IMPERIAL

MATH70092

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May 2024

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science

Multivariate Analysis

Date: Thursday, May 2, 2024
Time: 10:00 – 11:30 (BST)
Time Allowed: 1.5 hours
This paper has 3 Questions.
Please Answer All Questions in 1 Answer Booklet
Candidates should start their solutions to each question on a new sheet of paper.
Supplementary books may only be used after the relevant main book(s) are full.
Any required additional material(s) will be provided.
Allow margins for marking.
Credit will be given for all questions attempted.
Each question carries equal weight.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO

The open-book material allowed during the examinations consists of any material provided by the lecturers and annotated by the students, i.e. annotated lecture notes, annotated slides, and annotated problem class sheets. Books and electronic devices are not allowed.

1. (i) Let $\boldsymbol{X} \sim N_m(\mathbf{0}, \Sigma)$ where Σ is an $m \times m$ positive-definite matrix and $m \geq 2$. Let $\mathbf{v} \in \mathbb{R}^m$, $\mathbf{v} \neq 0$ be deterministic.

(a) Show that $Y = \mathbf{v}^\top \boldsymbol{X}$ is normally distributed on $\mathbb{R}$, and derive its mean and variance.

(b) What is the conditional distribution of $\boldsymbol{X} \mid Y = y$?

(c) Characterise the hyperplane on which the conditional distribution $\boldsymbol{X} \mid Y = y$ is defined.

(ii) Let $b \geq 0$. Define $\Sigma_b = \begin{pmatrix} 3 & b \\ b & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$.

(a) Justifying your answer, for what values of b is Σ_b is a covariance matrix?

(b) Suppose that Σ_b is a covariance matrix. Let $\boldsymbol{X} = (X_1, X_2)^\top \sim N_2(\boldsymbol{\mu}_0, \Sigma_b)$, where $\boldsymbol{\mu}_0 = (0, 0)^\top$. Let

$$\boldsymbol{Y} = \begin{pmatrix} X_1 + X_2 + 3 \\ X_1 - 2X_2 + 2 \\ X_2 - 3 \end{pmatrix}.$$

Show that $\boldsymbol{Y}$ has a multivariate normal distribution and calculate the mean and covariance matrix for $\boldsymbol{Y}$.

(c) Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be n iid samples from $N_2(\boldsymbol{\mu}, \Sigma_b)$ where b is a known parameter. Suppose that $\overline{\boldsymbol{X}} = (1, 0)^\top$ where $\overline{\boldsymbol{X}}$ is the sample mean. Show that the hypothesis $H_0 : \boldsymbol{\mu}^\top = (0, 0)$ will be rejected in favour of $H_0 : \boldsymbol{\mu}^\top \neq (0, 0)$ at size 95%, for b satisfying

$$3 - \frac{n}{c_2} \leq b^2 \leq 3,$$

where c_2 is the 95% quantile of the χ^2_2 distribution (i.e. the upper 5% point).

[Total 25 marks]

2. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ be independent $N_m(\boldsymbol{\mu}, \Sigma)$ random variables with $N > m$ and Σ positive definite. Ignoring the constant of proportionality, the likelihood function can be written as

$$L(\boldsymbol{\mu}, \Sigma) = (\det(\Sigma))^{-N/2} \mathrm{etr}\left(-\frac{1}{2}\Sigma^{-1}A\right) \exp\left(-\frac{1}{2}N(\overline{\boldsymbol{X}} - \boldsymbol{\mu})^\top \Sigma^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\right),$$

where $\overline{\boldsymbol{X}} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{X}_i$, $A = \sum_{i=1}^N (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)^\top$ and $\mathrm{etr}(\cdot)$ denotes $\exp(\mathrm{tr}(\cdot))$.

(i) Show that

$$\sup_{\boldsymbol{\mu}, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma) = N^{mN/2} \exp(-mN/2)(\det(A))^{-N/2}.$$

(ii) Let $m = 2$ so that $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, we wish to test the null hypothesis $H : \Sigma_{12} = \Sigma_{21}^T = 0$ against the alternative hypothesis K that says H is not true.

