## MAST 90138: MULTIVARIATE STATISTICAL TECHNIQUES

#### **Assessment:**

- Three assignments (15% each) due at dates TBD spread over the course.
- Final written examination (55%).

## Got question about course material?

- Ask your peers on Ed discussion
- or ask instructor/tutor during office hours and tutorials
- Questions are not to be asked by email.

## Office hours:

Regular weeks (subject to change, changes announced on LMS):

- Tu 3.00-4.00: room 189, Peter Hall building, start week 2,
- Wed 3.15-4.15: room 189, PHB
- Fri 11.00-12.00: Zoom, meeting ID 81136149049.

Aug 4-11: see subject overview on LMS.

#### Other comments

- The tutorials will involve the software R.
- You need to attend the tutorial that was assigned to you as the lists have been designed based on the computer rooms capacity.
- The official slides of the course will be the ones posted AFTER the class.
- I will try to post a version of the slides BEFORE the class. This will NOT be the official version of the slides.

I will often replace those slides after the class to post the official version of the slides.

## MAST 90138: MULTIVARIATE STATISTICAL TECHNIQUES

#### **1** Introduction

Goal of the course: Describe, understand and discover properties of data in p>1 dimensions. 为的数据,严肃的 每何是了第一

• We are interested in analysing a sample of n random vectors  $X_1, \ldots, X_n$ , also denoted by  $X_i$  for  $i = 1, \ldots, n$ , where  $X_1, X_2$ , etc all belong to  $\mathbb{R}^p$ :

$$X_1 = (X_{11}, \dots, X_{1p}) \in \mathbb{R}^p$$
  
 $X_2 = (X_{21}, \dots, X_{2p}) \in \mathbb{R}^p$   
:  
:  
 $X_n = (X_{n1}, \dots, X_{np}) \in \mathbb{R}^p$ .

• This means that we have a sample of n individuals, and that for each individual we observe p variables (sometimes called features).

•  $X_i = (X_{i1}, \dots, X_{ip})$  is a frow vector

$$X_i^T = \begin{pmatrix} X_{i1} \\ \dots \\ X_{ip} \end{pmatrix}$$
 is a column vector.

## Example:

- We consider a health study involving n = 100 patients.
- On each patient we measure p=4 quantities: age, weight, body mass index, systolic blood pressure.
- For the *i*th individual, where i = 1, ..., 100 we observe the vector  $X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$ .
- $X_{i1}$ =age of ith patient  $X_{i2}$ =weight of ith patient  $X_{i3}$ = body mass index of ith patient  $X_{i4}$ =systolic blood pressure of ith patient.

Often we gather all the observations into a single  $n \times p$  matrix

$$\mathcal{X} = \left(egin{array}{ccc} X_{11} & \dots & X_{1p} \ X_{21} & \dots & X_{2p} \ & dots \ X_{n1} & \dots & X_{np} \end{array}
ight)$$

- ightharpoonup Each  $X_{ij}$ , for  $i=1,\ldots,n$  and  $j=1,\ldots,p$  is a random variable.
- $\frown$   $\mathcal{X}$  has n rows and p columns.
- ightharpoonup The *i*th row represents the *p* variables corresponding to the *i*th individual,
- the jth column represents the jth variable for all n individuals.
- We call the value taken by  $X_i = (X_{i1}, \dots, X_{ip})$  the observed value or realisation.
- Often we use a lower case to denote the observed value:  $x_i = (x_{i1}, \dots, x_{ip})$  is the observed value (or realisation) of the random vector  $X_i$ .

## What we will do in the course:

- The technique learned in basic statistics courses are not focused on multivariate data.
- For ex, how would we do a scatterplot in more than p = 2 dimensions? How to represent graphically multivariate data?
- In this course we will learn techniques especially designed for multivariate data
- This course will introduce various CORE statistical methods for analysing and describing multivariate data.
- Topics such as neural networks, support vector machine, graphical models, etc will NOT be covered in this course.

## MATERIAL COVERED IN THE COURSE:

- Revision of basic matrix results and multivariate data needed in the course (tentative schedule: weeks 1 and 2).
  - As this is just a revision, we won't spend much time on examples but the reference book has examples.
- Principal component analysis, factor analysis, partial least squares.
- Classification (linear and quadratic discriminant, regression-based, including trees and forest).
- Clustering (*K*-means, *K*-medoids, hierarchical).
- If time permits, other topics such as correspondence analysis or more techniques for above topics.

## Main reference books:

• Härdle, W. and Simar, L (2015). Applied multivariate statistical analysis, 4th edition.

Available for download at the unimelb library website.

