UNSW Course Outline

COMP9417 Machine Learning and Data Mining - 2024

Published on the 28 Jan 2024

General Course Information

Course Code : COMP9417
Year : 2024
Term : Term 1
Teaching Period : T1
Is a multi-term course? : No
Faculty : Faculty of Engineering
Academic Unit : School of Computer Science and Engineering
Delivery Mode : Multimodal
Delivery Format : Standard
Delivery Location : Kensington
Campus : Sydney
Study Level : Undergraduate, Postgraduate
Units of Credit : 6

Useful Links

Handbook Class Timetable

Course Details & Outcomes

Course Description

Machine learning is the algorithmic approach to learning from data. The course also covers aspects of data mining, the application of machine learning to obtain insight from data. In this course machine learning algorithms are placed in the context of their theoretical foundations in

order to understand their derivation and correct application. Machine learning also is an empirical science, where performance of algorithms must be rigorously evaluated on datasets. Completion of this course will contribute to further learning in advanced topics such as deep learning, bioinformatics, computer vision, and robotics. Topics covered in the course include: linear models for regression and classification, local methods (nearest neighbour), tree learning, kernel machines, neural networks, unsupervised learning, ensemble learning, and learning theory. To expand and extend the development of theory and algorithms presented in lectures, practical examples will be given in tutorials and programming tasks during the project.

Course Aims

This course aims to expose students to the theory, algorithms and empirical work that are essential inter-dependent components of machine learning. In particular, the assignments are aimed at giving students an opportunity for active learning via practical experience in applying machine learning to real applications. The second assignment has a broad scope and should be treated as a small-scale project with submission of software and a written report. COMP9417 forms part of a suite of AI courses that together comprise the Artificial Intelligence major in the undergraduate computer science programs and the postgraduate information technology program.

Relationship to Other Courses

Before commencing this course, students should have completed the pre-requisite courses (or equivalent) and ensure they have acquired knowledge in the relevant areas:

- Mathematical assumed knowledge is the completion of basic university mathematics courses, such as the UNSW courses MATH1131 and MATH1231 (e.g. be familiar with vector calculus and linear algebra, or be willing to learn them independently).
- Additionally, in practice, some knowledge of basic probability, statistics (e.g. MATH2801, and the listed excluded equivalents MATH2089, MATH2099, MATH2859, MATH2901, BEES2041, ECON3209, CVEN2002), and logic will be the starting point for some course materials (e.g., as in a typical university course on discrete mathematics).
- Be able to program well in Python or be willing to learn it independently.

Course Learning Outcomes

Course Learning Outcomes
CLO1 : Construct a well-defined learning problem for a given task, selecting representations for the data input and output, the model, and the learning algorithm
CLO2 : Compare different algorithms according to the properties of their inputs and outputs, and the computational methods used
CLO3 : Develop and describe algorithms to solve a well-defined learning problem
CLO4 : Implement machine learning algorithms, apply them to realistic datasets and collect results to enable evaluation of their performance
CLO5 : Explain key concepts from the foundations of learning theory, describe their applicability, and express knowledge of the general limits of machine learning

Course Learning Outcomes	Assessment Item
CLO1 : Construct a well-defined learning problem for a given task, selecting representations for the data input and output, the model, and the learning algorithm	• Homework 1 • Machine learning project • EXAM3
CLO2 : Compare different algorithms according to the properties of their inputs and outputs, and the computational methods used	• Homework 2 • Homework 1 • Machine learning project • EXAM3
CLO3 : Develop and describe algorithms to solve a well-defined learning problem	• Homework 2 • Homework 1 • Machine learning project
CLO4 : Implement machine learning algorithms, apply them to realistic datasets and collect results to enable evaluation of their performance	• Homework 2 • Machine learning project
CLO5 : Explain key concepts from the foundations of learning theory, describe their applicability, and express knowledge of the general limits of machine learning	• EXAM3

Learning and Teaching Technologies

Moodle - Learning Management System

Learning and Teaching in this course

The delivery of this course is multimodal, with in-person/live-streaming lectures and in-person/online tutorials. All course materials can be found on the course Moodle page.

Additional Course Information

Student Conduct

The **Student Code of Conduct** (Information, Policy) sets out what the University expects from students as members of the UNSW community. As well as the learning, teaching and research environment, the University aims to provide an environment that enables students to achieve their full potential and to provide an experience consistent with the University's values and guiding principles. A condition of enrolment is that students *inform themselves* of the University's rules and policies affecting them, and conduct themselves accordingly.

