High Dimensional Statistical Analysis

$\boxed{\text{Assignment 1}}$

Vector and Matrix Algebra, Multivariate Normal Distribution

Problem 1 Consider the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$$

. Answer the following questions

1. Is $\mathbf{A}$ symmetric?
2. Perform the spectral decomposition of A.
3. One way of writing the spectral decomposition of $\mathbf{A}$ is

$$\lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T.$$

   Please, identify each matrix in the representation above.

4. Use the spectral decomposition of $\mathbf{A}$ given above and find $\sqrt{\mathbf{A}}$. Check that the matrix you found satisfies

$$\sqrt{\mathbf{A}}\sqrt{\mathbf{A}} = \mathbf{A}.$$

Problem 2 Consider the spectral decomposition of a positive definite matrix as given in Lecture 1:

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T.$$

The columns of $\mathbf{P}$ are made of eigenvectors $\mathbf{e}_i$, $i = 1, \dots, n$ and they are orthonormalized, i.e. their lengths are one and they are orthogonal (peripendicular) one to another. The diagonal matrix $\mathbf{\Lambda}$ has the corresponding (positive) eigenvalues on the diagonal. Provide argument for the following

1. $\mathbf{P}^T = \mathbf{P}^{-1}$
2. Determinant of $\mathbf{\Lambda}$ is equal to the product of the terms on the diagonal.
3. Dederminant of $\mathbf{A}$ is the same as that of $\mathbf{\Lambda}$.
4. Find the inverse matrix to $\mathbf{\Lambda}$, i.e. $\mathbf{\Lambda}^{-1}$.

5. A simple way to determine the inverse of a matrix $\mathbf{A}$ from its spectral decomposition is through
$$\mathbf{A}^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^T.$$
Verify that the right hand side of the above indeed define the inverse of $\mathbf{A}$.
6. Check all these statements on the little example of Problem 1.

Problem 3 In a medical study, length L and weight W of newborn children is considered. It was assumed that (L, W) will be modeled through a bivariate normal distribution. The following information has been known: the mean weight is 3343[g], with the standard deviation of 528[g], while the mean length is 49.8[cm], with the standard deviation of 2.5[cm]. Additionally the correlation between the length and the weight has been established and equal to 0.75. The joint distribution of (W, L) is bivariate normal, i.e. $(W, L) \sim N(\boldsymbol{\mu}, \Sigma)$. Perform the following tasks and answer the questions:

1. Write explicitly the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
2. Write explicitly the density of the joint distribution.
3. Find eigenvalues and eigenvectors of the covariance matrix $\boldsymbol{\Sigma}$. Sketch few elipses corresponding to the constant density contours of the joint distributions. Mark on the plot the eigenvectors scaled by the square roots of the corresponding eigenvalues and comment.
4. How many parameters total characterize a bivariate normal distribution? How many parameters total characterizes a p-dimensional normal distribution?
5. What is the distribution of L? Give its name and parameters.
6. Suppose that the hospital records of a new-born child was lost. Give a best guess for the value of his/her length. Provide with accuracy bounds of your 'educated' guess based on the 3-σ rule.

Problem 4 In the setup of the previous problem, assume that it was reported by the mother of the child that weight was 4025[g].

1. What is the distribution of L given this additional information? Give its name and parameters.
2. Improve your previous guess and provide with accuracy limits.
3. Compare the answers from this and previous problems and comment how additional information affected the prediction value and accuracy.

Problem 5 Let $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ be independent $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random vectors of a dimension p.

1. Find the distribution of each of the following vectors:

$$\mathbf{V}_1 = \frac{1}{4}\mathbf{X}_1 - \frac{1}{2}\mathbf{X}_2 + \frac{1}{4}\mathbf{X}_3$$
$$\mathbf{V}_2 = \frac{1}{4}\mathbf{X}_1 - \frac{1}{2}\mathbf{X}_2 - \frac{1}{4}\mathbf{X}_3$$

2. Find the joint distribution of the above vectors.