Multidimensional Scaling

Paula Andrea Pérez Toro
MSc. student in Telecommunications Engineering

Guberney Muñeton Santa
Ph.D. student in Electronic and Computation Engineering

GITA research group, University of Antioquia.

paula.perezt@udea.edu.co-guberney.muneton@udea.edu.co

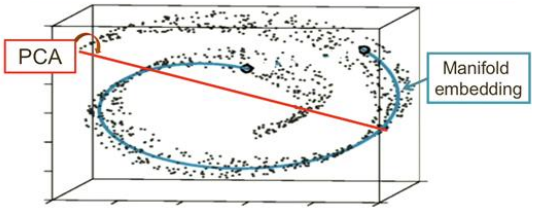
June 6, 2019

Outline

Manifold Learning

Manifold Learning

- Manifold as a surface of any shape, not necessarily have to be a plane.
- It reduces data dimensions and feature-sets from uneven weird surface by sub–optimal data representation.
- Attempt to generalize linear frameworks like PCA.



- Manifolds give a look of flat and featureless space that behaves like Euclidean space.
- The goal of the manifold learning algorithms is to recover the original domain structure, up to scaling rotation.
- Low-dimensional surface (manifold $\mathcal{M}$) embedded in high-dimensional space.
- $\mathcal{M} \subset \mathbb{R}^p$ is isomorphic to $\mathbb{R}^d$

Such that:

$$||\varphi(\boldsymbol{X_i}) - \varphi(\boldsymbol{X_{i'}})|| = \delta(\boldsymbol{X_i}, \boldsymbol{X_{i'}})$$

p: extrinsic dimension of dataset $\mathcal{X} = \{\boldsymbol{X_i}\}_1^n$
d: intrinsic dimension of $\mathcal{X}$
$d \leq p$

Aim: given $\boldsymbol{X_i} \in \mathcal{M} \subset \mathbb{R}^p$, find $\boldsymbol{Y_i} = \varphi(\boldsymbol{X_i}) \in \mathbb{R}^d$

Manifold learning is divided in two categories:

- Global methods: in the dimension reduction global properties are preserved. Methods as: **Multidimensional Scaling (MDS)** or Isomaps.
- Local methods: n the dimension reduction local properties are preserved. Methods as: Local linear embeddings or Laplacian eigenmap.

Multidimensional Scaling

Multidimensional Scaling (MDS)

- A multivariate technique to measure the similarity or dissimilarity in data.
- MDS leads to an Euclidean space.
- It transforms the data into a low–dimensional space, where distance between the original high–dimensional space gets reflected to the low–dimensional space.
- To reduce the dimensionality preserving the maximum amount of information.
- The distortion caused by the dimensional reduction will be as minimum as possible.

Multidimensional Scaling (MDS)

- Let be $\boldsymbol{X_1},\ \boldsymbol{X_2}, ...,\ \boldsymbol{X_N} \ \in \mathbb{R}^p$
- The Euclidean distance is chosen, but other distances can be used:

$$d_{i,j} = ||\boldsymbol{X_i} - \boldsymbol{X_j}||_{\mathbb{R}^p} = ||\boldsymbol{z_i} - \boldsymbol{z_{i'}}||_{\mathbb{R}^d}$$

- MDS only requires dissimilarities $d_{i,j}$.
- MDS is also known for identifying news representation by minimizing the quantity called Stress or Sstress

Multidimensional Scaling (MDS)

- MDS seeks values $\boldsymbol{z_1}, \boldsymbol{z_2}, ..., \boldsymbol{z_N} \in \mathbb{R}^d$ to minimize the so-called Stress function.

$$\mathrm{S_M}(\boldsymbol{z_1}, \boldsymbol{z_2}, ..., \boldsymbol{z_N}) = \sum_{i \neq i'} (d_{i,i'} - ||\boldsymbol{z_i} - \boldsymbol{z_{i'}}||)^2$$

- This method is known as least squared or Kruskal-Shepard scaling.
- A gradient descent algorithm or similar is used to minimize $\mathrm{S_M}$.
- To find a lower–dimensional representation that preserves the pairwise distances.

Multidimensional Scaling (MDS)

- A variation on least squares scaling is the so-called Sammon mapping which minimizes:

$$\mathrm{S_{sm}} = \sum_{i \neq i'} \frac{(d_{i,i'} - ||\boldsymbol{z_i} - \boldsymbol{z_{i'}}||)^2}{d_{i,i'}}$$

- Here more emphasis is put in preserving smaller pairwise distances.