In particular, students have the responsibility to observe standards of equity and respect in dealing with every member of the University community. This applies to all activities on UNSW premises and all external activities related to study and research. This includes behaviour in person as well as behaviour on social media, for example, Facebook groups set up for the purpose of discussing UNSW courses or coursework. Behaviour that is considered in breach of the Student Code Policy as discriminatory, sexually inappropriate, bullying, harassing, invading another's privacy or causing any person to fear for their personal safety is serious misconduct and can lead to severe penalties, including suspension or exclusion from UNSW.

If you have any concerns, you may raise them with your lecturer or approach the School Ethics Officer, Grievance Officer, or one of the student representatives.

Plagiarism is defined as using the words or ideas of others and presenting them as your own. UNSW and CSE treat plagiarism as academic misconduct, which means that it carries penalties as severe as being excluded from further study at UNSW. There are several online sources to help you understand what plagiarism is and how it is dealt with at UNSW:

- Plagiarism and Academic Integrity
- UNSW Plagiarism Procedure

Make sure that you read and understand these resources. Ignorance is not accepted as an excuse for plagiarism. In particular, you are also responsible that your assignment files are not accessible by anyone but you by setting the correct permissions in your CSE directory and code repository, if using. Note also that plagiarism includes paying or asking another person to do a piece of work for you and then submitting it as your own work.

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.

If you haven't done so yet, please take the time to read the full text of

- UNSW's policy regarding academic honesty and plagiarism

The pages below describe the policies and procedures in more detail:

- Student Code Policy
- Student Misconduct Procedure
- Plagiarism Policy Statement
- Plagiarism Procedure

You should also read the following page which describes your rights and responsibilities in the CSE context:

- Essential Advice for CSE Students

Assessments

Assessment Structure

Assessment Item	Weight	Relevant Dates
Homework 1 Assessment Format: Individual	7.5%	Start Date: Not Applicable Due Date: 04/03/2024 12:00 AM
Homework 2 Assessment Format: Individual	7.5%	Start Date: Not Applicable Due Date: 25/03/2024 12:00 AM
Machine learning project Assessment Format: Individual	30%	Start Date: Week 5 to 7 Due Date: Week 9 - 10
EXAM3 Assessment Format: Individual	55%	Start Date: Exam period Due Date: Not Applicable

Assessment Details

Homework 1

Assessment Overview

This homework deals with regression and optimization, and involves Python.

Marking is done with respect to a rubric, provided in the homework specification, and feedback will be provided with the online assessment.

Details of submission, deadlines and late penalties, etc. will be in the specification.

The overall course mark will be the sum of the marks for the course components

Course Learning Outcomes

- CLO1 : Construct a well-defined learning problem for a given task, selecting representations for the data input and output, the model, and the learning algorithm
- CLO2 : Compare different algorithms according to the properties of their inputs and outputs, and the computational methods used
- CLO3 : Develop and describe algorithms to solve a well-defined learning problem

Homework 2

Assessment Overview

This homework deals with the construction of models, and involves Python.

Marking is done with respect to a rubric, provided in the homework specification, and feedback will be provided with the online assessment.

Details of submission, deadlines and late penalties, etc. will be in the specification.

The overall course mark will be the sum of the marks for the course components

Course Learning Outcomes

- CLO2 : Compare different algorithms according to the properties of their inputs and outputs, and the computational methods used
- CLO3 : Develop and describe algorithms to solve a well-defined learning problem
- CLO4 : Implement machine learning algorithms, apply them to realistic datasets and collect results to enable evaluation of their performance

Machine learning project

Assessment Overview

This assignment has a broad scope and should be treated as a small-scale project with submission of software and a written report.

Marking is done with respect to a rubric, provided in the project specification, and feedback will be provided with the online assessment.

Details of submission, deadlines and late penalties, etc. will be in the specification.

The overall course mark will be the sum of the marks for the course components

Course Learning Outcomes

- CLO1 : Construct a well-defined learning problem for a given task, selecting representations for the data input and output, the model, and the learning algorithm
- CLO2 : Compare different algorithms according to the properties of their inputs and outputs, and the computational methods used
- CLO3 : Develop and describe algorithms to solve a well-defined learning problem
- CLO4 : Implement machine learning algorithms, apply them to realistic datasets and collect results to enable evaluation of their performance

EXAM3

Assessment Overview

This is a written examination which asks questions on a range of topics in the course.

