



UNSW SCIENCE
School of Maths and Statistics

Course outline

MATH5836
Data Mining

Term 3, 2022

Staff

Position	Name	Email	Room
Lecturer-in-charge	Dr Rohitash Chandra	Rohitash.chandra@unsw.edu.au	RC-2

Please refer to your Timetable on MyUNSW for your Lecture Tut, Lab enrolment days and times.

Timetable weblink: <https://timetable.unsw.edu.au/2022/MATH5836.html#S3S>

Note: We will use BB Collaborate for the Lecture/Tutorial sessions, these will be fully online.

Administrative Contacts

Please visit the School of Mathematics and Statistics website for a range of information on School Policies, Forms and Help for Students.

For information on Courses, please go to “Student Life & resources page” and either Undergraduate Courses and/or Postgraduate Courses for information on all course offerings,

The “Student Notice Board” can be located by going to the “Student Life & resources” page; Notices are posted regularly for your information here. Please familiarise yourself with the information found in these locations. The School web page is: <https://www.maths.unsw.edu.au>

If you cannot find the answer to your queries on the web you are welcome to contact the Student Services Office directly.

By email Postgraduate pg.mathsstats@unsw.edu.au

By phone: 9385 7053

Should we need to contact you, we will use your official UNSW email address of in the first instance. **It is your responsibility to regularly check your university email account. Please state your student number in all emails.**

Course Information

Assumed knowledge / Pre-Requisite: A prerequisite for this course is MATH2801/2901. A recommended prerequisite is MATH2831/2931 (Higher Linear Models). If you have not done the recommended prerequisite courses, make sure that you know sufficient statistical theory, can integrate and differentiate, and have competency in at least a programming language such as R or Python.

We are aware some course exclusions on the Handbook may be different to the School website. We are in the process of updating this information. Meanwhile, students should be following the Handbook course information with the School website information as a supplement.

Course Aims

This course is expected to give students an understanding of the fundamentals of machine learning and the basics of data mining, which is essential for anyone contemplating a career as a professional statistician or data analyst in industries reliant upon such expertise. The student should develop a working knowledge of the statistical and theoretical underpinnings of the topics covered.

Given this fundamental statistical understanding of these methodologies, this will allow the student to utilise these techniques with confidence on real-world data sets and scenarios. As such the student is expected to develop applied working knowledge of the methodologies covered, largely through practical applications. In addition, students will undertake additional reading of a collection of associated research papers in each topic, to further add context to the methodologies presented during the course. This will enhance the student's ability to utilise these techniques to solve real-world problems. It is stressed that this course is aimed at fundamental statistical properties of these methods, it is not a course on the application of computer software.

Course Description

Increasingly, organisations need to analyse enormous data sets to extract useful information. In response to this, a range of statistical and machine learning methods have been developed in recent times. This course covers the key techniques in data mining and machine learning with theoretical background and applications. The topics include methods such as linear and logistic regression, neural networks, Bayesian neural networks, clustering and dimensionality reduction, ensemble learning, and an introduction to deep learning. Emerging machine learning tools and libraries are used to illustrate the methods in programming environments that includes Python and R.

Rationale

New ideas and skills are introduced and demonstrated in lectures and through the recommended reading of supplementary material such as research papers, then students develop these skills by applying them to specific tasks in assessments. We believe that effective learning is best supported by a climate of inquiry, in which students are actively engaged in the learning process. Hence this course is structured with a strong emphasis on problem-solving tasks. Students are expected to devote the majority of their class and study time to solving such tasks. New ideas and skills are first introduced and demonstrated in lectures, and then students develop these skills by applying them to specific tasks in assessments. This course has a major focus on research, inquiry and analytical thinking as well as information literacy. We will also explore capacity and motivation for intellectual development through the solution of both simple and complex mathematical models of problems arising in finance, economics and engineering, and the interpretation and communication of the results.