• Hastie, T. Tibshirani, R. and Friedman, J. (2009). The elements of statistical learning, 2nd edition (available online):

https://hastie.su.domains/Papers/ESLII.pdf

## We might also use

- Koch, I. (2013). Analysis of Multivariate and High-Dimensional Data. Cambridge University Press.
- K. V. Mardia, J. T. Kent, & J. M. Bibby (1979). Multivariate Analysis. New York: Academic Press.
- Other resources as needed.

#### 2 REVIEW OF MATRIX PROPERTIES

Sections 2.1, 2.2, 2.3, 2.4 and 2.6 in Härdle and Simar (2015).

#### **2.1** ELEMENTARY OPERATIONS

• Let  $A = (a_{ij}) = (a_{ij})_{1 \le i \le n, 1 \le j \le p}$  be an  $n \times p$  matrix (matrix with n rows and p columns):

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{np} \end{pmatrix}$$

• We use  $A^T$  to denote the transpose of the matrix A:

$$A^{T} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & & & \\ a_{1p} & a_{2p} & \dots & a_{np} \end{pmatrix}$$

(rows become columns and columns become rows).

• Product of two matrices A, an  $n \times p$  matrix, and B, a  $p \times m$  matrix:

$$AB = C$$
,

where C is an  $n \times m$  matrix whose (i, j)th element is given by:

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$$

• Scalar product of two p dimensional vectors  $x = (x_1, \dots, x_p)^T$  and  $y = (y_1, \dots, y_p)^T$ :

$$x^T y = \sum_{j=1}^p x_j y_j.$$

• Norm of a p dimensional vector  $x = (x_1, \dots, x_p)^T$  or  $x = (x_1, \dots, x_p)$ :

$$||x|| = \sqrt{\sum_{j=1}^{p} x_j^2} = \begin{cases} \sqrt{x^T x} & \text{if } x = (x_1, \dots, x_p)^T \\ \sqrt{x x^T} & \text{if } x = (x_1, \dots, x_p) \end{cases}$$

# Some properties:

BUT:

$$A + B = B + A$$

$$A(BC) = (AB)C$$

$$(AB)^{T} = A$$

$$(AB)^{T} = B^{T}A^{T}$$

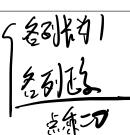
$$AC \neq CA$$

# Special matrices (taken from Härdle and Simar):

 Table 2.1 Special matrices and vectors

Name	Definition	Notation	Example
Scalar	p = n = 1	a	3
Column vector	p = 1	a	$\begin{pmatrix} 1 \\ 3 \end{pmatrix}$
Row vector	n = 1	$a^{\top}$	$\left  \begin{pmatrix} 1 & 3 \end{pmatrix} \right $
Vector of ones	$(\underbrace{1,\ldots,1}_n)^{\top}$	$1_n$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
Vector of zeros	$(\underbrace{0,\ldots,0}_{n})^{\top}$	$O_n$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
Square matrix	n = p	$\mathcal{A}(p \times p)$	$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$
Diagonal matrix	$a_{ij} = 0, i \neq j, n = p$	$diag(a_{ii})$	$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$

Identity matrix	$\operatorname{diag}(\underbrace{1,\ldots,1}_{p})$	$igg _{{\mathcal I}_p}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
Unit matrix	$a_{ij}=1, n=p$	$1_n 1_n^{\top}$	$ \left(\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array}\right) $
Symmetric matrix	$a_{ij}=a_{ji}$		$ \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} $
Null matrix	$a_{ij} = 0$	0	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
Upper triangular matrix	$a_{ij} = 0, i < j$		$ \begin{array}{ c c c } \hline \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{array} $
Idempotent matrix	AA = A		$ \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{array}\right) $
Orthogonal matrix	$\mathcal{A}^{\top}\mathcal{A} = \mathcal{I} = \mathcal{A}\mathcal{A}^{\top} \qquad \text{Algorithms}$	-1	$ \left(\begin{array}{cc} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{array}\right) $



• A set of k rows (or columns)  $a_1, \ldots, a_k$  of A are said to be linearly independent if none of them can be expressed as a nontrivial linear combination of the other k-1 rows (or columns). It is not possible to write

$$a_j = \sum_{i \neq j} c_i a_i$$
, 对作意识的都没法的类似。通过

where the  $c_i$ 's are all different from zero, or equivalently



• Rank: The rank of an  $n \times p$  matrix A, denoted by rank(A) is defined as the maximum number of linearly independent rows or columns).

