UNSW Course Outline

MATH3856 Introduction to Data and Machine Learning - 2024

Published on the 04 Sep 2024

General Course Information

Course Code : MATH3856
Year : 2024
Term : Term 3
Teaching Period : T3
Is a multi-term course? : No
Faculty : Faculty of Science
Academic Unit : School of Mathematics & Statistics
Delivery Mode : In Person
Delivery Format : Standard
Delivery Location : Kensington
Campus : Sydney
Study Level : Undergraduate
Units of Credit : 6

Useful Links

Handbook Class Timetable

Course Details & Outcomes

Course Description

A wide range of statistical methods and computational tools have been developed in the past few decades to gather information from data. This undergraduate course covers the key techniques in data mining and machine learning with theoretical background and applications,

delivered through a series of lectures and tutorials. 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 include Python and R.

Course Aims

This undergraduate 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 an 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 on 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.

Relationship to Other Courses

MATH5836 covers about 67 - 70 percent of the topics covered by "COMP9417 Machine Learning" and 30 - 35 percent of topics in "COMP9414 Artificial Intelligence" offered by the UNSW School of Computer Science and Engineering. MATH5836 requires intermediate to advanced Python/R programming skills and introductory statistics and linear algebra.

Course Learning Outcomes

Course Learning Outcomes
CLO1 : Identify the fundamentals of machine learning and data mining through model development.
CLO2 : Select appropriate statistical and machine learning approaches to analyse data.
CLO3 : Demonstrate an applied working knowledge of the methodologies covered with practical assignments.
CLO4 : Develop models for solving data mining problems that include clustering, regression, and classification.
CLO5 : Build machine learning models using real-world data sets and use evaluation metrics to compare their performance.

Course Learning Outcomes	Assessment Item
CLO1 : Identify the fundamentals of machine learning and data mining through model development.	• Quiz • Model Building and Evaluation • Final Exam
CLO2 : Select appropriate statistical and machine learning approaches to analyse data.	• Model Building and Evaluation • Final Exam
CLO3 : Demonstrate an applied working knowledge of the methodologies covered with practical assignments.	• Machine Learning Project • Model Building and Evaluation • Final Exam
CLO4 : Develop models for solving data mining problems that include clustering, regression, and classification.	• Machine Learning Project • Final Exam
CLO5 : Build machine learning models using real-world data sets and use evaluation metrics to compare their performance.	• Machine Learning Project • Final Exam

Learning and Teaching Technologies

Moodle - Learning Management System | Blackboard Collaborate | EdStem

Learning and Teaching in this course

TBA

Additional Course Information

Course Schedule

1. Data Processing and Linear Regression with Gradient Descent
2. Logistic Regression and Evaluation Metrics
3. Introduction to Neural Networks
4. Advances in Neural Networks
5. Introduction to Bayesian linear models and Bayesian neural Networks
6. Break
7. Decision Trees and Forests
8. Ensemble Learning: Boosting and Stacking
9. Unsupervised Learning
10. Emerging topics in AI: Deep learning and Ethics

Assessments

Assessment Structure

Assessment Item	Weight	Relevant Dates
Quiz Assessment Format: Individual	5%	Due Date: Week 3
Model Building and Evaluation Assessment Format: Individual	15%	Due Date: Week 5
Machine Learning Project Assessment Format: Group	25%	Due Date: Week 10
Final Exam Assessment Format: Individual	55%	Due Date: Exam Period

Assessment Details

Quiz

Assessment Overview

You will engage in an online open book quiz covering the first three weeks of the course content. This is a 30 minute online quiz that is held before the end of Week 3.

Feedback is provided within 2 weeks of completing the task.

Course Learning Outcomes

- CLO1 : Identify the fundamentals of machine learning and data mining through model development.

Detailed Assessment Description

Refer to Edstem

Assignment submission Turnitin type

This is not a Turnitin assignment

Generative AI Permission Level

No Assistance

This assessment is designed for you to complete without the use of any generative AI. You are not permitted to use any generative AI tools, software or service to search for or generate information or answers.
For more information on Generative AI and permitted use please see here.

Model Building and Evaluation

Assessment Overview

The assignment provides an opportunity for you to apply existing machine learning-based model code on benchmark datasets and provides an opportunity to learn about the model evaluation and reporting of results. This is due by the end of week 5.

Feedback is available within 2 weeks of submission.

Course Learning Outcomes

- CLO1 : Identify the fundamentals of machine learning and data mining through model development.
- CLO2 : Select appropriate statistical and machine learning approaches to analyse data.
- CLO3 : Demonstrate an applied working knowledge of the methodologies covered with practical assignments.

Assessment information

This assessment will require group work of minimum of 2 students.

Assignment submission Turnitin type

This assignment is submitted through Turnitin and students do not see Turnitin similarity reports.

Generative AI Permission Level

No Assistance

This assessment is designed for you to complete without the use of any generative AI. You are not permitted to use any generative AI tools, software or service to search for or generate information or answers.
For more information on Generative AI and permitted use please see here.

Machine Learning Project

Assessment Overview

In this project, you will work in a small group to apply machine learning methods in real-world applications. It will provide the opportunity to learn about the model evaluation and reporting of results. It will also enhance skills in technical report writing that incorporates literature review with a comprehensive presentation of results. The task involves the submission of code and a technical report which is at least a thousand words long and includes at least 10 references. The report also includes the results from experiments presented as Tables and Figures that are

discussed in detail. The report is accompanied by code and data that are submitted online. This will be due end of week 10. Feedback is provided two weeks after submission.

The laboratory classes are designed to provide you with practical experience and the tools to design experiments in the lab. They will be essential to build this project. You are required to attend a minimum of 6/9 laboratory classes to meet the hurdle pass requirement for this course.

Course Learning Outcomes

- CLO3 : Demonstrate an applied working knowledge of the methodologies covered with practical assignments.
- CLO4 : Develop models for solving data mining problems that include clustering, regression, and classification.
- CLO5 : Build machine learning models using real-world data sets and use evaluation metrics to compare their performance.

Assessment information

This assessment will require group work of minimum of 2 students.

Assignment submission Turnitin type

This assignment is submitted through Turnitin and students do not see Turnitin similarity reports.

Hurdle rules

A hurdle requirement or hurdle rule is a course requirement that must be fulfilled in order to pass the course. 7/9 lab attendance is mandatory.

Generative AI Permission Level

No Assistance

This assessment is designed for you to complete without the use of any generative AI. You are not permitted to use any generative AI tools, software or service to search for or generate information or answers.
For more information on Generative AI and permitted use please see here.

Final Exam

Assessment Overview

The final exam is designed to test your learning and problem-solving skills on all topics delivered across the term. The exam is 2 hours in duration and consists of MCQ, short answer responses, and practical, i.e coding and problem solving components. The details will be confirmed during

the course. The examination will occur during the official university examination period. Feedback is available through inquiry with the course convenor. Hurdle requirement: must achieve 50% to receive a passing grade in the course.

Course Learning Outcomes

- CLO1 : Identify the fundamentals of machine learning and data mining through model development.
- CLO2 : Select appropriate statistical and machine learning approaches to analyse data.
- CLO3 : Demonstrate an applied working knowledge of the methodologies covered with practical assignments.
- CLO4 : Develop models for solving data mining problems that include clustering, regression, and classification.
- CLO5 : Build machine learning models using real-world data sets and use evaluation metrics to compare their performance.

Assessment information

Final exam will have coding component

Assignment submission Turnitin type

This assignment is submitted through Turnitin and students do not see Turnitin similarity reports.

Hurdle rules

A hurdle requirement or hurdle rule is a course requirement that must be fulfilled in order to pass the course. At least 28/55 in Final Exam and attendance of 7/9 labs required to pass the course.

Generative AI Permission Level

No Assistance

This assessment is designed for you to complete without the use of any generative AI. You are not permitted to use any generative AI tools, software or service to search for or generate information or answers.
For more information on Generative AI and permitted use please see here.

General Assessment Information

Assessment 1: Quiz (5 %). You will engage in an online open-book quiz covering the first three weeks of the course content. This is a 30-minute online quiz that is held before the end of Week 3.

Assesment 2: Model building and evaluation (15 %). The assignment provides an opportunity

for you to apply existing machine learning-based model code on benchmark datasets and provides an opportunity to learn about the model evaluation and reporting of results. This is due by the end of week 5.