Assessment and Deadlines

Assessment	Week	Weighting %	Due date if applicable
Assessment 1: Online Quiz. Duration 30 minutes	3	5%	Wednesday 6pm
Assessment 2	5	20%	Saturday 10pm
Assessment 3	10	25%	Saturday 10pm
Assessment 4: Final Exam. Duration 3 hours		50%	Exam period

Note that Python is the designated language for the course and minimum support on R will be given. You can submit the assessments in either R or Python. You need to get a 40% minimum in the final exam to pass the course; i.e if the final exam has a total of 100 marks, you will need to get at least 40/100 to pass the course.

Late Submission of Assessment Tasks

Assessment 1: No late submission accepted (apply for special consideration in special cases)
Assessments 2 and 3: A late penalty of 10% of the awarded mark will be applied per day. Any assessment task submitted 5 or more days late will be given zero.

Course Learning Outcomes (CLO)

- CLO1- Demonstrate an understanding of the fundamentals of machine learning and basics of data mining.
- CLO1- Demonstrate a working knowledge of the statistical and theoretical underpinnings of the methods.
- CLO3- Demonstrate an applied working knowledge of the methodologies covered with practical assignments.

Course Schedule

The course will include material taken from some of the following topics. This is should only serve as a guide as it is not an extensive list of the material to be covered and the timings are approximate. The course content is ultimately defined by the material covered in lectures.

Weeks	Topic
0	Python and R Tutorials (no Lectures)
1	Data Processing and Introduction to Data mining
2	Logistic Regression and Evaluation
3	Intro to Neural Networks
4	Advances in Neural Networks
5	Bayesian Neural Networks
7	Trees and Forests
8	Ensemble Learning
9	Unsupervised Learning and Dimensionality Reduction
10	Emerging Topics in Data Mining

Textbooks

Géron. A, 2019, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems, O'Reilly, second edition: <https://www.bookshop.unsw.edu.au/details.cgi?ITEMNO=9781492032649> (a copy would be handy but not required)

EXTRA READING MATERIALS (optional)

Mitchell. Tom, 1997, Machine Learning, McGraw-Hill. (additional textbook for reference): <https://www.amazon.com.au/Machine-Learning-Thomas-Mitchell/dp/0070428077>

Kroese, Botev, Tamire & Vaisman (2020), Data Science and Machine Learning, Chapman and Hall: <https://www.bookshop.unsw.edu.au/details.cgi?ITEMNO=9781138492530>

Moodle

Log in to Moodle to find announcements, general information, notes, lecture slide, classroom tutorial and assessments etc.

<https://moodle.telt.unsw.edu.au>

School and UNSW Policies

The School of Mathematics and Statistics has adopted a number of policies relating to enrolment, attendance, assessment, plagiarism, cheating, special consideration etc. These are in addition to the Policies of The University of New South Wales. Individual courses may also adopt other policies in addition to or replacing some of the School ones. These will be clearly notified in the Course Initial Handout and on the Course Home Pages on the Maths Stats web site.

Students in courses run by the School of Mathematics and Statistics should be aware of the School and Course policies by reading the appropriate pages on the Maths Stats web site starting at:

<https://www.maths.unsw.edu.au/currentstudents/assessment-policies>

The School of Mathematics and Statistics will assume that all its students have read and understood the School policies on the above pages and any individual course policies on the Course Handout and Course Home Page. Lack of knowledge about a policy will not be an excuse for failing to follow the procedure in it.

Academic Integrity and Plagiarism

UNSW has an ongoing commitment to fostering a culture of learning informed by academic integrity. All UNSW staff and students have a responsibility to adhere to this principle of academic integrity. Plagiarism undermines academic integrity and is not tolerated at UNSW. *Plagiarism at UNSW is defined as using the words or ideas of others and passing them off as your own.*

The **UNSW Student Code** provides a framework for the standard of conduct expected of UNSW students with respect to their academic integrity and behaviour. It outlines the primary obligations of students and directs staff and students to the Code and related procedures.

In addition, it is important that students understand that it is not permissible to buy essay/writing services from third parties as the use of such services constitutes plagiarism because it involves using the words or ideas of others and passing them off as your own. Nor is it permissible to sell copies of lecture or tutorial notes as students do not own the rights to this intellectual property.

If a student breaches the Student Code with respect to academic integrity, the University may take disciplinary action under the **Student Misconduct Procedure**.

