Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Emilio Balda

Exercise 6

Friday, November 24, 2017

Problem 1. *(MDS vs. PCA)* Suppose that $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ and $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n] \in \mathbb{R}^{p\times n}$. Let $\mathbf{E}_n$ be the centering matrix defined as $\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$.

a) Show that the sample covariance matrix $\mathbf{S}_n$ is equal to $\frac{1}{n-1}\mathbf{X}\mathbf{E}_n\mathbf{X}^T$.

b) Show that if the projection matrix in PCA is $\mathbf{Q}$ then the projected points are given by $\mathbf{Q}\mathbf{X}\mathbf{E}_n$.

c) Consider n points presented as $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n] \in \mathbb{R}^{p\times n}$. Show that the if PCA analysis is applied to have the best projection on a k-dimensional space, the output is given by:

$$\begin{bmatrix} \sqrt{\lambda_1}\mathbf{v}_1^T \\ \vdots \\ \sqrt{\lambda_k}\mathbf{v}_k^T \end{bmatrix}$$

where $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_p]$ comes from the singular value decomposition of $\mathbf{X}\mathbf{E}_n$ which is :

$$\mathbf{X}\mathbf{E}_n = \mathbf{U}_{p\times p}\mathbf{\Lambda}\mathbf{V}_{n\times p}^T.$$

d) Show that applying MDS on the distance matrix $\mathbf{D}(\mathbf{X})$ provides the same result as PCA.

Problem 2. *(MDS in 3-dimensional space)* Consider four samples in $\mathbb{R}^3$ given as follows:

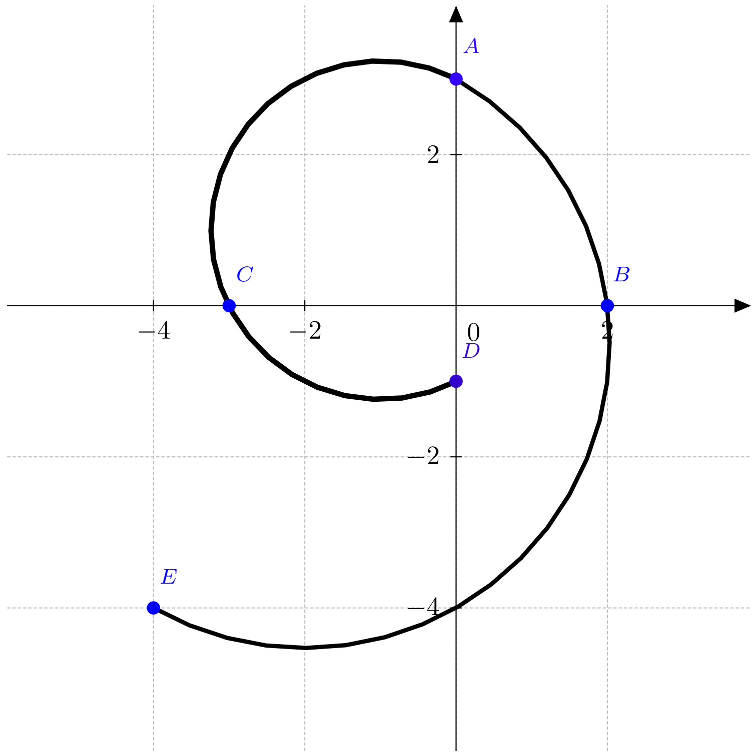
$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix} \mathbf{x}_3 = \begin{bmatrix} -4 \\ 2 \\ 2 \end{bmatrix} \mathbf{x}_4 = \begin{bmatrix} -3 \\ -1 \\ 4 \end{bmatrix}.$$

Using MDS, find the best Euclidean embedding for presenting the data in two dimensional space. Explain each step.

Problem 3. *(Isomap)* In this exercise, we examine what happens if the dataset is not large enough to find its geometry. Consider five vectors $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ and $\mathbf{E}$ given as follows

$$\mathbf{A} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \mathbf{C} = \begin{pmatrix} -3 \\ 0 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \mathbf{E} = \begin{pmatrix} -4 \\ 4 \end{pmatrix}.$$

a) Construct the graph for the Isomap based on 1-nearest neighbor and 2-nearest neighbor criteria and the mutual distances as the weights. Estimate the geodesic distance of $\mathbf{E}$ and $\mathbf{D}$.



b) Construct the graph for the Isomap using an ϵ. Discuss the choice of ϵ.