(a) Define the likelihood ratio statistic Λ_X for the null hypothesis $H : \Sigma_{12} = \Sigma_{21}^T = 0$ in terms of the likelihood function $L(\cdot, \cdot)$.

(b) Show

$$\Lambda_X^{2/N} = \frac{A_{11}A_{22} - A_{12}^2}{A_{11}A_{22}},$$

where $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$.

(iii) Consider the transformed random vectors $\boldsymbol{Y} = T\boldsymbol{X}$, where $T = \begin{pmatrix} t_{11} & 0 \\ 0 & t_{22} \end{pmatrix}$.

(a) Write down $\widetilde{A} = \sum_{i=1}^N (\boldsymbol{Y}_i - \overline{\boldsymbol{Y}}_i)(\boldsymbol{Y}_i - \overline{\boldsymbol{Y}}_i)^\top$ in terms of A.

(b) Let Λ_Y be the likelihood ratio statistic for the null hypothesis $H : \widetilde{\Sigma}_{12} = \widetilde{\Sigma}_{21}^T = 0$, where $\widetilde{\Sigma} = \mathrm{Cov}(\boldsymbol{Y})$. Show that $\Lambda_X = \Lambda_Y$.

[Total 25 marks]

3. (i) Let $\boldsymbol{X} \sim N_m(\mathbf{0}, \sigma^2 I_m)$ and let $\boldsymbol{Y} = \boldsymbol{X} + Z\mathbf{u}$, where $Z \sim N_1(0, \epsilon^2)$ is an independent Gaussian random variable, and $\mathbf{u} \in \mathbb{R}^m$, $||\mathbf{u}|| = 1$ be deterministic.

(a) Show that $\boldsymbol{Y} \sim N_m(\mathbf{0}, \Sigma)$, where

$$\Sigma = \sigma^2 I_m + \epsilon^2 \mathbf{u}\mathbf{u}^\top.$$

(b) Show that $\mathbf{u}$ is a principal loading vector with maximal variance.

(ii) Consider a population comprised of two classes C_1 and C_2, with proportion π_1 in C_1 and $\pi_2 = 1 - \pi_1$ in C_2. We assume that objects from class C_i are $N_m(\boldsymbol{\mu}_i, \Sigma)$ distributed with associated density $f_i(x)$, $i = 1, 2$. We define the region C_1 such that

$$\frac{f_1(x)}{f_2(x)} > \frac{\pi_2}{\pi_1}$$

holds, and define C_2 to be the complement of C_1.

(a) Taking $\boldsymbol{X}$ to be an m-dimensional normal vector from $N_m(\boldsymbol{\mu}_i, \Sigma)$ depending on the class it belongs to, show that the decision boundary takes the form $\boldsymbol{\alpha}^\top \boldsymbol{X} = c$, and provide explicit expressions for $\boldsymbol{\alpha} \in \mathbb{R}^m$ and $c \in \mathbb{R}$ in terms of Σ, $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and π_1.

(b) For $\boldsymbol{X} \sim N_m(\boldsymbol{\mu}_i, \Sigma)$, with $i = 1, 2$, let us define $\boldsymbol{Z} = T\boldsymbol{X} + \mathbf{b}$, where $\mathbf{b} \in \mathbb{R}^m$ and T is a non-singular $m \times m$ matrix. How is the decision boundary affected by this transformation?

(c) Suppose that $\boldsymbol{\mu}_1 = (0, 2)^\top$, $\Sigma = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$ and $\boldsymbol{\mu}_2 = (0, -2)^\top$ and $\pi_1 = \pi_2$. On a single plot, sketch the:

- contours of equal density for the distribution of C_1.
- contours of equal density for the distribution of C_2.
- the decision boundary.
- the vector $\boldsymbol{\alpha}$.

(d) Suppose that $\pi_1 \neq \pi_2$. Compute the decision boundary. Using the previous sketch, interpret how the class proportions affect the position of the boundary when $\pi_1 > \pi_2$ and $\pi_2 < \pi_1$?

