STA 437/2005: Methods for Multivariate Data

Week 9: Non-linear Dimension Reduction Techniques

Piotr Zwiernik

University of Toronto

Why Principal Component Analysis may not be enough?

Why go beyond PCA?
PCA captures variance through linear projections but struggles with:

- Non-linear relationships.
- Complex manifolds.

We explore MDS, UMAP and its relationship with PCA.

Multi-dimensional Scaling (MDS)

Problem Setup

Consider a dissimilarity matrix $\Delta = (\delta_{ij}) \in \mathbb{R}^{n \times n}$: $\delta_{ii} = 0$ for all i, $\delta_{ij} \geq 0$ for all $i \neq j$.

In classical MDS: there exist $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$ such that $\delta_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$.

In general we have n objects and δ_{ij} is a measure of their dissimilarity (small if similar). There need not a Euclidean distance defining this metric.

Classical MDS: $\delta_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$

Classical MDS: $\delta_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$

If $\delta_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$, we have:

$$\delta_{ij}^2 \;=\; (\mathbf{x}_i - \mathbf{x}_j)^\top(\mathbf{x}_i - \mathbf{x}_j) \;=\; (\mathbf{X}\mathbf{X}^\top)_{i,i} + (\mathbf{X}\mathbf{X}^\top)_{j,j} - 2(\mathbf{X}\mathbf{X}^\top)_{i,j}.$$

The Hadamard product $\Delta \odot \Delta = [\delta_{ij}^2]$ can be written as:

$$\Delta \odot \Delta = \mathrm{diag}(\mathbf{X}\mathbf{X}^\top)\mathbf{1}\mathbf{1}^\top + \mathbf{1}\mathbf{1}^\top\mathrm{diag}(\mathbf{X}\mathbf{X}^\top) - 2\mathbf{X}\mathbf{X}^\top$$

Reintroducing the centering matrix $H = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$, we obtain

$$B \;:=\; -\frac{1}{2}H(\Delta \odot \Delta)H \;=\; H\mathbf{X}(H\mathbf{X})^\top \;=\; \tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top.$$

This matrix contains all inner products $\langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle$.

Centering the Distance Matrix

$$H = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$$

$$B = -\frac{1}{2}H(\Delta \odot \Delta)H$$

Eigen-decomposition for Dimensionality Reduction

$$B = V\Lambda V^\top$$

$$\mathbf{Y} = U_d \Lambda_d^{1/2}$$

U_d: top d eigenvectors. Λ_d: top d eigenvalues.

Duality Between MDS and PCA

Classical MDS and PCA are closely connected. Here is the key insight:

- **PCA**: Finds principal components from the eigenvectors of $(H\mathbf{X})^\top H\mathbf{X}$.
- **MDS**: Finds embeddings from the eigenvectors of $H\mathbf{X}(H\mathbf{X})^\top$.

Both methods rely on the singular value decomposition (SVD) of $H\mathbf{X}$.

Detailed Explanation of Duality

Singular Value Decomposition (SVD):

$$H\mathbf{X} = U\tilde{\Lambda}^{1/2}V^\top$$

- MDS uses U (left singular vectors) and $\tilde{\Lambda}$ (singular values).
- PCA uses V (right singular vectors) and $\tilde{\Lambda}$ (singular values).

This shows that MDS and PCA are dual methods, analyzing complementary covariance structures.

Key Result

Theorem: Classical MDS on distances is equivalent to PCA on the centered data matrix.

$$H\mathbf{X}(H\mathbf{X})^\top = U\tilde{\Lambda}U^\top$$

$$(H\mathbf{X})^\top H\mathbf{X} = V\tilde{\Lambda}V^\top$$

Conclusion: The MDS embedding and PCA scores are both derived from $H\mathbf{X}$ but use different components of the SVD.