The UNSW Student Code and the Student Misconduct Procedure can be found at:

<https://student.unsw.edu.au/plagiarism>

An online Module “[Working with Academic Integrity](https://student.unsw.edu.au/aim)” (<https://student.unsw.edu.au/aim>) is a six-lesson interactive self-paced Moodle module exploring and explaining all of these terms and placing them into your learning context. It will be the best one-hour investment you’ve ever made.

Plagiarism

Plagiarism is presenting another person's work or ideas as your own. Plagiarism is a serious breach of ethics at UNSW and is not taken lightly. So how do you avoid it? A one-minute video for an overview of how you can avoid plagiarism can be found <https://student.unsw.edu.au/plagiarism>.

Additional Support

ELISE (Enabling Library and Information Skills for Everyone)

ELISE is designed to introduce new students to studying at UNSW.

Completing the ELISE tutorial and quiz will enable you to:

- analyse topics, plan responses and organise research for academic writing and other assessment tasks
- effectively and efficiently find appropriate information sources and evaluate relevance to your needs
- use and manage information effectively to accomplish a specific purpose
- better manage your time
- understand your rights and responsibilities as a student at UNSW
- be aware of plagiarism, copyright, UNSW Student Code of Conduct and Acceptable Use of UNSW ICT Resources Policy
- be aware of the standards of behaviour expected of everyone in the UNSW community
- locate services and information about UNSW and UNSW Library

Some of these areas will be familiar to you, others will be new. Gaining a solid understanding of all the related aspects of ELISE will help you make the most of your studies at UNSW.

The *ELISE* training webpages:

<https://subjectguides.library.unsw.edu.au/elise/aboutelise>

Equitable Learning Services (ELS)

If you suffer from a chronic or ongoing illness that has, or is likely to, put you at a serious disadvantage, then you should contact the Equitable Learning Services (previously known as SEADU) who provide confidential support and advice.

They assist students:

- living with disabilities
- with long- or short-term health concerns and/or mental health issues
- who are primary carers
- from low SES backgrounds
- of diverse genders, sexes and sexualities
- from refugee and refugee-like backgrounds
- from rural and remote backgrounds
- who are the first in their family to undertake a bachelor-level degree.

Their web site is: <https://student.unsw.edu.au/els/services>

Equitable Learning Services (ELS) may determine that your condition requires special arrangements for assessment tasks. Once the School has been notified of these, we will make every effort to meet the arrangements specified by ELS.

Additionally, if you have suffered significant misadventure that affects your ability to complete the course, please contact your Lecturer-in-charge in the first instance.

Academic Skills Support and the Learning Centre

The Learning Centre offers academic support programs to all students at UNSW Australia. We assist students to develop approaches to learning that will enable them to succeed in their academic study. For further information on these programs please go to:

<http://www.lc.unsw.edu.au/services-programs>

Applications for Special Consideration for Missed Assessment

Please adhere to the Special Consideration Policy and Procedures provided on the web page below when applying for special consideration.

<https://student.unsw.edu.au/special-consideration>

Please note that the application is not considered by the Course Authority, it is considered by a centralised team of staff at the Nucleus Student Hub.

The School will contact you (via student email account) after special consideration has been granted to reschedule your missed assessment, for a *lab test or paper-based test* only.

For applications for special consideration for *assignment extensions*, please note that the new submission date and/or outcome will be communicated through the special consideration web site only, no communication will be received from the School.

For Dates on Final Term Exams and Supplementary Exams please check the “Key Dates for Exams” ahead of time to avoid booking holidays or work obligations.

<https://student.unsw.edu.au/exam-dates>

If you believe your application for Special Consideration has not been processed, you should email specialconsideration@unsw.edu.au immediately for advice.

Course Evaluation and Development (MyExperience)

Student feedback is very important to continual course improvement. This is demonstrated within the School of Mathematics and Statistics by the implementation of the UNSW online student survey *myExperience*, which allows students to evaluate their learning experiences in an anonymous way. *myExperience* survey reports are produced for each survey. They are released to staff after all student assessment results are finalised and released to students. Course convenor will use the feedback to make ongoing improvements to the course.