[Total 25 marks]

End of examination paper

Imperial College London

Module:	MATH70092
Setter:	Duncan
Checker:	Kantas
Editor:	Varty
External:	Woods
Date:	March 15, 2024

MSc EXAMINATIONS (STATISTICS)
May 2023

MATH70092 Multivariate Analysis
Time: 1 hour and 30 minutes

Setter's signature	Checker's signature	Editor's signature
. . . .	. . . .	. . . .

1. (i) Let $\boldsymbol{X} \sim N_m(\mathbf{0}, \Sigma)$ where Σ is an $m \times m$ positive-definite matrix and $m \geq 2$. Let $\mathbf{v} \in \mathbb{R}^m$, $\mathbf{v} \neq 0$ be deterministic.

(a) Show that $Y = \mathbf{v}^\top \boldsymbol{X}$ is normally distributed on $\mathbb{R}$, and derive its mean and variance.

ANSWER: (SEEN) By definition, $\boldsymbol{X}$ is multivariate normal if any linear combination of the components is univariate normal. In particular $\mathbf{v}^\top \boldsymbol{X}$ is univariate normal. Taking expectations we see that $E[Y] = \mathbf{v}^\top \boldsymbol{\mu} = \mathbf{0}$ and $\mathrm{Var}(\mathbf{v}^\top \boldsymbol{X}) = \mathbf{v}^\top \Sigma \mathbf{v}$. [3 marks]

(b) What is the conditional distribution of $\boldsymbol{X} \mid Y = y$?

ANSWER: (UNSEEN) The random vector $(\boldsymbol{X}, Y)^\top$ is jointly Gaussian with zero-mean and covariance operator

$$\begin{pmatrix} \Sigma & \Sigma\mathbf{v} \\ \mathbf{v}^\top\Sigma & \mathbf{v}^\top\Sigma\mathbf{v} \end{pmatrix}.$$

Conditioning, the distribution of $\boldsymbol{X}$ given Y is multivariate normal with conditional distribution

$$N_m(\Sigma\mathbf{v}(\mathbf{v}^\top\Sigma\mathbf{v})^{-1}y, \Sigma - \Sigma\mathbf{v}(\mathbf{v}^\top\Sigma\mathbf{v})^{-1}\mathbf{v}^\top\Sigma),$$

which we can rewrite as

$$N_m\left(\frac{\Sigma\mathbf{v}y}{\mathbf{v}^\top\Sigma\mathbf{v}}, \Sigma - \frac{\Sigma\mathbf{v}\mathbf{v}^\top\Sigma}{\mathbf{v}^\top\Sigma\mathbf{v}}\right).$$

[3 marks]

(c) Characterise the hyperplane on which the conditional distribution $\boldsymbol{X} \mid Y = y$ is defined.

ANSWER: (UNSEEN) Let $\boldsymbol{X}_y$ denote the random variable $\boldsymbol{X}$ conditioned on $Y = y$. Consider

$$\mathrm{Var}[\mathbf{v} \cdot \mathbf{X}_y] = \mathbf{v}^\top\left(\Sigma - \frac{\Sigma\mathbf{v}\mathbf{v}^\top\Sigma}{\mathbf{v}^\top\Sigma\mathbf{v}}\right)\mathbf{v} = \mathbf{v}^\top\Sigma\mathbf{v} - \mathbf{v}^\top\Sigma\mathbf{v} = 0.$$

This implies that $\mathbf{v}^\top\boldsymbol{X}_y = \mathbf{v}^\top E[\boldsymbol{X}_y] = \mathbf{v}^\top\left(\frac{\Sigma\mathbf{v}y}{\mathbf{v}^\top\Sigma\mathbf{v}}\right) = y$. It follows that the support of the conditional distribution lies in the $m - 1$ dimensional hyperplane

$$\left\{\mathbf{x} \,:\, \sum_{i=1}^m v_i x_i = y\right\}.$$

[Students may also argue directly using the definition of conditional distribution, and this is also fine.] [6 marks]

(ii) Let $b \geq 0$. Define $\Sigma_b = \begin{pmatrix} 3 & b \\ b & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$.