Assesment 3: Project (25%). In this project, you will work in a small group to apply machine learning methods in real-world applications. It will provide the opportunity to learn about the model evaluation and reporting of results. It will also enhance skills in technical report writing that incorporate a literature review with a comprehensive presentation of results. The task involves the submission of code and a technical report which is at least a thousand words long and includes at least 10 references. The report also includes the results from experiments presented as Tables and Figures that are discussed in detail. The report is accompanied by code and data that are submitted online. This will be due end of week 10.

Assesment 4: Final Exam (55%). The final exam is designed to test your learning and problem-solving skills on all topics delivered across the term. The exam consists of MCQ, short answer responses, and practical, i.e. – coding and problem-solving components. The details will be confirmed during the course. The examination will occur during the official university examination period.

Grading Basis

Standard

Requirements to pass course

At least 28/55 in Final Exam and attendance of 7/9 labs required to pass the course.

Course Schedule

Teaching Week/Module	Activity Type	Content
Week 0 : 2 September - 8 September	Other	N/A
Week 1 : 9 September - 15 September	Module	Data Processing and Linear Regression with Gradient Descent
Week 2 : 16 September - 22 September	Module	Logistic Regression and Evaluation Metrics
Week 3 : 23 September - 29 September	Module	Introduction to Neural Networks
Week 4 : 30 September - 6 October	Module	Advances in Neural Networks
Week 5 : 7 October - 13 October	Module	Introduction to Bayesian linear models and Bayesian neural Networks
Week 6 : 14 October - 20 October	Other	Non teaching week
Week 7 : 21 October - 27 October	Module	Decision Trees and Forests
Week 8 : 28 October - 3 November	Module	Ensemble Learning: Boosting and Stacking
Week 9 : 4 November - 10 November	Module	Unsupervised Learning
Week 10 : 11 November - 17 November	Module	Emerging topics in AI: Deep learning and Ethics

Attendance Requirements

Students are strongly encouraged to attend all classes and review lecture recordings.

General Schedule Information

Course Schedule

1. Data Processing and Linear Regression with Gradient Descent
2. Logistic Regression and Evaluation Metrics
3. Introduction to Neural Networks
4. Advances in Neural Networks
5. Introduction to Bayesian linear models and Bayesian neural Networks
6. Break
7. Decision Trees and Forests
8. Ensemble Learning: Boosting and Stacking
9. Unsupervised Learning
10. Emerging topics in AI: Deep learning and Ethics

Course Resources

Prescribed Resources

Hands-On Machine Learning with Scikit-Learn Keras and Tensor Flow

Print:

https://www.bookshop.unsw.edu.au/details.cgi?ITEMNO=9781098125974

https://www.bookshop.unsw.edu.au/details.cgi?ITEMNO=9781138492530

Digital:

https://unswbookshop.vitalsource.com/products/-v9781098122461

https://unswbookshop.vitalsource.com/products/-v9781000731071

Recommended Resources

TBA

Additional Costs

NA

Course Evaluation and Development

TBA

Staff Details

Position	Name	Email	Location	Phone	Availability	Equitable Learning Services Contact	Primary Contact
Convenor	Rohitash Chandra		Anita B Lawrence Centre (H13)	+61413071839	TBA	Yes	Yes

Other Useful Information

Academic Information

Upon your enrolment at UNSW, you share responsibility with us for maintaining a safe, harmonious and tolerant University environment.

You are required to:

- Comply with the University's conditions of enrolment.
- Act responsibly, ethically, safely and with integrity.
- Observe standards of equity and respect in dealing with every member of the UNSW community.
- Engage in lawful behaviour.
- Use and care for University resources in a responsible and appropriate manner.
- Maintain the University's reputation and good standing.

For more information, visit the UNSW Student Code of Conduct Website.

Academic Honesty and Plagarism

Referencing is a way of acknowledging the sources of information that you use to research your assignments. You need to provide a reference whenever you draw on someone else's words, ideas or research. Not referencing other people's work can constitute plagiarism.
Further information about referencing styles can be located at https://student.unsw.edu.au/referencing

Academic integrity is fundamental to success at university. Academic integrity can be defined as a commitment to six fundamental values in academic pursuits: honesty, trust, fairness, respect, responsibility and courage. At UNSW, this means that your work must be your own, and others'

ideas should be appropriately acknowledged. If you don't follow these rules, plagiarism may be detected in your work.
Further information about academic integrity, plagiarism and the use of AI in assessments can be located at:

- The Current Students site,
- The ELISE training site, and
- The Use of AI for assessments site.