It is run on the Inspera platform.

No feedback is provided on exams, except as a final mark.

Course Learning Outcomes

- CLO1 : Construct a well-defined learning problem for a given task, selecting representations for the data input and output, the model, and the learning algorithm
- CLO2 : Compare different algorithms according to the properties of their inputs and outputs, and the computational methods used
- CLO5 : Explain key concepts from the foundations of learning theory, describe their applicability, and express knowledge of the general limits of machine learning

General Assessment Information

All assessment work items (except maybe for the final exam) will involve electronic submission via the Moodle page. Details of submission, deadlines, late penalties, etc., will be in the specifications.

The overall course mark will be the sum of the marks for the course components.

Grading Basis

Standard

Requirements to pass course

The requirement to pass the course is to acquire a minimum mark of 50 out of 100.

Course Schedule

Attendance Requirements

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

General Schedule Information

Week	Lecture	Tutorial	Assignment	Other
1	Regression	No tutorial	-	-
2	Classification	Regression	-	-
3	Classification	Classification	-	-
4	Tree Learning	Classification	Homework 1 due	Project Group Formation due
5	Kernel Methods	Tree Learning	-	-
6	Flexibility Week	Flexibility Week	-	-
7	Ensemble Learning	Kernel Methods	Homework 2 due	-
8	Neural Network	Ensemble Learning	-	-
9	Unsupervised Learning	Neural Network	-	-
10	Learning Theory	Unsupervised Learning	-	Group Project due

Course Resources

Recommended Resources

Owing to the expansion of machine learning in recent years, and the wide availability of online materials, it is no longer possible to recommend a single textbook for this course. However, below is a list of books (those with an asterisk have copies freely available online) that can be consulted to back up and expand on the course content. If you plan to continue with machine learning, any of these (and many others) are worth reading:

- Hastie, T., Tibshirani, R. and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. * Springer, 2009.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. An Introduction to Statistical Learning with Applications in R . * Springer, 2017.
- Flach, P. Machine learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, 2012.
- Rogers, S. and Girolami, M. A First Course in Machine Learning (2nd Edition) . Chapman and Hall/CRC, 2016.
- Bishop, C. Pattern recognition and Machine Learning. Springer, 2006.
- Charniak, E. Introduction to Deep Learning . MIT Press, 2019.
- Blum, A., Hopcroft, J. and Kannan, R. Foundations of Data Science .* 2018.
- Shalev-Shwartz, S. and Ben-David, S. Understanding Machine Learning: From Theory to Algorithms . Cambridge University Press, 2014.
- Deisenroth. M. P., Faisal, A. A., and Ong, C. S., Mathematics for Machine Learning. *Cambridge University Press, 2020.
- Geron, A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition . O'Reilly Media, Inc, 2019.

Other resources (e.g. links to online documentation) will be made available in the relevant course materials.

Course Evaluation and Development

This course is evaluated each session using myExperience.

Following the previous offerings of this course, students requested increasing homework weights to be proportional to the amount of effort going into it. Therefore, we have increased the homework's weight to address this feedback since the last offering. Students also asked that tutors explain the material rather than Q&A sessions, which we will implement in this offering. Students found the weekly assignments very time-consuming, therefore in this offering, we will not have weekly assignments.

Staff Details

Position	Name	Email	Location	Phone	Availability	Equitable Learning Services Contact	Primary Contact
Head lecturer	Gelareh Moh ammadi					Yes	No
Administrator	Omar Ghattas					No	Yes

Other Useful Information

Academic Information

I. Special consideration and supplementary assessment

If you have experienced an illness or misadventure beyond your control that will interfere with your assessment performance, you are eligible to apply for Special Consideration prior to, or within 3 working days of, submitting an assessment or sitting an exam.

Please note that UNSW has a Fit to Sit rule, which means that if you sit an exam, you are declaring yourself fit enough to do so and cannot later apply for Special Consideration.

For details of applying for Special Consideration and conditions for the award of supplementary assessment, please see the information on UNSW's Special Consideration page.