[This question continues on the next page . . .]

(a) Justifying your answer, for what values of b is Σ_b is a covariance matrix?
ANSWER: (SEEN SIMILAR) Characteristic equation for eigenvalues

$$\left|\begin{pmatrix} 3-\lambda & b \\ b & 1-\lambda \end{pmatrix}\right| = 0,$$

so that $(3-\lambda)(1-\lambda) - b^2 = 0$ so that

$$3 - 4\lambda + \lambda^2 - b^2 = 0,$$

i.e.

$$\lambda = \frac{4 \pm \sqrt{4^2 - 4(3-b^2)}}{2} = 2 \pm \sqrt{1+b^2},$$

so that both eigenvalues are non-negative if and only if $1 + b^2 \leq 4$, i.e. $b^2 \leq 3$. [3 marks]

(b) Suppose that Σ_b is a covariance matrix. Let $\boldsymbol{X} = (X_1, X_2)^\top \sim N_2(\boldsymbol{\mu}_0, \Sigma_b)$, where $\boldsymbol{\mu}_0 = (0,0)^\top$. Let

$$\boldsymbol{Y} = \begin{pmatrix} X_1 + X_2 + 3 \\ X_1 - 2X_2 + 2 \\ X_2 - 3 \end{pmatrix}.$$

Show that $\boldsymbol{Y}$ has a multivariate normal distribution and calculate the mean and covariance matrix for $\boldsymbol{Y}$.
ANSWER: (SEEN SIMILAR) We can write

$$\boldsymbol{Y} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \\ 0 & 1 \end{pmatrix} \boldsymbol{X} + \begin{pmatrix} 3 \\ 2 \\ -3 \end{pmatrix}$$

And so, by Theorem 4.2.1 (no need to explicitly call the theorem out), the $\boldsymbol{Y}$ is $N_m(\mathbf{b}, \widetilde{\Sigma})$,

$$\mathbf{b} = \begin{pmatrix} 3 \\ 2 \\ -3 \end{pmatrix}$$

and

$$\begin{aligned}
\widetilde{\Sigma} &= \begin{pmatrix} 1 & 1 \\ 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & b \\ b & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & -2 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 \\ 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3+b & 3-2b & b \\ b+1 & b-2 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 4+2b & 1-b & b+1 \\ 1-b & 7-4b & b-2 \\ b+1 & b-2 & 1 \end{pmatrix}
\end{aligned}$$

[4 marks]

[This question continues on the next page . . .]

(c) Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be n iid samples from $N_2(\boldsymbol{\mu}, \Sigma_b)$ where b is a known parameter. Suppose that $\overline{\boldsymbol{X}} = (1, 0)^\top$ where $\overline{\boldsymbol{X}}$ is the sample mean. Show that the hypothesis $H_0 : \boldsymbol{\mu}^\top = (0, 0)$ will be rejected in favour of $H_0 : \boldsymbol{\mu}^\top \neq (0, 0)$ at size 95%, for b satisfying

$$3 - \frac{n}{c_2} \leq b^2 \leq 3,$$

where c_2 is the 95% quantile of the χ_2^2 distribution (i.e. the upper 5% point).
ANSWER: (UNSEEN) Since the variance matrix Σ is known, we can use the test-statistic

$$n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})^\top \Sigma^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}),$$

which has a χ_2^2 distribution. Plugging in the observed values

$$\begin{aligned} n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})^\top \Sigma^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}) &= n\begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & b \\ b & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= n\begin{pmatrix} 1 & 0 \end{pmatrix} \frac{1}{3 - b^2} \begin{pmatrix} 1 & -b \\ -b & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \frac{n}{3 - b^2}. \end{aligned}$$

