows of $\mathbf{Z}$ are the coordinates in $\mathbb{R}^k$ that best preserve the pairwise distances.

Objective Functions in MDS

Strain Function: (used in Classical MDS)

$$\text{Strain}_D(\mathbf{z}_1, \ldots, \mathbf{z}_n) = \sqrt{\frac{\|\mathbf{X}\mathbf{X}^\top - \mathbf{Z}\mathbf{Z}^\top\|_F^2}{\|\mathbf{X}\mathbf{X}^\top\|_F^2}}$$

- Measures the difference between the original inner product matrix and the low-dimensional one.
- Optimized directly in classical MDS via eigen-decomposition.

Stress Function (Used in Metric MDS)

Stress Function:

$$\text{Stress}_D(\mathbf{z}_1, \ldots, \mathbf{z}_n) = \sqrt{\frac{\sum_{i,j} \left(d_{ij} - \|\mathbf{z}_i - \mathbf{z}_j\|\right)^2}{\sum_{i,j} d_{ij}^2}}$$

- Measures the discrepancy between original distances d_{ij} and embedded distances.
- Used in **metric MDS**, optimized using iterative methods (e.g., SMACOF).
- More flexible, accommodates non-Euclidean input distances.

Mapping with Isometric Transformation

- The low-dimensional representation $\mathbf{Z} \in \mathbb{R}^{n \times k}$ can be written as:

$$\mathbf{Z} = \mathbf{V}_k \mathbf{\Lambda}_k^{1/2} \mathbf{R}$$

  where:
  - $\mathbf{V}_k$: eigenvectors (principal directions)
  - $\mathbf{\Lambda}_k$: top k eigenvalues (diagonal matrix)
  - $\mathbf{R}$: rotation matrix (orthogonal transformation)
- Both rotation and translation preserve Euclidean distances — this is called an **isometric transformation.**

Distance Preservation

1. Inner product matrix is invariant under rotation:

$$\mathbf{Z}\mathbf{Z}^\top = (\mathbf{V}_k \boldsymbol{\Lambda}_k^{1/2} \mathbf{R})(\mathbf{V}_k \boldsymbol{\Lambda}_k^{1/2} \mathbf{R})^\top = \mathbf{V}_k \boldsymbol{\Lambda}_k \mathbf{V}_k^\top$$

2. Translation of $\mathbf{Z}$ does not change pairwise distances:

$$\|\mathbf{z}_i + \mathbf{c} - (\mathbf{z}_j + \mathbf{c})\| = \|\mathbf{z}_i - \mathbf{z}_j\|$$

   for any constant vector $\mathbf{c} \in \mathbb{R}^k$.

Conclusion: The geometry of the embedded space is preserved under orthogonal and translational transformations.

Recovering Geographic Positions

- The coordinates from MDS can be rotated and aligned with actual geographic maps using Procrustes alignment.
- This allows for an approximate reconstruction of the cities' locations.
- **Interpretation:** Axes in MDS have no inherent meaning, but after rotation, they can align with real-world longitude and latitude.
- This mapping demonstrates how structural information (distances) alone can recover meaningful spatial configurations.

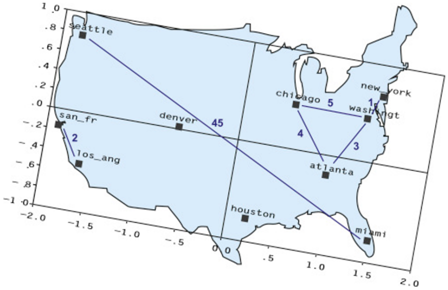

Figure: source: Krabbe, Paul. The measurement of health and health status: concepts, methods and applications from a multidisciplinary perspective. Academic Press, 2016.

Nonmetric MDS

From Inner Product to Dissimilarity

- Classical MDS solves:
$$\min_{\mathbf{Z}} \left\| \mathbf{X}\mathbf{X}^\top - \mathbf{Z}\mathbf{Z}^\top \right\|_F^2$$
- The matrix $\mathbf{X}\mathbf{X}^\top$ contains centered inner products:
$$\mathbf{x}_i^\top \mathbf{x}_j = \|\mathbf{x}_i\| \|\mathbf{x}_j\| \cos\theta$$
- These inner products encode information about dissimilarities between variables or observations.

Generalized MDS Problem

- Given a dissimilarity matrix $[\delta_{ij}]$ defined in a high-dimensional space,
- Find low-dimensional points $\mathbf{z}_1, \ldots, \mathbf{z}_n \in \mathbb{R}^k$ such that:

$$\|\mathbf{z}_i - \mathbf{z}_j\|^2 \approx \delta_{ij}$$

- That is, distances in the embedded space approximate the given dissimilarities.
- This formulation applies even when no explicit coordinates $\mathbf{x}_i$ exist, only pairwise dissimilarities.

Why Non-metric MDS? [Agarwal, 2007]

- Suppose we want to visualize the perceptual similarity of visual stimuli.
- Participants are asked to rate the similarity between objects on a scale from 1 to 10.
- Although these scores can be used for embedding and visualization, the internal criteria used by participants may vary.
- For example, evaluations may depend on the order in which stimuli are presented.

Ordinal Data and Robust Representation

- Absolute values provided by users are often unreliable.
- However, the relative orderings (ranking of distances or similarities) tend to be more consistent.
- **Question:** Can we develop an MDS method that only uses the ordering of dissimilarities rather than their absolute magnitudes?
- **Answer:** Yes — this leads to the formulation of **Non-metric MDS**.