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

Exercise 7
- Proposed Solution -

Friday, November 30, 2018

Solution of Problem 1

a) Note that:

$$\|\mathbf{x}_i - \mathbf{x}_j\|^2 = \mathbf{x}_i^T\mathbf{x}_i + \mathbf{x}_j^T\mathbf{x}_j - 2\mathbf{x}_i^T\mathbf{x}_j^T.$$

It is easy to check that:

$$(\mathbf{X}\mathbf{X}^T)_{ij} = \mathbf{x}_i\mathbf{x}_j^T.$$

Consider $\hat{\mathbf{x}} = \frac{1}{2}[\mathbf{x}_1^T\mathbf{x}_1, \dots, \mathbf{x}_n^T\mathbf{x}_n]^T$. We have:

$$\mathbf{1}_n\hat{\mathbf{x}}^T = \begin{bmatrix} \frac{1}{2}\mathbf{x}_1^T\mathbf{x}_1 & \dots & \frac{1}{2}\mathbf{x}_n^T\mathbf{x}_n \\ \frac{1}{2}\mathbf{x}_1^T\mathbf{x}_1 & \dots & \frac{1}{2}\mathbf{x}_n^T\mathbf{x}_n \\ \vdots & \ddots & \vdots \\ \frac{1}{2}\mathbf{x}_1^T\mathbf{x}_1 & \dots & \frac{1}{2}\mathbf{x}_n^T\mathbf{x}_n \end{bmatrix}$$

This means that $(\mathbf{1}_n\hat{\mathbf{x}}^T)_{ij} = \frac{1}{2}\mathbf{x}_j^T\mathbf{x}_j$ and moreover $(\hat{\mathbf{x}}\mathbf{1}_n^T)_{ij} = \frac{1}{2}\mathbf{x}_i^T\mathbf{x}_i$

Therefore:

$$\left(-\frac{1}{2}\mathbf{D}^{(2)}(\mathbf{X})\right)_{ij} = (\mathbf{X}\mathbf{X})_{ij} - (\mathbf{1}_n\hat{\mathbf{x}}^T)_{ij} - (\hat{\mathbf{x}}\mathbf{1}_n^T)_{ij}.$$

The element-wise identity implies the desired identity.

b) Since $-\frac{1}{2}\mathbf{E}_n\boldsymbol{\Delta}^{(2)}\mathbf{E}_n$ is non-negative definite and has the rank $\mathrm{rk}(-\frac{1}{2}\mathbf{E}_n\boldsymbol{\Delta}^{(2)}\mathbf{E}_n) \le k$, it can be written as:

$$-\frac{1}{2}\mathbf{E}_n\boldsymbol{\Delta}^{(2)}\mathbf{E}_n = \sum_{i=1}^{k}\lambda_i\mathbf{v}_i\mathbf{v}_i^T,$$

where $\lambda_1 \ge \dots \ge \lambda_k$ are top k eigenvalues of the matrix $-\frac{1}{2}\mathbf{E}_n\boldsymbol{\Delta}^{(2)}\mathbf{E}_n$ with corresponding orthonormal eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. This can be obtained from spectral decomposition of $-\frac{1}{2}\mathbf{E}_n\boldsymbol{\Delta}^{(2)}\mathbf{E}_n$. Using this representation, the matrix $\mathbf{X}$ can be constructed as $\mathbf{X} = [\sqrt{\lambda_1}\mathbf{v}_1, \dots, \sqrt{\lambda_k}\mathbf{v}_k]$. It can be seen that:

$$\mathbf{X}\mathbf{X}^T = \sum_{i=1}^{k}\lambda_i\mathbf{v}_i\mathbf{v}_i^T = -\frac{1}{2}\mathbf{E}_n\boldsymbol{\Delta}^{(2)}\mathbf{E}_n.$$