It follows that the null hypothesis is rejected if

$$\frac{n}{3 - b^2} \geq c_2,$$

where c_2 is the 95% quantile of a χ_2^2 distribution. This is rejected if

$$\frac{n}{c_2} \geq 3 - b^2 \Rightarrow b^2 \geq 3 - n/c_2.$$

The upper bound on b^2 comes from the requirement that Σ is positive definite. [6 marks]

[Total 25 marks]

2. Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_N$ be independent $N_m(\boldsymbol{\mu}, \Sigma)$ random variables with $N > m$ and Σ positive definite. Ignoring the constant of proportionality, the likelihood function can be written as

$$L(\boldsymbol{\mu}, \Sigma) = (\det(\Sigma))^{-N/2} \mathrm{etr}\left(-\frac{1}{2}\Sigma^{-1}A\right) \exp\left(-\frac{1}{2}N(\overline{\boldsymbol{X}} - \boldsymbol{\mu})^\top \Sigma^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})\right),$$

where $\overline{\boldsymbol{X}} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{X}_i$, $A = \sum_{i=1}^N (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)^\top$ and $\mathrm{etr}(\cdot)$ denotes $\exp(\mathrm{tr}(\cdot))$.

(i) Show that

$$\sup_{\boldsymbol{\mu}, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma) = N^{mN/2} \exp(-mN/2)(\det(A))^{-N/2}.$$

ANSWER: (SEEN) $\overline{\boldsymbol{X}}$ and $N^{-1}A$ are the maximum likelihood estimators, therefore

$$\begin{aligned}
\sup_{\boldsymbol{\mu} \in \mathbb{R}^m, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma) &= L(\overline{\boldsymbol{X}}, N^{-1}A) \\
&= (\det(N^{-1}A))^{-N/2} \mathrm{etr}\left(-\frac{N}{2}A^{-1}A\right) \\
&= (N^{-m}\det(A))^{-N/2} \mathrm{etr}\left(-\frac{N}{2}I\right) \\
&= N^{mN/2} \exp(-mN/2)(\det(A))^{-N/2}.
\end{aligned}$$

[5 marks]

(ii) Let $m = 2$ so that $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, we wish to test the null hypothesis $H : \Sigma_{12} = \Sigma_{21}^T = 0$ against the alternative hypothesis K that says H is not true.

(a) Define the likelihood ratio statistic Λ_X for the null hypothesis $H : \Sigma_{12} = \Sigma_{21}^T = 0$ in terms of the likelihood function $L(\cdot, \cdot)$.
ANSWER: (SEEN SIMILAR) The likelihood ratio statistic Λ_X is given by

$$\Lambda_X = \frac{\sup_{\boldsymbol{\mu}, \Sigma_{11}, \Sigma_{22}} L\left(\boldsymbol{\mu}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right)}{\sup_{\boldsymbol{\mu}, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma)}.$$

[2 marks]

(b) Show

$$\Lambda_X^{2/N} = \frac{A_{11}A_{22} - A_{12}^2}{A_{11}A_{22}},$$

where $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$.
ANSWER: (UNSEEN) For the numerator we note that

$$\sup_{\boldsymbol{\mu}, \Sigma_{11}, \Sigma_{22}} L\left(\boldsymbol{\mu}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right) = \sup_{\Sigma_{11}, \Sigma_{22}} L\left(\overline{\boldsymbol{X}}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right),$$

[This question continues on the next page . . .]

which is equal to

$$\sup_{\Sigma_{11},\Sigma_{22}} \Sigma_{11}^{-N/2}\Sigma_{22}^{-N/2}\mathrm{etr}\left(-\frac{1}{2}(\Sigma_{11}^{-1}A_{11}+\Sigma_{22}^{-1}A_{22})\right).$$

Splitting this maximisation problem into two separate terms for Σ_{11} and Σ_{22} we have that

$$\sup_{\Sigma_{11}} \Sigma_{11}^{-N/2}\exp(-\Sigma_{11}^{-1}A_{11}/2) \text{ and } \sup_{\Sigma_{22}} \Sigma_{22}^{-N/2}\exp(-\Sigma_{22}^{-1}A_{22}/2).$$