We always have

$$\operatorname{rank}(A) \leq \min(n, p)$$

• Trace: The trace of a square  $p \times p$  matrix A, denoted by tr(A), is the sum of its diagonal elements:

$$\operatorname{tr}(A) = \sum_{i=1}^{p} a_{ii}.$$

• Determinant: The determinant of a square  $p \times p$  matrix A, denoted by [et(A) or |A|], is a number computed from the matrix and which plays an important role in all sorts of problems. For a  $2 \times 2$  matrix

$$A = \left(\begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array}\right)$$

it is computed by

$$|A| = a_{11}a_{22} - a_{21}a_{12}$$
.

For larger matrices, we compute this recursively. If you forgot how a determinant is computed, see

https://en.wikipedia.org/wiki/Determinant

• If rank of 
$$p \times p$$
 matrix  $A$  not full  $(\operatorname{rank}(A) < p)$  then  $\det(A) = 0$ .

• Inverse: If  $|A| \neq 0$ , the inverse of a square  $p \times p$  matrix A exists. It is denoted by  $A^{-1}$  and is such that  $A^{-1} = A + 0$  detays  $A^{-1} = A + 0$ 

full runk 
$$\Rightarrow$$
 [A] to det (D)  $\Rightarrow$  non Singular.
$$AA^{-1} = A^{-1}A = I_p$$

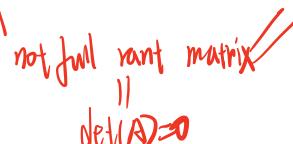
 $(I_p = p \times p \text{ identity matrix}).$ 

See Härdle and Simar, page 57, for how to compute the inverse.

We have

$$|A^{-1}| = 1/|A|$$
.

• A non invertible square matrix is also said to be singular.



• Eigenvalues and eigenvectors of a square  $p \times p$  matrix A:

The (non zero)  $p \times 1$  vector v is an eigenvector of A with eigenvalue  $\lambda$  if it is such that

$$Av = \lambda v$$
.

Note that  $\lambda$  is a real number (not a vector).

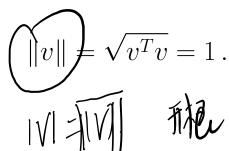
- ightharpoonup If A is symmetric then there are p eigenvalues and eigenvectors.
- All eigenvalues satisfy

$$|A - \lambda I_p| = 0$$

(they are the p roots of the above polynomial of order p in  $\lambda$ ).

The eigenvalues are not necessarily all different from each other.

- Constant multiples of an eigen vector v with eigenvalue  $\lambda$  are also eigenvectors with eigenvalue  $\lambda$ .



- $\blacksquare$  In practice we often compute them with a software, e.g. R.
- Suppose the square  $p \times p$  matrix A has eigenvalues  $\lambda_1, \ldots, \lambda_p$ .

Let  $\Lambda$  be the diagonal matrix  $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_p)$  with the  $\lambda_i$ 's on the diagonal and 0 everywhere else.

Then we have

$$\det(A) = \underbrace{(A|=|\Lambda)} = \prod_{i=1}^{p} \lambda_i \quad \text{iff} = \det(A)$$

$$\underbrace{\text{product}}_{\text{of. eigenvalues}}$$

and

$$\operatorname{tr}(A) = \operatorname{tr}(\Lambda) = \sum_{i=1}^p \lambda_i$$
 .

• If *A* and *B* are two invertible matrices of the same dimension, then

$$(A \cdot B)^{-1} = B^{-1}A^{-1}.$$

Some other properties of matrix characteristics (from Härdle and Simar):

$$\mathcal{A}(n \times p)$$
,  $\mathcal{B}(p \times q)$ ,  $\mathcal{C}(q \times n)$ 

$$=\operatorname{tr}(\mathcal{ABC})=\operatorname{tr}(\mathcal{CAB})=\operatorname{tr}(\mathcal{BCA})$$

$$\mathcal{A}(n \times n), \ \mathcal{B}(n \times n), \ c \in \mathbb{R}$$

$$\operatorname{tr}(\mathcal{A} + \mathcal{B}) = \operatorname{tr} \mathcal{A} + \operatorname{tr} \mathcal{B}$$

$$tr(cA) = c tr A$$

$$|AB| = |BA| = |A||B|$$

$$\mathcal{A}(n \times p), \ \mathcal{B}(p \times n)$$

$$\operatorname{tr}(\mathcal{A} \cdot \mathcal{B}) = \operatorname{tr}(\mathcal{B} \cdot \mathcal{A})$$

$$\operatorname{rank}(\mathcal{A}) \leq \min(n, p)$$

$$\operatorname{rank}(\mathcal{A}) \geq 0 \quad \text{for ank}(\mathcal{A}) \geq 0 \quad \text{for ank}(\mathcal{A}) = \operatorname{rank}(\mathcal{A}) \quad \text{for ank}(\mathcal{A}) \quad \text{for ank}(\mathcal{A}) \quad \text{for ank}(\mathcal{A}) \quad \text{for ank}(\mathcal{A}) \leq \operatorname{rank}(\mathcal{A}) + \operatorname{rank}(\mathcal{B})$$

$$\operatorname{rank}(\mathcal{A}\mathcal{B}) \leq \min\{\operatorname{rank}(\mathcal{A}), \operatorname{rank}(\mathcal{B})\}$$