Moreover the image of $-\frac{1}{2}\mathbf{E}_n\boldsymbol{\Delta}^{(2)}\mathbf{E}_n$ is a subset of the image of $\mathbf{E}_n$. Therefore for all non-zero λ_i, the corresponding eigenvector $\mathbf{v}_i$ belongs to the image of $\mathbf{E}_n$ and since it is an orthogonal projection:

$$\mathbf{E}_n\mathbf{v}_i = \mathbf{v}_i.$$

If $\lambda_i = 0$, then trivially $\mathbf{E}_n\sqrt{\lambda_i}\mathbf{v}_i = \sqrt{\lambda_i}\mathbf{v}_i = 0$. This means that:

$$\mathbf{E}_n\mathbf{X} = \mathbf{X} \implies \mathbf{X}^T\mathbf{E}_n = \mathbf{X}^T.$$

c) The direction where $\mathbf{A} = 0$ is trivial. Let us assume $\mathbf{E}_n\mathbf{A}\mathbf{E}_n = 0$. This means that the matrix $\mathbf{A}$ takes each vector in the image of $\mathbf{E}_n$ to the kernel of $\mathbf{E}_n$. Note that the kernel of $\mathbf{E}_n$ is spanned by $\mathbf{1}_n$, so for each $\mathbf{v}$ such that $\mathbf{v}^T\mathbf{1}_n = 0$, we have:

$$\exists \alpha \in \mathbb{R}; \mathbf{A}\mathbf{v} = \alpha \mathbf{1}_n.$$

Pich $\mathbf{v} = \mathbf{e}_i - \mathbf{e}_j$. The equation above implies that $(\mathbf{A}\mathbf{v})_i = (\mathbf{A}\mathbf{v})_j$. But $(\mathbf{A}\mathbf{v})_k = a_{ki} - a_{kj}$. Therefore:

$$a_{ii} - a_{ij} = a_{ji} - a_{jj}.$$

But $a_{kk} = 0$ for all $1 \leq k \leq n$ and $\mathbf{A}$ is symmetric. Therefore $a_{ij} = 0$ for all i, j which means that $\mathbf{A} = 0$.

Solution of Problem 2

a) First of all, note that:

$$\overline{\mathbf{x}} = \frac{1}{n}\mathbf{X}\mathbf{1}_n.$$

Moreover:

$$\mathbf{S}_n = \frac{1}{n-1}(\mathbf{X} - \overline{\mathbf{x}}\mathbf{1}_n^T)(\mathbf{X} - \overline{\mathbf{x}}\mathbf{1}_n^T)^T.$$

Therefore:

$$\mathbf{S}_n = \frac{1}{n-1}(\mathbf{X} - \frac{1}{n}\mathbf{X}\mathbf{1}_n\mathbf{1}_n^T)(\mathbf{X} - \frac{1}{n}\mathbf{X}\mathbf{1}_n\mathbf{1}_n^T)^T = \frac{1}{n-1}\mathbf{X}\mathbf{E}_n\mathbf{E}_n^T\mathbf{X}^T.$$

Using $\mathbf{E}_n\mathbf{E}_n = \mathbf{E}_n$, we have $\mathbf{S}_n$ is equal to $\frac{1}{n-1}\mathbf{X}\mathbf{E}_n\mathbf{X}^T$.

b) The result of PCA is $\mathbf{Q}(\mathbf{x}_i - \overline{\mathbf{x}})$. This is indeed equal to $\mathbf{Q}(\mathbf{x}_i - \frac{1}{n}\mathbf{X}\mathbf{1}_n)$. Constructing the matrix $\mathbf{X}$ as suggested, the projected points can be written as:

$$\mathbf{Q}(\mathbf{X} - \frac{1}{n}\mathbf{X}\mathbf{1}_n\mathbf{1}_n^T) = \mathbf{Q}\mathbf{X}\mathbf{E}_n.$$

c) Let the singular value decomposition of $\mathbf{X}\mathbf{E}_n$ be:

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

It is known that:

$$\mathbf{S}_n = \frac{1}{n-1}\mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^T,$$

and top k eigenvectors of $\mathbf{S}_n$ are given therefore by picking first k columns of $\mathbf{U}$, denoted by $\mathbf{U}_k$. In any case, we have:

$$\mathbf{U}_k^T\mathbf{X} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{x}_1 & \dots & \mathbf{u}_1^T\mathbf{x}_n \\ \vdots & \ddots & \vdots \\ \mathbf{u}_k^T\mathbf{x}_1 & \dots & \mathbf{u}_k^T\mathbf{x}_n \end{bmatrix} = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_n],$$

where $\hat{\mathbf{x}}_i$ is the projected point into the k dimensional subspace. From the previous point, the projected points are given by $\mathbf{U}_k^T\mathbf{X}\mathbf{E}_n$.