II. Administrative matters and links

All students are expected to read and be familiar with UNSW guidelines and polices. In particular, students should be familiar with the following:

- Attendance
- UNSW Email Address
- Special Consideration
- Exams
- Approved Calculators
- Academic Honesty and Plagiarism
- Equitable Learning Services

III. Equity and diversity

Those students who have a disability that requires some adjustment in their teaching or learning environment are encouraged to discuss their study needs with the course convener prior to, or at the commencement of, their course, or with the Equity Officer (Disability) in the Equitable Learning Services. Issues to be discussed may include access to materials, signers or note-takers, the provision of services and additional exam and assessment arrangements. Early notification is essential to enable any necessary adjustments to be made.

IV. Professional Outcomes and Program Design

Students are able to review the relevant professional outcomes and program designs for their streams by going to the following link: https://www.unsw.edu.au/engineering/student-life/student-resources/program-design.

Note: This course outline sets out the description of classes at the date the Course Outline is published. The nature of classes may change during the Term after the Course Outline is published. Moodle or your primary learning management system (LMS) should be consulted for the up-to-date class descriptions. If there is any inconsistency in the description of activities between the University timetable and the Course Outline/Moodle/LMS, the description in the Course Outline/Moodle/LMS applies.

Academic Honesty and Plagarism

UNSW has an ongoing commitment to fostering a culture of learning informed by academic integrity. All UNSW 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.*

Plagiarism is a type of intellectual theft. It can take many forms, from deliberate cheating to accidentally copying from a source without acknowledgement. UNSW has produced a website with a wealth of resources to support students to understand and avoid plagiarism, visit: student.unsw.edu.au/plagiarism. The Learning Centre assists students with understanding academic integrity and how not to plagiarise. They also hold workshops and can help students one-on-one.

You are also reminded that careful time management is an important part of study and one of the identified causes of plagiarism is poor time management. Students should allow sufficient time for research, drafting and the proper referencing of sources in preparing all assessment tasks.

Repeated plagiarism (even in first year), plagiarism after first year, or serious instances, may also be investigated under the Student Misconduct Procedures. The penalties under the procedures can include a reduction in marks, failing a course or for the most serious matters (like plagiarism in an honours thesis or contract cheating) even suspension from the university. The Student Misconduct Procedures are available here:

www.gs.unsw.edu.au/policy/documents/studentmisconductprocedures.pdf

Submission of Assessment Tasks

Work submitted late without an approved extension by the course coordinator or delegated authority is subject to a late penalty of five percent (5%) of the maximum mark possible for that assessment item, per calendar day.

The late penalty is applied per calendar day (including weekends and public holidays) that the assessment is overdue. There is no pro-rata of the late penalty for submissions made part way through a day. This is for all assessments where a penalty applies.

Work submitted after five days (120 hours) will not be accepted and a mark of zero will be awarded for that assessment item.

For some assessment items, a late penalty may not be appropriate. These will be clearly indicated in the course outline, and such assessments will receive a mark of zero if not completed by the specified date. Examples include:

- Weekly online tests or laboratory work worth a small proportion of the subject mark;
- Exams, peer feedback and team evaluation surveys;
- Online quizzes where answers are released to students on completion;
- Professional assessment tasks, where the intention is to create an authentic assessment that has an absolute submission date; and,
- Pass/Fail assessment tasks.

Faculty-specific Information

Engineering Student Support Services – The Nucleus - enrolment, progression checks, clash requests, course issues or program-related queries

Engineering Industrial Training – Industrial training questions

UNSW Study Abroad – study abroad student enquiries (for inbound students)

UNSW Exchange – student exchange enquiries (for inbound students)

UNSW Future Students – potential student enquiries e.g. admissions, fees, programs, credit transfer

Phone

(+61 2) 9385 8500 – Nucleus Student Hub

(+61 2) 9385 7661 – Engineering Industrial Training

(+61 2) 9385 3179 – UNSW Study Abroad and UNSW Exchange (for inbound students)

School Contact Information

CSE Help! - on the Ground Floor of K17

- For assistance with coursework assessments.

The Nucleus Student Hub - https://nucleus.unsw.edu.au/en/contact-us

- Course enrolment queries.

Grievance Officer - grievance-officer@cse.unsw.edu.au

- If the course convenor gives an inadequate response to a query or when the course convenor does not respond to a query about assessment.

Student Reps - stureps@cse.unsw.edu.au

- If some aspect of a course needs urgent improvement. (e.g. Nobody responding to forum queries, cannot understand the lecturer)