Prof. Dr. Rudolf Mathar, Dr. Gholamreza Alirezaei, Dr. Arash Behboodi

Exercise 3
- Proposed Solution -

Friday, November 18, 2016

Solution of Problem 1

a) For each component x_i of $\mathbf{x}$, we find $\frac{\partial \mathbf{y}^T\mathbf{x}}{\partial x_i}$:

$$\frac{\partial \mathbf{y}^T\mathbf{x}}{\partial x_i} = \frac{\partial(\sum_{j=1}^{n} y_j x_j)}{\partial x_i} = y_i.$$

This implies that $\frac{\partial \mathbf{y}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{y}$.

b) Similar to previous step:

$$\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial x_i} = \frac{\partial\left(\sum_{i,j=1}^{n} A_{ij}x_i x_j\right)}{\partial x_i} = \sum_{j=1}^{n} A_{ij}x_j + \sum_{j=1}^{n} A_{ji}x_j = (\mathbf{A}\mathbf{x})_i + (\mathbf{A}^T\mathbf{x})_i.$$

And this yields the result.

c) Note that:

$$\operatorname{tr}(\mathbf{X}\mathbf{A}) = \sum_{i=1}^{n}(\mathbf{X}\mathbf{A})_{ii} = \sum_{i=1}^{n}(\sum_{j=1}^{n} x_{ij}a_{ji}).$$

Therefore:

$$\frac{\partial}{\partial x_{ij}}\operatorname{tr}(\mathbf{X}\mathbf{A}) = a_{ji}.$$

It easily implies the result.

d) Note that if $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n]$, then:

$$\operatorname{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) = \sum_{i=1}^{n} \mathbf{x}_i^T\mathbf{A}\mathbf{x}_i$$

Using (b), we have:

$$\frac{\partial}{\partial \mathbf{x}_j}\operatorname{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) = \frac{\partial}{\partial \mathbf{x}_j}(\mathbf{x}_j^T\mathbf{A}\mathbf{x}_j) = (\mathbf{A}^T + \mathbf{A})\mathbf{x}_j,$$

Therefore $\frac{\partial}{\partial \mathbf{X}}\operatorname{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) = (\mathbf{A}^T + \mathbf{A})\mathbf{X}$.

e) An interesting property of Frobenius norm is that:

$$\|\mathbf{X}\|_F^2 = \text{tr}(\mathbf{X}^T\mathbf{X}).$$

Then applying the previous result with $\mathbf{A} = \mathbf{I}$ yields the result.

f) Note that the Laplace expansion of a matrix is given by:

$$\det(\mathbf{X}) = \sum_{j=1}^{n}(-1)^{i+j}x_{ij}\det(\mathbf{X}_{ij})$$

where $\mathbf{X}_{ij}$ is a matrix obtained by deleting the row i and column j of $\mathbf{X}$ and its called its cofactor. The derivation is calculated as:

$$\frac{\partial}{\partial x_{ij}}\det(\mathbf{X}) = (-1)^{i+j}\det(\mathbf{X}_{ij}).$$

However the matrix having $(-1)^{i+j}\det(\mathbf{X}_{ij})$ as its (j,i) element is called adjoint of $\mathbf{X}$, denoted by $\text{adj}(\mathbf{X})$. It is easy to see that:

$$\mathbf{X}.\text{adj}(\mathbf{X}) = \det(\mathbf{X})\mathbf{I}$$

Therefore:

$$\frac{\partial}{\partial x_{ij}}\det(\mathbf{X}) = (\text{adj}(\mathbf{X}))_{ji} \implies \frac{\partial}{\partial \mathbf{X}}\det(\mathbf{X}) = \text{adj}(\mathbf{X})^T = \det(\mathbf{X})(\mathbf{X}^{-1})^T.$$

g) This is easily obtained by the chain rule and taking the derivative of logarithm.

Solution of Problem 2

Note that an estimator $\hat{X}$ of a parameter X is unbiased if its expected value equals X. Therefore it is enough to show:

$$\mathbb{E}(\overline{\mathbf{X}}) = \boldsymbol{\mu} = \mathbb{E}(\mathbf{X}), \quad \mathbb{E}(\mathbf{S}_n) = \boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}).$$

First see that:

$$\mathbb{E}(\overline{\mathbf{X}}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\mathbf{X}_i\right) = \frac{1}{n}n\mathbb{E}\left(\mathbf{X}\right) = \mathbb{E}\left(\mathbf{X}\right).$$