See that:

$$\mathbf{U}_k^T \mathbf{X}\mathbf{E}_n = \mathbf{U}_k^T \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T.$$

But :

$$\mathbf{U}_k^T \mathbf{U} = \begin{bmatrix} \mathbf{u}_1^T \mathbf{u}_1 & \dots & \mathbf{u}_1^T \mathbf{u}_p \\ \vdots & \ddots & \vdots \\ \mathbf{u}_k^T \mathbf{u}_1 & \dots & \mathbf{u}_k^T \mathbf{u}_p \end{bmatrix} = [\mathbf{I}_k \quad \mathbf{0}_{k\times p-k}].$$

Using the fact that $\boldsymbol{\Lambda}_{ii}^2 = \lambda_i$, we have:

$$\mathbf{U}_k^T \mathbf{U}\boldsymbol{\Lambda} = [\mathbf{I}_k \quad \mathbf{0}_{k\times p-k}]\boldsymbol{\Lambda} = [\text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_k})_{k\times k} \quad \mathbf{0}_{k\times p-k}]$$

Now write $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_p]$ where $\mathbf{v}_i \in \mathbb{R}^n$. We have:

$$\mathbf{U}_k^T \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T = [\text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_k})_{k\times k} \quad \mathbf{0}_{k\times p-k}]\mathbf{V}^T = \begin{bmatrix} \sqrt{\lambda_1}\mathbf{v}_1^T \\ \vdots \\ \sqrt{\lambda_k}\mathbf{v}_k^T \end{bmatrix}$$

d) MDS starts by finding $-\frac{1}{2}\mathbf{E}_n\mathbf{D}^{(2)}\mathbf{E}_n$ which is $\mathbf{E}_n\mathbf{X}^T\mathbf{X}\mathbf{E}_n$ for Euclidean distance matrix. The spectral decomposition of $\mathbf{E}_n\mathbf{X}^T\mathbf{X}\mathbf{E}_n$ is then found by $\hat{\mathbf{V}}\,\text{diag}(\lambda_1, \dots, \lambda_n)\hat{\mathbf{V}}^T$ where $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1 \dots \hat{\mathbf{v}}_n]$ is the eigenvector matrix. Using SVD of $\mathbf{X}\mathbf{E}_n$ above we get:

$$\mathbf{E}_n\mathbf{X}^T\mathbf{X}\mathbf{E}_n = \mathbf{V}\boldsymbol{\Lambda}^2\mathbf{V}^T.$$

Therefore if $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_p]$, then for $i = 1, \dots, p$ we have:

$$\hat{\mathbf{v}}_i = \mathbf{v}_i.$$

The solution to MDS is then $\mathbf{X}^{*T} = [\sqrt{\lambda_1}\mathbf{v}_1, \dots, \sqrt{\lambda_k}\mathbf{v}_k] \in \mathbb{R}^{n\times k}$. This means that:

$$\mathbf{U}_k^T \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T = \mathbf{X}^*.$$

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

Remark: There is another way of showing this equivalence. Note that $\mathbf{S}_n = \frac{1}{n-1}\mathbf{X}\mathbf{E}_n\mathbf{X}^T$ and let $\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ be its spectral decomposition. Suppose that $(\lambda, \mathbf{u})$ is eigenvalue-eigenvector of $\mathbf{X}\mathbf{E}_n\mathbf{X}^T = (\mathbf{X}\mathbf{E}_n)(\mathbf{X}\mathbf{E}_n)^T$. Then:

$$(\mathbf{X}\mathbf{E}_n)^T(\mathbf{X}\mathbf{E}_n)(\mathbf{X}\mathbf{E}_n)^T\mathbf{u} = \lambda(\mathbf{X}\mathbf{E}_n)^T\mathbf{u}.$$

This means that $(\mathbf{X}\mathbf{E}_n)^T\mathbf{u} = \mathbf{E}_n\mathbf{X}^T\mathbf{u}$ is an eigenvector of $(\mathbf{X}\mathbf{E}_n)^T(\mathbf{X}\mathbf{E}_n) = \mathbf{E}_n\mathbf{X}^T\mathbf{X}\mathbf{E}_n$ and λ is its eigenvalue. So top k eigenvalues of $\mathbf{X}\mathbf{E}_n\mathbf{X}^T$ remains the same for $\mathbf{E}_n\mathbf{X}^T\mathbf{X}\mathbf{E}_n$. Therefore $\mathbf{E}_n\mathbf{X}^T\mathbf{u}_1, \dots \mathbf{E}_n\mathbf{X}^T\mathbf{u}_k$ are top k eigenvectors of $\mathbf{E}_n\mathbf{X}^T\mathbf{X}\mathbf{E}_n$. They are orthogonal but they do not have unit norm:

$$\mathbf{u}^T(\mathbf{X}\mathbf{E}_n)(\mathbf{X}\mathbf{E}_n)^T\mathbf{u} = \mathbf{u}^T\lambda\mathbf{u} = \lambda.$$

