MSc I: Statistical Analysis (Spring Term)
Multivariate Analysis

Multivariate Analysis: Exercise 3

1. [Exam 2008]

   (a) Let a sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of size n be drawn on a random $p \times 1$ vector $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Show that the union-intersection procedure to test the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, a specified constant, against a the alternative hypothesis $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, without making any assumptions about $\boldsymbol{\Sigma}$, leads to the statistic known as Hotelling's T^2,

   $$T^2 = n(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu}_0),$$

   where $\overline{\mathbf{x}}$ is the sample mean $\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ and $\mathbf{S}$ is the *unbiased* sample covariance matrix $\frac{1}{n-1}\sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T$, assumed to be invertible.

   (b) Suppose that $n_1 = 11$ and $n_2 = 12$ observations are sampled from two different bivariate normal distributions that have a common covariance matrix $\boldsymbol{\Sigma}$ and possibly different mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. The sample mean vectors and pooled covariance matrix are:

   $$\overline{\mathbf{x}}_1 = \begin{bmatrix} -1 \\ -2 \end{bmatrix} \qquad \overline{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} 13 & 6 \\ 6 & 22 \end{bmatrix}$$

   (i) Use the Hotelling two-sample T^2 statistic to test for a difference in the population mean vectors.

   (ii) What is the estimate of the formula for Fisher's linear discriminant function? Explain how this is used to classify new observations. Consider an observation $\mathbf{x}_0 = \begin{bmatrix} 2, & 1 \end{bmatrix}^T$ on a new experimental unit. Was this unit more likely to have come from population 1 or population 2? (Assume equal misclassification costs and equal prior probabilities).

   [It may be assumed that if $T^2 \sim T_q^2(f)$, then $\frac{f-q+1}{fq} T^2 \sim F_{q, f-q+1}$].