- In classical scaling, we instead start with similarities $s_{i,i'}$.
- Using the Euclidean distance, its squared can be define as inner products:

$$d_{i,i'}^2 = ||\boldsymbol{X_i} - \boldsymbol{X_{i'}}||^2$$

- From the previous equation, thus:

$$s_{i,i'} := < \boldsymbol{X_i}, \boldsymbol{X_{i'}} >$$

- Often used the centered inner product:

$$s_{i,i'} := < \boldsymbol{X_i} - \overline{\boldsymbol{X}}, \boldsymbol{X_{i'}} - \overline{\boldsymbol{X}} >$$

- The problem then is to minimize:

$$S_c(\boldsymbol{z_1},\ \boldsymbol{z_2}, ...,\ \boldsymbol{z_N}) = \sum_{i \neq i'} (S_{i,i'} - < \boldsymbol{z_i} - \overline{\boldsymbol{z}}, \boldsymbol{z_{i'}} - \overline{\boldsymbol{z}} >)^2$$

- If we have distances rather than inner–products we can convert them to centered inner–products, if the distances are Euclidean.
- If the similarities are in fact centered inner products, classical scaling is exactly equivalent to principal components.
- The algorithm is based on the spectral decomposition of Stress function:

$$||\boldsymbol{z_i} - \boldsymbol{z_{i'}}||^2 = ||\boldsymbol{X_i} - \boldsymbol{X_{i'}}||^2$$

- There is an explicit solution in terms of eigenvalues. The embedding is given by the eigenvectors corresponding to the d largest eigenvalues of the spectral decomposition of Stress:

$$S = \mathrm{V}\Lambda\mathrm{V}^T = (\mathrm{V}\Lambda^{1/2})(\Lambda^{1/2}\mathrm{V}^T)$$

- Then, if S is non-negative definite, the largest eigenvalues will be positive.

$$S = (\mathrm{V}\Lambda^{1/2})(\Lambda^{1/2}\mathrm{V}^T)$$

$$= \boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{z}$$

- Thus,

$$\boldsymbol{z} = (\boldsymbol{z_1}, \boldsymbol{z_2}, ..., \boldsymbol{z_N}) = \Lambda^{1/2}\mathrm{V}^T = (\lambda_1^{1/2}\mathcal{V}_1, \lambda_2^{1/2}\mathcal{V}_2, ..., \lambda_d^{1/2}\mathcal{V}_d)^T$$

Bonus: non–metric MDS

Seeks to minimize the Stress function:

$$\mathrm{S_{NM}}(\boldsymbol{z_1},\ \boldsymbol{z_2}, ...,\ \boldsymbol{z_N}) = \sum_{i \neq i'} \frac{[||\boldsymbol{z_i} - \boldsymbol{z_{i'}}|| - \theta(d_{i,i'})]^2}{||\boldsymbol{z_i} - \boldsymbol{z_{i'}}||^2}$$

θ: an arbitrary increasing function.

Ordinal data. Increasing monotonic relation between the proximities.

Review: Classical MDS

1. We must know the dissimilarities $d_{i,i'}$ of $\boldsymbol{X_1}, \boldsymbol{X_2}, ..., \boldsymbol{X_N} \in \mathbb{R}^p$
2. Seeks values $\boldsymbol{z_1}, \boldsymbol{z_2}, ..., \boldsymbol{z_N} \in \mathbb{R}^d$ to minimize Stress function by gradient descent algorithm or similar.

$$\mathrm{S_M}(\boldsymbol{z_1},\ \boldsymbol{z_2}, ...,\ \boldsymbol{z_N}) = \sum_{i \neq i'} (d_{i,i'} - ||\boldsymbol{z_i} - \boldsymbol{z_{i'}}||)^2$$

We instead start with similarities $s_{i,i'}$ often using the centering inner–product.

$$S_c(\boldsymbol{z_1},\ \boldsymbol{z_2}, ...,\ \boldsymbol{z_N}) = \sum_{i \neq i'} (S_{i,i'} - < \boldsymbol{z_i} - \overline{\boldsymbol{z}}, \boldsymbol{z_{i'}} - \overline{\boldsymbol{z}} >)^2$$

3. Compute the eigenvalue decomposition $S = \mathrm{V}\Lambda\mathrm{V}^T$.
4. Get $\boldsymbol{z} = \boldsymbol{z_1}, \boldsymbol{z_2}, ..., \boldsymbol{z_N} = \Lambda^{1/2}\mathrm{V}^T$

References

- Alan J Izenman .Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning.
- Trevor Hastie, Robert Tibshirani, Jerome H. Friedman.The Elements of Statistical Learning.
- Santiago Gallón. High-Dimensional Data and Non-Linear Dimensionality Reduction (Slides).