# Course Outline

## Contents


- Course Details
- Course Summary
- Assumed Knowledge
- Student Learning Outcomes
- Teaching Strategies
- Teaching Rationale
- Student Conduct
- Assessment
- Course Schedule
- Resources for Students
- Course Evaluation and Development


## Course Details

| | |
|---|---|
| **Course Code** | COMP9417 |
| **Course Title** | Machine Learning and Data Mining |
| **Convenor** | Michael Bain (/users/z9400013) |
| **Admin** | Omar Ghattas (mailto:mailto:omarghattas1991@gmail.com) |
| **Classes** | Timetable for all classes (http://timetable.unsw.edu.au/2019/COMP9417.html) |
| **Consultations** | TBA (online) |
| **Units of Credit** | 6 |
| **Course Website** | http://www.cse.unsw.edu.au/~cs9417/ (http://www.cse.unsw.edu.au/~cs9417/) |
| **Handbook Entry** | http://www.handbook.unsw.edu.au/postgraduate/courses/current/COMP9417.html (http://www.handbook.unsw.edu.au/postgraduate/courses/current/COMP9417.html) |

## 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:

- 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, definitions and methods
- Tutorials: reinforce concepts and provide examples
- Homeworks, Project: introduce practical applications of methods and 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 (https://student.unsw.edu.au/conduct) , Policy (https://www.gs.unsw.edu.au/policy/documents/studentcodepolicy.pdf) ) 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 (mailto:ethics-officer@cse.unsw.edu.au) , Grievance Officer (mailto:grievance-officer@cse.unsw.edu.au) , or one of the student representatives.

**Plagiarism** is defined as (https://student.unsw.edu.au/plagiarism) 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 (https://student.unsw.edu.au/plagiarism)
- UNSW Plagiarism Procedure (https://www.gs.unsw.edu.au/policy/documents/plagiarismprocedure.pdf)

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 (https://student.unsw.edu.au/plagiarism)

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

- Student Code Policy (https://www.gs.unsw.edu.au/policy/documents/studentcodepolicy.pdf)
- Student Misconduct Procedure (https://www.gs.unsw.edu.au/policy/documents/studentmisconductprocedures.pdf)
- Plagiarism Policy Statement (https://www.gs.unsw.edu.au/policy/documents/plagiarismpolicy.pdf)
- Plagiarism Procedure (https://www.gs.unsw.edu.au/policy/documents/plagiarismprocedure.pdf)

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

- Essential Advice for CSE Students (https://www.engineering.unsw.edu.au/computer-science-engineering/about-us/organisational-structure/student-services/policies/essential-advice-for-cse-students)

# Assessment

| Item | Topics | Due | Marks | Contributes to |
|---|---|---|---|---|
| Homework 0 | Background for machine learning | Week 1 | 1% | |
| Homework 1 | Applications of machine learning | Week 3 | 7% | 1,2,3,4,5 |
| Homework 2 | Applications of machine learning | Week 7 | 7% | 1,2,3,4,5,6 |
| Assignment | Machine learning project | Week 9 | 30% | 1-6,9 |
| Final Exam | All topics | Exam period | 55% | 1-8 |

All assessment items will involve electronic submission. 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.

# Course Schedule

Note: this schedule may be subject to change !

| Week | Lecture | Tutorial | Assignment | Quiz |
|---|---|---|---|---|
| 1 | Regression | - | Homework 0 due | - |
| 2 | Classification | Regression | - | - |
| 3 | Classification | