Australian National University | PROGRAMS AND COURSES

CLASS NUMBER 5612 TERM CODE 3360

CLASS INFO
Unit Value
6 units
Mode of Delivery
In Person
COURSE CONVENER
AsPr Dale Roberts

CLASS DATES
Class Start Date
24/07/2023
Class End Date
27/10/2023
Census Date
31/08/2023

Big Data Statistics

An Undergraduate course offered by the **Rsch Sch of Finance, Actuarial Studies & App Stats**.

STAT3017

CLASS NUMBER 5612 TERM CODE 3360

CLASS INFO
Unit Value
6 units
Mode of Delivery
In Person
COURSE CONVENER
AsPr Dale Roberts
LECTURER
AsPr Dale Roberts

CLASS DATES
Class Start Date
24/07/2023
Class End Date
27/10/2023
Census Date
31/08/2023
Last Date to Enrol
31/07/2023

Overview | Assessment | Submission | Contacts

This research-led course provides an introduction to recent developments in Random Matrix Theory and Online Learning that addresses the challenges and opportunities posed by the availability of large amounts of data.

In the first instance, we will review some classic results from multivariate statistical theory, matrix analysis, and probability theory. Then we will present the salient statistical features of big data (e.g., heterogeneity, noise accumulation, spurious correlation, and incidental endogeneity) and show how this impacts on traditional statistical methods and theory.

We follow with an introduction to modern Random Matrix theory and its application in statistics. Applications presented may include topics such as high-dimensional statistical inference, large covariance matrices, large-scale statistical learning through subsampling, sparsification of large matrices, principal component analysis, and dimension reduction.

We conclude with an introduction to the theory of online learning (aka. sequential prediction) to handle the situation of streaming data.

Students will use and learn about the latest computational tools to work with big and streaming data sets. Example data sets may be drawn from areas such finance, web analytics, digital marketing, and satellite imagery data.

Learning Outcomes

Upon successful completion, students will have the knowledge and skills to:

1. Explain how statistical features of big data impact traditional statistical methods and theory.
2. Discuss Random Matrix theory and its application in statistics on large scale.
3. Summarise the theory of sequential prediction and management of streaming data.
4. Demonstrate the use of computational tools to work with big and streaming data sets.

Research-Led Teaching

This course is based on recent research papers and surveys applications of random matrix theory in statistics. The topic is rapidly advancing and recent results may be introduced into the course

as they appear in the literature.

Examination Material or equipment

Working version of RStudio and R.

Required Resources

Working version of RStudio and R. Resources (research papers, etc) and lecture notes will be provided throughout the semester on Wattle.

Staff Feedback

Students will be given feedback in the following forms in this course:

- Written comments
- Verbal comments
- Feedback to the whole class, to groups, to individuals, focus groups

Student Feedback

ANU is committed to the demonstration of educational excellence and regularly seeks feedback from students. Students are encouraged to offer feedback directly to their Course Convener or through their College and Course representatives (if applicable). The feedback given in these surveys is anonymous and provides the Colleges, University Education Committee and Academic Board with opportunities to recognise excellent teaching, and opportunities for improvement. The Surveys and Evaluation website provides more information on student surveys at ANU and reports on the feedback provided on ANU courses.

Class Schedule

WEEK/SESSION	SUMMARY OF ACTIVITIES	ASSESSMENT
1	Introduction to the challenges of Big Data and overview of the course. Review of some prerequisite concepts.	
2	Further matrix analysis, eigenvalues and eigenvectors, the multivariate normal distribution.	
3	Fundamental tools for studying limiting spectral distributions, Marcenko-Pastur distributions, Fisher spectral distribution.	Assessment 1 Due (20%)
4	CLT for linear spectral statistics: Introduction and integration tools.	
5	Moments and statistics of the Marcenko-Pastur distribution.	
6	CLT for linear spectral statistics: Sample covariance matrix, Bai and Silverstein's CLT, CLT for random Fisher matrices.	Assessment 2 Due (20%)
7	Generalised variance in higher dimensions.	Assessment 3 Due (20%)
8	Multiple correlation coefficient.	
9	Multivariate linear regression in the high-dimensional setting.	Assessment 4 Due (20%)
10	PCA and high-dimensional spiked population models.	
11	Applications and recent theoretical results.	
12	Applications and recent theoretical results.	Final Project Due (20%)

Tutorial Registration

Tutorial registration will be available two weeks prior to the beginning of the semester and will close at the end of week 1. More details can be found on the Timetable webpage. https://www.anu.edu.au/students/program-administration/timetabling.

Assessment Summary

ASSESSMENT TASK	VALUE	DUE DATE	RETURN OF ASSESSMENT	LEARNING OUTCOMES
Assessment 1	20 %	08/08/2023	22/08/2023	1,2,3,4
Assessment 2	20 %	29/08/2023	12/09/2023	1,2,3,4
Assessment 3	20 %	19/09/2023	03/10/2023	1,2,3,4
Assessment 4	20 %	04/10/2023	18/10/2023	1,2,3,4
Final Project	20 %	02/11/2023	30/11/2023	1,2,3,4

* If the Due Date and Return of Assessment date are blank, see the Assessment Tab for specific Assessment Task details

Policies

ANU has educational policies, procedures and guidelines, which are designed to ensure that staff and students are aware of the University's academic standards, and implement them. Students are expected to have read the Academic Misconduct Rule before the commencement of their course. Other key policies and guidelines include:

- Student Assessment (Coursework) Policy and Procedure
- Special Assessment Consideration Policy and General Information
- Student Surveys and Evaluations
- Deferred Examinations
- Student Complaint Resolution Policy and Procedure

Responsible Officer: Registrar, Student Administration / **Page Contact:** Website Administrator / Frequently Asked Questions