The function $f(x) = x^{-N/2}exp(-\frac{1}{2}A_{11}/x)$ satisfies

$$\log f(x) = -(N/2)\log x - \frac{1}{2}A_{11}/x.$$

The first order optimality condition $-\frac{N/2}{x} + \frac{A_{11}}{2}/x^2 = 0$, so that $\Sigma_{11} = \frac{1}{N}A_{11}$, and similarly $\Sigma_{22} = \frac{1}{N}A_{22}$. Plugging these in we get

$$\Lambda = \frac{(A_{11}A_{22} - A_{12}^2)^{N/2}}{A_{11}^{N/2}A_{22}^{N/2}},$$

from which the result follows. [Students may also do a second derivative test, but its ok if not.] [10 marks]

(iii) Consider the transformed random vectors $\boldsymbol{Y} = T\boldsymbol{X}$, where $T = \begin{pmatrix} t_{11} & 0 \\ 0 & t_{22} \end{pmatrix}$.

(a) Write down $\widetilde{A} = \sum_{i=1}^N(\boldsymbol{Y}_i - \overline{\boldsymbol{Y}}_i)(\boldsymbol{Y}_i - \overline{\boldsymbol{Y}}_i)^\top$ in terms of A.
ANSWER: (SEEN SIMILAR) We have that

$$\widetilde{A} = \begin{pmatrix} t_{11}^2A_{11} & t_{11}t_{22}A_{12} \\ t_{11}t_{22}A_{21} & t_{22}^2A_{22} \end{pmatrix}$$

(b) Let Λ_Y be the likelihood ratio statistic for the null hypothesis $H : \widetilde{\Sigma}_{12} = \widetilde{\Sigma}_{21}^T = 0$, where $\widetilde{\Sigma} = \mathrm{Cov}(\boldsymbol{Y})$. Show that $\Lambda_X = \Lambda_Y$.
ANSWER: (UNSEEN) Computing the likelihood ratio

$$\Lambda_Y = \frac{t_{11}^2t_{22}^2A_{11}A_{22} - t_{11}^2t_{22}^2A_{12}^2}{t_{11}^2A_{11}t_{22}^2A_{22}} = \frac{A_{11}A_{22} - A_{12}^2}{A_{11}A_{22}} = \Lambda_X.$$

[8 marks]

[Total 25 marks]

3. (i) Let $\boldsymbol{X} \sim N_m(\mathbf{0}, \sigma^2 I_m)$ and let $\boldsymbol{Y} = \boldsymbol{X} + Z\mathbf{u}$, where $Z \sim N_1(0, \epsilon^2)$ is an independent Gaussian random variable, and $\mathbf{u} \in \mathbb{R}^m$, $||\mathbf{u}|| = 1$ be deterministic.

(a) Show that $\boldsymbol{Y} \sim N_m(\mathbf{0}, \Sigma)$, where

$$\Sigma = \sigma^2 I_m + \epsilon^2 \mathbf{u}\mathbf{u}^\top.$$

ANSWER: (SEEN SIMILAR) We can write

$$\boldsymbol{Y} = \begin{pmatrix} I & \mathbf{u} \end{pmatrix} \begin{pmatrix} \boldsymbol{X} \\ Z \end{pmatrix},$$

where $(\boldsymbol{X}, Z)^\top$ is a jointly normal random variable. $\boldsymbol{Y}$ will have zero mean and covariance

$$\begin{pmatrix} I & \mathbf{u} \end{pmatrix} \begin{pmatrix} \sigma^2 I & 0 \\ \mathbf{0} & \epsilon^2 \end{pmatrix} \begin{pmatrix} I \\ \mathbf{u} \end{pmatrix} = \sigma^2 I + \epsilon^2 \mathbf{u}\mathbf{u}^\top.$$

[2 marks]

