Machine Learning & Data Mining T3 2019

Course Outline

Course Details

Course Code	COMP9417
Course Title	Machine Learning and Data Mining
Convenor	Gelareh Mohammadi
Admin	Anahita Namvar
Classes	Timetable for all classes
Units of Credit	6
Handbook Entry	http://www.handbook.unsw.edu.au/postgraduate/courses/current/COMP9417.html

Special Consideration

If your work in this course is affected by unforeseen adverse circumstances, you should apply for Special Consideration. If your request is reasonable and your work has clearly been impacted, then .

- for an assignment, you may be granted an extension
- for the Final Exam, you may be offered a Supplementary Exam

Note the use of the word "may". None of the above is guaranteed. It depends on you making a convincing case that the circumstances have clearly impacted your ability to work.

UNSW handles special consideration requests centrally (in the Student Lifecycle division), so all special consideration requests must be submitted via the UNSW Special Consideration website.

Special consideration requests must be accompanied by documentation, which will be verified by Student Lifecycle. Do not email the course convenor directly about special consideration.

If you cannot attend the Final Exam because of illness or misadventure, then you must submit a Special Consideration request, with documentation, through MyUNSW within 24 hours of the exam. If your request is reasonable, then you will be awarded a Supplementary Exam.

Note that UNSW expects you to be available to sit Supplementary Exams if required. If you are awarded a Supplementary Exam and do not attend, then your exam mark will be zero.

For further details on special consideration, see the UNSW Student website .

If you are registered with Disability Services, please forward your documentation to Course email "cs9417@cse.unsw.edu.au" within the first two weeks of term.

Course Summary

This course explores machine learning as the algorithmic approach to learning from data. The course also covers key aspects of data mining, which is understood as the application of machine learning tools to obtain insight from data. Algorithms are placed in the context of their theoretical foundations in order to understand their derivation and correct application. Topics include: linear models for regression and classification, local methods (nearest neighbour), neural networks, tree learning, kernel machines, unsupervised learning, ensemble learning, computational and statistical learning theory, and Bayesian learning. To expand and extend the development of theory and algorithms presented in lectures, practical applications will be given in tutorials and programming tasks during the project.

Assumed Knowledge

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

- HRRF0001
- WHSR2015
- OHSA1402
- ENGG4999-5196_00946
- Courses
  - (hidden)
    - (hidden)
      - COMPSC - School of Computer Science and Engineering
        - 2013 Session 1
        - 2013 Session 2
        - 2014 Session 1
        - 2014 Session 2
        - 2015 Session 1
        - 2015 Session 2
        - 2016 Session 1
        - 2016 Session 2
        - 2016 Summer
        - 2017 Session 1
        - Sandpit
        - COMP3231-5206_01495
        - ENGG1811-5206_01494
        - COMP9991-5206_01493
        - COMP9447-5206_01492
        - COMP9444-5206_01491
        - COMP9323-5206_01490
        - COMP9302-5206_01489
        - COMP9301-5206_01488
        - COMP9020-5206_01487
        - COMP6452-5206_01486
        - COMP6448-5206_01485
        - COMP9417-COMP9417-5199_00104
          - Participants
          - Badges
          - Competencies
          - Grades
          - General
          - Welcome Message and course outline
            - Course Outline
          - Homework & Assignment
          - Week 1: Regression
          - Week 2: Classification 1
          - Week 3: Classification 2
          - Week 4
          - Week 5: Tree Learning
          - Week 6: Kernel Methods
          - Week 7: Ensemble Learning
          - Week 8: Neural Net
          - Week 9
          - Week 10
- Sandpit Request Form
- Accessibility report

- Prerequisite is COMP1927 Computing 2 or equivalent. Waivers may be granted where applicable (see Course Coordinator)
- Mathematical assumed knowledge is completion of basic university mathematics courses, such as the UNSW courses MATH1131 and MATH1231
- Additionally, in practice, some knowledge of basic probability, statistics and logic will be the starting point for some course materials (e.g., as in a typical university course on discrete mathematics).
- Ability to program and construct working software in a general-purpose programming language (e.g., C, Java, Perl, Python, etc.) is assumed. An important part of practical data mining is "data wrangling", i.e., the pre-processing, filtering, cleaning, etc. of datasets; for this you need to have mastered Unix tools such as those taught in COMP2041 Software Construction, or equivalents such as can be found in Python, R, Matlab/Octave, etc.

