FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG
SCHOOL OF ENGINEERING

Lecture Pattern Analysis

Part 13: Multidimensional Scaling

Christian Riess
IT Security Infrastructures Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
June 3, 2021

Introduction

- Multidimensional Scaling (MDS) is equivalent to PCA, but operates on distances between samples
- This variant may be useful if only relative information about samples is available, e.g., from a learned similarity measure
- MDS operates on a distance matrix

$$D^2 = [d_{ij}^2] \text{ where } 1 \leq i, j \leq N, \tag{1}$$

and

$$d_i j = (\mathbf{x}_i - \mathbf{x}_j)^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j) \ . \tag{2}$$

- However, the absolute locations $\mathbf{x}_i$, $\mathbf{x}_j \in \mathbb{R}^d$ are the unknown quantity
- Hence, we seek to reconstruct $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \mathbb{R}^{N \times d'}$, where $d' \leq d$ is an orthogonal projection onto a d'-dimensional subspace
- Since we can not distinguish spatial translations of $\mathbf{X}$, we require the unique solution $\mathbf{X}$ to be mean-free, i.e., its mean value is 0

Linking Distances to Coordinates

- The components of the distance matrix can be written in matrix notation

$$d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j) \tag{3}$$

$$= \mathbf{x}_i^\mathsf{T}\mathbf{x}_i + \mathbf{x}_j^\mathsf{T}\mathbf{x}_j - 2\mathbf{x}_i^\mathsf{T}\mathbf{x}_j \tag{4}$$

$$\Rightarrow D^2 = \mathrm{diag}(\mathbf{X}^\mathsf{T}\mathbf{X}) \cdot \mathbb{1}\mathbb{1}^\mathsf{T} + \mathbb{1}\mathbb{1}^\mathsf{T} \cdot \mathrm{diag}(\mathbf{X}^\mathsf{T}\mathbf{X})^\mathsf{T} - 2\mathbf{X}^\mathsf{T}\mathbf{X} \tag{5}$$

  where $\mathbb{1} = (1, \ldots, 1)^\mathsf{T} \in \mathbb{R}^N$ and $\mathrm{diag}(\mathbf{X})$ is a matrix with the diagonal elements of $\mathbf{X}$ and 0 everywhere else
- This equation links the distances to the actual coordinates
- $\mathbf{X}$ can be recovered after multiplying $\mathbf{D}^2$ by a centering matrix $\mathbf{C}$ with

$$\mathbf{C} = (\mathbf{I} - \frac{1}{N}\mathbb{1}\mathbb{1}^\mathsf{T}) \tag{6}$$

  where $\mathbf{I}$ is the $n \times n$ identity matrix

Multiplication by the Centering Matrix

- The multiplication itself is somewhat messy, but not difficult. The equation is

$$-\frac{1}{2}\mathbf{C}\mathbf{D}^2\mathbf{C} = -\frac{1}{2}\left(\mathbf{I} - \frac{1}{N}\mathbb{1}\mathbb{1}^\mathsf{T}\right)\cdot \tag{7}$$

$$\Bigg(\underbrace{\text{diag}\left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)\cdot\mathbb{1}\mathbb{1}^\mathsf{T}}_{(1)} + \underbrace{\mathbb{1}\mathbb{1}^\mathsf{T}\cdot\text{diag}\left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)^\mathsf{T}}_{(2)} - \underbrace{2\mathbf{X}^\mathsf{T}\mathbf{X}}_{(3)}\Bigg)\left(\mathbf{I} - \frac{1}{N}\mathbb{1}\mathbb{1}^\mathsf{T}\right) \tag{8}$$

- The sum in the middle has three elements, we will look at them one by one
- Element (1):

$$\left(\mathbf{I} - \frac{1}{N}\mathbb{1}\mathbb{1}^\mathsf{T}\right)\left(\text{diag}\left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)\cdot\mathbb{1}\mathbb{1}^\mathsf{T}\right)\left(\mathbf{I} - \frac{1}{N}\mathbb{1}\mathbb{1}^\mathsf{T}\right) \tag{9}$$

$$= \left(\mathbf{I} - \frac{1}{N}\mathbb{1}\mathbb{1}^\mathsf{T}\right)\text{diag}\left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)\cdot\left(\mathbb{1}\mathbb{1}^\mathsf{T}\cdot\mathbf{I} - \frac{1}{N}\mathbb{1}\mathbb{1}^\mathsf{T}\cdot\mathbb{1}\cdot\mathbb{1}^\mathsf{T}\right) \tag{10}$$

Solving Elements (1), (2) and (3)

- Element (1) evaluates to 0, since the rightmost brackets are 0:

$$\mathbb{1}^\mathsf{T}\mathbb{1} = N \ , \text{ hence } \frac{1}{N}\mathbb{1}^\mathsf{T}\mathbb{1} = 1 \ , \text{ and finally } \mathbb{1}\mathbb{1}^\mathsf{T} - \mathbb{1}\mathbb{1}^\mathsf{T} = \mathbf{0} \ . \tag{11}$$

- With the analogous calculation, element (2) also becomes 0
- Hence, if there will be anything interesting, it must be in element (3):

$$-\frac{1}{2}\left(\mathbf{I} - \frac{1}{N}\mathbb{1}\mathbb{1}^\mathsf{T}\right) \cdot \left(-2\mathbf{X}^\mathsf{T}\mathbf{X}\right)\left(\mathbf{I} - \frac{1}{N}\mathbb{1}\mathbb{1}^\mathsf{T}\right) \tag{12}$$

$$= \left(\mathbf{I}\cdot\mathbf{X}^\mathsf{T} - \underbrace{\frac{1}{N}\mathbb{1}\cdot\mathbb{1}^\mathsf{T}\cdot\mathbf{X}^\mathsf{T}}_{=0 \text{ (zero mean!)}}\right) \cdot \left(\mathbf{X}\cdot\mathbf{I} - \underbrace{\frac{1}{N}\mathbf{X}\cdot\mathbb{1}\cdot\mathbb{1}^\mathsf{T}}_{=0 \text{ (zero mean!)}}\right) \tag{13}$$

$$= \mathbf{X}^\mathsf{T}\mathbf{X} \tag{14}$$

Obtaining the Coordinates

- To summarize, we showed that

$$-\frac{1}{2}\mathbf{C}\mathbf{D}^2\mathbf{C} = \mathbf{X}^\mathsf{T}\mathbf{X} \tag{15}$$

- Hence, **X** can be obtained via singular value decomposition (SVD),

$$\mathrm{SVD}\left(-\frac{1}{2}\mathbf{C}\mathbf{D}^2\mathbf{C}\right) = \mathrm{SVD}\left(\mathbf{X}^\mathsf{T}\mathbf{X}\right) = \mathbf{U}^\mathsf{T}\Sigma\mathbf{V} \tag{16}$$

$$\Rightarrow \mathbf{X} = \Sigma^{\frac{1}{2}}\mathbf{V} \tag{17}$$

- As usual, Σ is a diagonal matrix, so the square root is just the square root of its elements
- Analogously to PCA, we could now also select the number of dimensions from the magnitudes of the singular values