3.5.3 Algorithm (MDS)

Input: square distance matrix $D \in \mathbb{R}^{n\times n}$

Output: a configuration of sample pts using "artificial features" $x^{(1)}, \cdots, x^{(n)} \in \mathbb{R}^d$.

Step 1: Compute $K \overset{\text{def}}{=} -\frac{1}{2} HDH$ where $H = I - \frac{1}{n} \mathbb{1}\mathbb{1}^T$ is the centering matrix.

Step 2: Compute the EVD:

$$K_{n\times n} = V_{n\times n} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}_{n\times n} V^T$$

with $V = \begin{pmatrix} | & & | \\ v^{(1)} & \cdots & v^{(n)} \\ | & & | \end{pmatrix}$ orthogonal

Step 3: Take $X = \begin{pmatrix} | & & | \\ x^{(1)} & \cdots & x^{(n)} \\ | & & | \end{pmatrix}_{d\times n} \text{"="} \begin{pmatrix} \sqrt{\lambda_1} & & & & & \\ & \ddots & & & & \\ & & \sqrt{\lambda_d} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}_{n\times n} V^T$

$$= \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_d} \end{pmatrix}_{d\times d} \begin{pmatrix} | & & | \\ v^{(1)} & \cdots & v^{(d)} \\ | & & | \end{pmatrix}^T_{d\times n}$$

Remark: Comparison between PCA and MDS.

Let $X = U \Sigma V^T$ be the SVD of the sample matrix where

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}$$

with singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$.

PCA: makes use of the EVD of the covariance matrix $XX^T = U \underbrace{\Sigma \Sigma^T}_{\text{square, diagonal}} U^T$.

The first d columns of U determine the d-dimensional subspace with the largest projected data variance.

MDS: makes use of the EVD of the kernel matrix ("Gram matrix")

$$X^T X = V \underbrace{\Sigma^T \Sigma}_{\text{square, diagonal}} V^T$$

The first d columns of V (and the first d singular values of X) determine the "artificial features".

Summary of this section:

- What is MDS? Def
- How to implement? Algorithm.

Reference: "Generalized PCA" by Vidal-Ma-Sastry, Section 4.2.