Errata: The dimensions of A and B have to match for A + B to be defined in

$$rank(\mathcal{A} + \mathcal{B}) \le rank(\mathcal{A}) + rank(\mathcal{B}),$$

#### **2.2** Spectral decompositions

# Spectral decomposition

• Spectral decomposition: Suppose A is a square and symmetric  $p \times p$  matrix and let

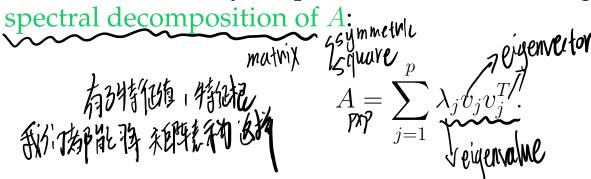
ម្រែក្រាស្ត្រ  $\lambda_1,\ldots,\lambda_p$  denote its p eigenvalues and  $\lambda_1$  ិ

 $v_1, \ldots, v_p$  denote the associated  $p \times 1$  eigenvectors of norm 1 and orthogonal to each other (each  $v_k$  is the eigenvector associated to  $\lambda_k$ ).

Note: two  $p \times 1$  column vectors v and w are orthogonal if

$$v^T w = \sum_{i=1}^p v_i w_i = 0.$$

Then we can always express A in the following way, which is called



This can also be written in matrix form, if we let

and

$$\begin{array}{c} \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p) \\ \text{Gamma} \\ \Gamma = (v_1, \ldots, v_p) \end{array}$$

the  $p \times p$  orthogonal matrix whose columns are the p eigenvectors standardised to be of norm 1: A HAMPED AND POPPER TO THE PROPERTY OF THE

$$A = \Gamma \Lambda \Gamma .$$

• In the above notation, if  $A = \Gamma \Lambda \Gamma^T$  then if we take an integer power of A, for example  $A^{\alpha}$ , we find

This is because the  $v_j$ 's are orthonormal (i.e. orthogonal and of norm 1).

For example

$$A^2 = \Gamma \Lambda \Gamma^T \Gamma \Lambda \Gamma^T = \Gamma \Lambda^2 \Gamma^T$$
 .

This also works for negative powers if (A is invertible) (which happens if and only if the eigenvalues are all honzero). For example,

$$A^{-1} = \Gamma \Lambda^{-1} \Gamma^T$$

 $(A^{-1}$ : the inverse of the matrix A).

# Singular value decomposition

More generally, a similar decomposition exists for matrices that are not especially square matrices. In particular, any  $n \times p$  matrix A with rank r can be decomposed as

$$A = \prod_{n \neq p} \prod_{n \neq p}$$

where the  $n \times r$  matrix  $\Gamma$  and the  $p \times r$  matrix  $\Delta$  are column orthonormal, i.e. their columns are orthonormal (i.e., orthogonal and of norm 1), that is

$$\Gamma^T \Gamma = \Delta^T \Delta = I_r \quad \Gamma \quad \Delta = 1$$

and

$$\Lambda = \operatorname{diag}(\lambda_1^{1/2}, \dots, \lambda_r^{1/2})$$

$$\lambda_r = \lambda_r = \lambda_r = \lambda_r = \lambda_r$$

where each  $\lambda_i > 0$ .

The  $\lambda_i$ 's are the nonzero eigenvalues of the matrices  $AA^T$  or  $A^TA$ ; the columns of  $\Gamma$  and  $\Delta$  are the corresponding r eigenvectors of those matrices, respectively.

$$A = \binom{1}{1} \binom{0}{2}$$

$$Yunk(0) = 2$$
because its impossible to express  $(1,1,0) = 6 \cdot (0,1,1)$ 

$$AA^{T} = \binom{1}{0} \binom{1}{1} \binom{1}{1} = \binom{1}{1} \binom{1}{2} \binom$$

 $\left\{ \begin{pmatrix} 2 \\ 1 2 \end{pmatrix} \begin{pmatrix} Y_{4} \\ Y_{2} \end{pmatrix} = \frac{3}{3} \begin{pmatrix} Y_{4} \\ V_{12} \end{pmatrix} \right\}$ 

VI= (VI) associated to 71.=1

 $A = [ ] / \Delta^T$