Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Markus Rothe

Exercise 7

Friday, November 30, 2018

Problem 1. *(Characterization of Euclidean Distance Matrices)*

a) Show that if $\mathbf{D}(\mathbf{X}) \in \mathbb{R}^{n\times n}$ is a distance matrix, then show that:

$$-\frac{1}{2}\mathbf{D}^{(2)}(\mathbf{X}) = \mathbf{X}\mathbf{X}^T - \mathbf{1}_n\hat{\mathbf{x}}^T - \hat{\mathbf{x}}\mathbf{1}^T$$

where $\hat{\mathbf{x}} = \frac{1}{2}[\mathbf{x}_1^T\mathbf{x}_1, \ldots, \mathbf{x}_n^T\mathbf{x}_n]^T$.

b) Consider $-\frac{1}{2}\mathbf{E}_n\mathbf{\Delta}^{(2)}\mathbf{E}_n$, which is non-negative definite and $\text{rk}(-\frac{1}{2}\mathbf{E}_n\mathbf{\Delta}^{(2)}\mathbf{E}_n) \leq k$, then there exists $n \times k$ matrix $\mathbf{X}$ such that

$$-\frac{1}{2}\mathbf{E}_n\mathbf{\Delta}^{(2)}\mathbf{E}_n = \mathbf{X}\mathbf{X}^T, \text{ and } \mathbf{X}^T\mathbf{E}_n = \mathbf{X}^T.$$

c) A matrix with zero diagonal elements is called hollow matrix. Prove that if $\mathbf{A}$ is a symmetric hollow matrix, then $\mathbf{A} = 0$ if and only if $\mathbf{E}_n\mathbf{A}\mathbf{E}_n = 0$.

Problem 2. *(MDS vs. PCA)* Suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ and $\mathbf{X} = [\mathbf{x}_1 \ldots \mathbf{x}_n] \in \mathbb{R}^{p\times n}$. 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 $\mathbf{X} = [\mathbf{x}_1 \ldots \mathbf{x}_n] \in \mathbb{R}^{p\times n}$. 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 $\mathbf{V} = [\mathbf{v}_1 \ldots \mathbf{v}_p]$ 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 3. *(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.