Therefore a normalization by $\frac{1}{\sqrt{\lambda}}$ is needed. So top k eigenvectors of $\mathbf{E}_n\mathbf{X}^T\mathbf{X}\mathbf{E}_n$ is given by $\frac{1}{\sqrt{\lambda_1}}\mathbf{E}_n\mathbf{X}^T\mathbf{u}_1, \dots, \frac{1}{\sqrt{\lambda_k}}\mathbf{E}_n\mathbf{X}^T\mathbf{u}_k$. Therefore $\mathbf{X}^*_{\text{MDS}}$ is given by:

$$\mathbf{X}^*_{\text{MDS}} = \begin{bmatrix} \sqrt{\lambda_1}\mathbf{v}_1^T \\ \vdots \\ \sqrt{\lambda_k}\mathbf{v}_k^T \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1}(\frac{1}{\sqrt{\lambda_1}}\mathbf{E}_n\mathbf{X}^T\mathbf{u}_1)^T \\ \vdots \\ \sqrt{\lambda_k}(\frac{1}{\sqrt{\lambda_k}}\mathbf{E}_n\mathbf{X}^T\mathbf{u}_k)^T \end{bmatrix} = \begin{bmatrix} \mathbf{X}\mathbf{E}_n\mathbf{u}_1^T \\ \vdots \\ \mathbf{X}\mathbf{E}_n\mathbf{u}_k^T \end{bmatrix} = \mathbf{U}_k^T\mathbf{X}\mathbf{E}_n.$$

But we have seen above that $\mathbf{U}_k^T\mathbf{X}\mathbf{E}_n$ is X^*_{PCA} and therefore the desired result follows.

Solution of Problem 3

Consider four samples in $\mathbb{R}^3$ given as follows:

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

MDS steps are as follows:

a) Find $\mathbf{E}_n\mathbf{X}^T\mathbf{X}\mathbf{E}_n$ where $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n]$.
In this step, $\mathbf{X}$ is obtained as:

$$\mathbf{X} = \begin{bmatrix} 1 & 3 & -4 & -3 \\ 2 & -1 & 2 & -1 \\ -3 & -2 & 2 & 4 \end{bmatrix}$$

We have:

$$\mathbf{E}_n\mathbf{X}^T\mathbf{X}\mathbf{E}_n = \begin{bmatrix} 15.875 & 11.625 & -9.125 & -18.375 \\ 11.625 & 21.375 & -18.375 & -14.625 \\ -9.125 & -18.375 & 15.875 & 11.625 \\ -18.375 & -14.625 & 11.625 & 21.375 \end{bmatrix}$$

b) Find spectral decomposition of $\mathbf{E}_n\mathbf{X}^T\mathbf{X}\mathbf{E}_n = \mathbf{V}\,\mathrm{diag}(\lambda_1, \dots, \lambda_n)\mathbf{V}^T$.
For this example eigenvalues and eigenvectors are given by:

$$\mathrm{diag}(\lambda_1, \dots, \lambda_n) = \begin{bmatrix} 61 \\ 13.5 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} -0.45267873 & 0.5 & -0.65666815 & 0.25502096 \\ -0.54321448 & -0.5 & -0.31320188 & -0.63102251 \\ 0.45267873 & 0.5 & -0.28197767 & -0.71157191 \\ 0.54321448 & -0.5 & -0.62544394 & 0.17447155 \end{bmatrix}$$

c) $\mathbf{X}^*$ is given by $[\sqrt{\lambda_1}\mathbf{v}_1, \dots, \sqrt{\lambda_k}\mathbf{v}_k]^T$.

$$\mathbf{X}^{*T} = \begin{bmatrix} -3.53553391 & 1.83711731 \\ -4.24264069 & -1.83711731 \\ 3.53553391 & 1.83711731 \\ 4.24264069 & -1.83711731 \end{bmatrix}$$

Checking with PCA process, similar output is found.