(b) Show that $\mathbf{u}$ is a principal loading vector with maximal variance.
ANSWER: (UNSEEN) Clearly we have that

$$\left(\sigma^2 I + \epsilon^2 \mathbf{u}\mathbf{u}^\top\right) \mathbf{u} = (\sigma^2 + \epsilon^2)\mathbf{u},$$

so that $\mathbf{u}$ is an eigenvector of the covariance, with corresponding eigenvalue $\sigma^2 + \epsilon^2$. The remaining $m - 1$ eigenvectors $\mathbf{v}$ will be orthogonal to $\mathbf{u}$, and so $\left(\sigma^2 I + \epsilon^2 \mathbf{u}\mathbf{u}^\top\right) \mathbf{v} = \sigma^2 \mathbf{v}$. It follows that $\mathbf{u}$ has maximal variance [4 marks]

(ii) Consider a population comprised of two classes C_1 and C_2, with proportion π_1 in C_1 and $\pi_2 = 1 - \pi_1$ in C_2. We assume that objects from class C_i are $N_m(\boldsymbol{\mu}_i, \Sigma)$ distributed with associated density $f_i(x)$, $i = 1, 2$. We define the region C_1 such that

$$\frac{f_1(x)}{f_2(x)} > \frac{\pi_2}{\pi_1}$$

holds, and define C_2 to be the complement of C_1.

(a) Taking $\boldsymbol{X}$ to be an m-dimensional normal vector from $N_m(\boldsymbol{\mu}_i, \Sigma)$ depending on the class it belongs to, show that the decision boundary takes the form $\boldsymbol{\alpha}^\top \boldsymbol{X} = c$, and provide explicit expressions for $\boldsymbol{\alpha} \in \mathbb{R}^m$ and $c \in \mathbb{R}$ in terms of Σ, $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and π_1.
ANSWER: (SEEN) The decision boundary takes the form

$$\log f_1(\mathbf{x}) - \log f_2(\mathbf{x}) = \log \pi_2 - \log \pi_1.$$

In our Gaussian setting, with common covariance, we obtain

$$-(\mathbf{x}^\top - \boldsymbol{\mu}_1)^\top \Sigma^{-1} (\mathbf{x}^\top - \boldsymbol{\mu}_1) + (\mathbf{x}^\top - \boldsymbol{\mu}_2)^\top \Sigma^{-1} (\mathbf{x}^\top - \boldsymbol{\mu}_2) = \log \pi_2 - \log \pi_1.$$

[This question continues on the next page . . .]

So that
$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \Sigma^{-1}\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2).$$
We therefore define $\boldsymbol{\alpha} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ so that
$$\boldsymbol{\alpha}^\top \left(\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\right) = \log \pi_2 - \log \pi_1,$$
so that
$$\boldsymbol{\alpha}^\top \mathbf{x} = \log \pi_2 - \log \pi_1 + \frac{1}{2}\boldsymbol{\alpha}^\top(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) =: c.$$
This defines an $m-1$-dimensional hyperplane in $\mathbb{R}^m$, with points sitting on the side $\boldsymbol{\alpha}^\top\mathbf{x} > c$ classified as C_1 and points on the other-side classified as C_2. [4 marks]

(b) For $\boldsymbol{X} \sim N_m(\boldsymbol{\mu}_i, \Sigma)$, with $i = 1, 2$, let us define $\boldsymbol{Z} = T\boldsymbol{X} + \mathbf{b}$, where $\mathbf{b} \in \mathbb{R}^m$ and T is a non-singular $m \times m$ matrix. How is the decision boundary affected by this transformation?