The Student Conduct and Integrity Unit provides further resources to assist you to understand your conduct obligations as a student: https://student.unsw.edu.au/conduct

Submission of Assessment Tasks

Penalty for Late Submissions

UNSW has a standard late submission penalty of:

- 5% per day,
- for all assessments where a penalty applies,
- capped at five days (120 hours) from the assessment deadline, after which a student cannot submit an assessment, and
- no permitted variation.

Any variations to the above will be explicitly stated in the Course Outline for a given course or assessment task.

Students are expected to manage their time to meet deadlines and to request extensions as early as possible before the deadline.

Special Consideration

If circumstances prevent you from attending/completing an assessment task, you must officially apply for special consideration, usually within 3 days of the sitting date/due date. You can apply by logging onto myUNSW and following the link in the My Student Profile Tab. Medical documentation or other documentation explaining your absence must be submitted with your application. Once your application has been assessed, you will be contacted via your student email address to be advised of the official outcome and any actions that need to be taken from there. For more information about special consideration, please visit: https://student.unsw.edu.au/special-consideration

Important note: UNSW has a "fit to sit/submit" rule, which means that if you sit an exam or

submit a piece of assessment, you are declaring yourself fit to do so and cannot later apply for Special Consideration. This is to ensure that if you feel unwell or are faced with significant circumstances beyond your control that affect your ability to study, you do not sit an examination or submit an assessment that does not reflect your best performance. Instead, you should apply for Special Consideration as soon as you realise you are not well enough or are otherwise unable to sit or submit an assessment.

Faculty-specific Information

Additional support for students

- The Current Students Gateway
- Student Support
- Academic Skills and Support
- Student Wellbeing, Health and Safety
- Equitable Learning Services
- UNSW IT Service Centre
- Science EDI Student Initiatives, Offerings and Guidelines

School-specific Information

School of Mathematics and Statistics 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 web site starting at: The School of Mathematics and Statistics 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 Initial Handout and Course Home Page. Lack of knowledge about a policy will not be an excuse for failing to follow the procedure in it.

Special Consideration - Short Extension Policy

The School of Mathematics and Statistics has carefully reviewed its range of assignments and

projects to determine their suitability for automatic short extensions as set out by the UNSW Short Extension Policy. Upon comprehensive examination of our course offerings that incorporate these types of assessments, we have concluded that our current deadline structures already accommodate the possibility of unexpected circumstances that may lead students to require additional days for submission. Consequently, the School of Mathematics and Statistics has decided to universally opt out of the Short Extension provision for all its courses, having pre-emptively integrated flexibility into our assessment deadlines. The decision is subject to revision in response to the introduction of new course offerings. Students may still apply for Special Consideration via the usual procedures.

Computing Lab

The main computing laboratory is room G012 of the Anita B.Lawrence Centre (formerly Red Centre). You can get to this lab by entering the building through the main entrance to the School of Mathematics (on the Mezzanine Level) and then going down the stairs to the Ground Level. A second smaller lab is Room M020, located on the mezzanine level through the glass door (and along the corridor) opposite the School's entrance.

For more information, including opening hours, see the computing facilities webpage. Remember that there will always be unscheduled periods when the computers are not working because of equipment problems and that this is not a valid excuse for not completing assessments on time.

School Contact Information

Please visit the School of Mathematics and Statistics website for a range of information.

For information on Courses, please go to “Student life & resources” and either Undergraduate and/or Postgraduate and respective "Undergraduate courses" and "Postgraduate courses" for information on all course offerings.

All school policies, forms and help for students can be located by going to the "Student Services" within “Student life & resources” page. We also post notices in "Student noticeboard" for your information. Please familiarise yourself with the information found in these locations. If you cannot find the answer to your queries on the web you are welcome to contact the Student Services Office directly.

Undergraduate

E: ug.mathsstats@unsw.edu.au

P: 9385 7011 or 9385 7053

Postgraduate

E: pg.mathsstats@unsw.edu.au

P: 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 use your UNSW student email and state your student number in all emails to us.**