Student Learning Outcomes

After completing this course, students will be able to:

1. set up a well-defined learning problem for a given task
2. select and define a representation for data to be used as input to a machine learning algorithm
3. select and define a representation for the model to be output by a machine learning algorithm
4. compare different algorithms according to the properties of their inputs and outputs
5. compare different algorithms in terms of similarities and differences in the computational methods used
6. develop and describe algorithms to solve a learning problem in terms of the inputs, outputs and computational methods used
7. express key concepts from the foundations of computational and statistical learning theory and demonstrate their applicability
8. express knowledge of general capabilities and limitations of machine learning from computational and statistical theory
9. use or extend or invent algorithms in applications to real-world data sets and collect results to enable evaluation and comparison of their performance

This course contributes to the development of the following graduate capabilities:

Graduate Capability	Acquired in
Scholars capable of independent and collaborative enquiry, rigorous in their analysis, critique and reflection, and able to innovate by applying their knowledge and skills to the solution of novel as well as routine problems	lectures, tutorials, homework, project and exam
Entrepreneurial leaders capable of initiating and embracing innovation and change, as well as engaging and enabling others to contribute to change	tutorials, homework and project
Professionals capable of ethical, self- directed practice and independent lifelong learning	suggested references, tutorials and project
Global citizens who are culturally adept and capable of respecting diversity and acting in a socially just and responsible way	lectures, tutorials and project

Teaching Strategies

- Lectures ... introduce concepts, show examples
- Tutorials ... reinforce concepts and provide additional examples
- Homework and Project .. allow students to solve significant problems

Teaching Rationale

This course is taught to emphasise that theory, algorithms and empirical work are essential inter-dependent components of machine learning. Teaching is mainly focused on lectures and assessed practical work on topics in machine learning, with tutorials to expand and reinforce the lecture content. Assessment is by two marked homeworks, a project and a final exam. The assignments are aimed at giving students an opportunity for active learning in a structured way with submission deadlines. The purpose is to give students practical experience of machine learning and relate lecture material 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.

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 course work. 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 on-line 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. 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

Assessment

Item	Topics	Due	Marks	Contributes to
Homework 1	Applications of machine learning	Week 4	5%	1,2,3,4,5
Homework 2	Applications of machine learning	Week 7	5%	1,2,3,4,5,6
Assignment	Machine learning project	Week 10	30%	1-9
Final Exam	All topics	Exam period	60%	1-9

All assessment work items (except for the final exam) will involve electronic submission via the CSE <t>give</t> system. Details of submission, deadlines and late penalties, etc. will be in the specifications.

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

The outcome of attaining a course mark in the range 45-49 will be decided on a case-by-case basis.

Course Schedule

Note: this schedule may be subject to change !

Week	Lecture	Tutorial	Assignment	Quiz
1	Regression	**No Tutorial**	-	-
2	Classification	Regression	-	-
3	Classification	Classification	-	-
4	No Lectures	No Tutorial	Homework 1 due	-
5	Tree Learning	Classification (group formation for the project)	-	-
6	Kernel Methods	Tree Learning	-	-
7	Ensemble Learning	Kernel Methods	Homework 2 due	-
8	Neural networks	Ensemble Learning	-	-
9	Unsupervised Learning	Neural networks	-	-
10	Learning Theory	Unsupervised Learning	Project due	-
Exam period				Final Exam

Resources for Students

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.
- Murphy, K. Machine learning: a probabilistic perspective. MIT press, 2012.
- Barber, D. Bayesian reasoning and machine learning. * Cambridge University Press, 2012.
- Bishop, C. Pattern recognition and Machine Learning. Springer, 2006.
- Charniak, E. Introduction to Deep Learning . MIT Press, 2019.
- Goodfellow, I., Bengio, Y. and Courville, A. Deep Learning. * MIT Press, 2016.
- 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.

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

Course Evaluation and Development

This course is evaluated each session using the myExperience system.

In the previous offering of this course, students requested that tutorials should be made available sooner, more examples should be given in lectures and feedback on assessment results should be made available more quickly.

Based on these comments, we are restructuring some of the tutorial material and assessment methods to better address the needs of a larger class.

Last modified: Wednesday, 13 November 2019, 9:06 PM