ANSWER: (UNSEEN) Under the transformation, the two distributions map to multivariate normal distributions $N_m(\widetilde{\boldsymbol{\mu}}_i, \widetilde{\Sigma})$, where
$$\widetilde{\boldsymbol{\mu}}_i = T\boldsymbol{\mu}_i + \mathbf{b}, \text{ and } \widetilde{\Sigma} = T\Sigma T^\top.$$
The distributions have the same form as the original, and so we can calculate the decision boundary in a similar manner, as
$$\widetilde{\boldsymbol{\alpha}}^\top \mathbf{x} = \widetilde{c},$$
where
$$\widetilde{\boldsymbol{\alpha}} = (T^\top)^{-1}\Sigma^{-1}T^{-1}T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = (T^\top)^{-1}\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = (T^\top)^{-1}\boldsymbol{\alpha} = (T^{-1})^\top\boldsymbol{\alpha}.$$
and
$$\widetilde{c} = \log \pi_2 - \log \pi_1 + \frac{1}{2}\boldsymbol{\alpha}^\top T^{-1}T(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = c.$$
In conclusion, the decision boundary takes the form
$$\boldsymbol{\alpha}^\top T^{-1}\mathbf{x} = c.$$
[6 marks]

(c) Suppose that $\boldsymbol{\mu}_1 = (0, 2)^\top$, $\Sigma = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$ and $\boldsymbol{\mu}_2 = (0, -2)^\top$ and $\pi_1 = \pi_2$. On a single plot, sketch the:

- contours of equal density for the distribution of C_1.
- contours of equal density for the distribution of C_2.
- the decision boundary.

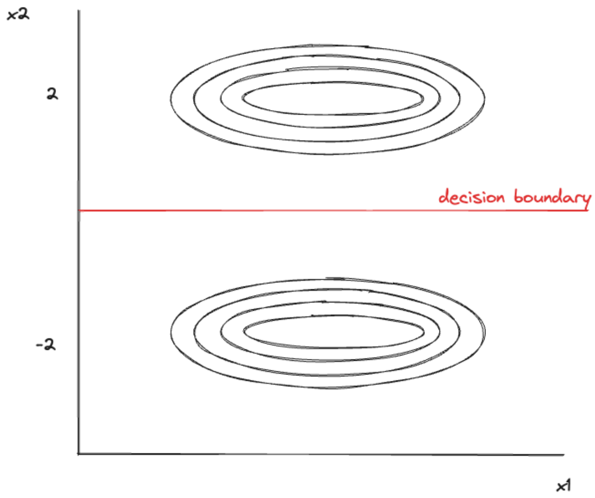
[This question continues on the next page . . .]

- the vector α.

ANSWER: (SEEN SIMILAR) $\alpha = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \end{pmatrix}$.
In addition, $c = 0$, so that the decision boundary is

$$(0, 4)^\top \mathbf{x} = 0 \Rightarrow x_2 = 0.$$

Plotting the above, we obtain



[6 marks]

(d) Suppose that $\pi_1 \neq \pi_2$. Compute the decision boundary. Using the previous sketch, interpret how the class proportions affect the position of the boundary when $\pi_1 > \pi_2$ and $\pi_2 < \pi_1$?
ANSWER: (UNSEEN) The class proportions will only affect the constant c. In this case, the decision boundary becomes

$$x_2 = \log(\pi_2/\pi_1),$$

so that the red line will move up if $\pi_2 > \pi_1$ and move down towards C_2 if $\pi_1 > \pi_2$. If π_1 is much bigger than π_2 this means that it is much more important to mitigate the misclassification of samples from C_1 instead of C_2. And so, we make the boundary much closer to the mode of the second distribution. [3 marks]

[Total 25 marks]

MATH70092 Multivariate Analysis

Question	Marker's comment
1	Overall students answered this one well, but quite a few students struggled with the computation of the conditional Gaussian distribution which identifies the hyperplane.nbsp; nbsp;This issue propagated down through parts of the question.nbsp; nbsp;Most students answered the second park well.
2	This question was quite similar to exercises done during class, so most students have a fairly clear idea of how to approach this.nbsp; nbsp;Some common mistakes involved actually defining the likelihood ratio, i.e. what are we taking supremums over.nbsp; The last part was mostly answered correctly.
3	Students did reasonably well in this question, though many struggled with the first part on principal component analysis.nbsp; nbsp;Students were able to identify the principal loading vector, but then some struggled to actually show it was the PLV.nbsp; nbsp;Some students struggled with the calculation of the decision boundary, but broadly everyone followed the correct strategy