For the sample covariance matrix, we have:

$$\begin{aligned}\mathbb{E}(\mathbf{S}_n) &= \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T\right)\\ &= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left((\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T\right)\end{aligned}$$

Next see that:

$$\mathbb{E}\left((\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T\right) = \mathbb{E}\left((\mathbf{X}_i - \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j)(\mathbf{X}_i - \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j)^T\right)$$
$$= \mathbb{E}\left((\mathbf{X}_i - \boldsymbol{\mu} - \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu}))(\mathbf{X}_i - \boldsymbol{\mu} - \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu}))^T\right)$$
$$= \mathbb{E}\left(\left(\frac{n-1}{n}(\mathbf{X}_i - \boldsymbol{\mu}) - \frac{1}{n}\sum_{j=1, j\neq i}^{n}(\mathbf{X}_j - \boldsymbol{\mu})\right)\left(\frac{n-1}{n}(\mathbf{X}_i - \boldsymbol{\mu}) - \frac{1}{n}\sum_{j=1, j\neq i}^{n}(\mathbf{X}_j - \boldsymbol{\mu})\right)^T\right)$$

It is easy to see that:

$$\mathbb{E}\left((\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_j - \boldsymbol{\mu})^T\right) = \delta_{ij}\mathbf{\Sigma}.$$

Using this fact, it is easy to see that:

$$\mathbb{E}\left(\left(\frac{n-1}{n}(\mathbf{X}_i - \boldsymbol{\mu}) - \frac{1}{n}\sum_{j=1, j\neq i}^{n}(\mathbf{X}_j - \boldsymbol{\mu})\right)\left(\frac{n-1}{n}(\mathbf{X}_i - \boldsymbol{\mu}) - \frac{1}{n}\sum_{j=1, j\neq i}^{n}(\mathbf{X}_j - \boldsymbol{\mu})\right)^T\right)$$
$$= \frac{(n-1)^2}{n^2}\mathbb{E}\left((\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^T\right) + \frac{1}{n^2}\sum_{j=1, j\neq i}^{n}\mathbb{E}\left((\mathbf{X}_j - \boldsymbol{\mu})(\mathbf{X}_j - \boldsymbol{\mu})^T\right)$$
$$= \frac{(n-1)^2}{n^2}\mathbf{\Sigma} + \frac{n-1}{n^2}\mathbf{\Sigma} = \frac{n-1}{n}\mathbf{\Sigma}.$$

Therefore $\mathbb{E}\left((\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T\right) = \frac{n-1}{n}\mathbf{\Sigma}$. We can finally find the expected value of sample covariance as follows:

$$\mathbb{E}(\mathbf{S}_n) = \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left((\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T\right) = \frac{1}{n-1}\sum_{i=1}^{n}\frac{n-1}{n}\mathbf{\Sigma} = \mathbf{\Sigma}.$$

Solution of Problem 3

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

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

a) The sample mean can be easily found as:

$$\overline{\mathbf{x}} = \begin{bmatrix} -0.75 \\ 0.5 \\ 0.25 \end{bmatrix}$$

To find the sample covariance, we have:

$$\mathbf{S}_n = \frac{1}{3}\sum_{i=1}^{4}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T = \frac{1}{3}\begin{bmatrix} 32.75 & -4.5 & -28.25 \\ -4.5 & 9 & -4.5 \\ -28.25 & -4.5 & 32.75 \end{bmatrix}.$$

b) Step 1: find the sample covariance matrix $\mathbf{S}_n$ (previous part)

Step 2: find the eigenvalues and eigenvectors of the matrix. Sort them out and pick 2 orthonormal eigenvectors corresponding to 2 highest eigenvalues

$$\lambda_1 = 20.333333, \lambda_2 = 4.5, \lambda_3 = 0.$$

$$\mathbf{v}_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}.$$

Step 3: Construct $\mathbf{Q} = \mathbf{v}_1\mathbf{v}_1^T + \mathbf{v}_2\mathbf{v}_2^T$.
Following this procedure, we have:

$$\mathbf{Q} = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

c) Note that all the points are already on the same plane $x + y + z = 0$, so intuitively, the projection should be the projection on the same plane. This projection leaves those points untouched (Check!). Each $\mathbf{y} \in \text{Im}(\mathbf{Q})$ is also on this plane. To see that assume that $\mathbf{y} = \mathbf{Q}\mathbf{x}$. Then $y_1 + y_2 + y_3 = 0$. Another way, is to observe that the kernel of $\mathbf{Q}$ is spanned by the vector $(1, 1, 1)$, the last eigenvalue. Therefore its image is the orthogonal complement of this vector which is the plane $x + y + z = 0$.