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

Exercise 4

Friday, November 9, 2018

Problem 1. *(Bivariate Distribution)*

Suppose that $(Y_1, Y_2) \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Then obtain an expression in terms of $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho \in \mathbb{R}$ for the following distributions.

a) The joint distribution $f_{Y_1,Y_2}(y_1, y_2)$.

b) The distribution of Y_1 and the distribution of Y_2.

c) The conditional density $f_{Y_1|Y_2}(y_1|y_2)$.

Problem 2. *(Unbiased Covariance Estimator)* If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are random i.i.d. draws from a multivariate distribution, prove that the sample mean $\overline{\mathbf{X}}$ and sample covariance matrix $\mathbf{S}_n$ are unbiased estimators of expected value $\mathbb{E}(\mathbf{X})$ and covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X})$ of the multivariate distribution. The sample mean and covariance are defined as follows:

$$\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i, \quad \mathbf{S}_n = \frac{1}{n-1}\sum_{i=1}^{n} (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T.$$

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

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

a) Find the sample mean and the sample covariance matrix.

b) Using PCA, find the best orthogonal projection matrix $\mathbf{Q}$ for presenting the data in two dimensional space. Explain each step.

c) Show that the image of $\mathbf{Q}$ is the plane $x + y + z = 0$.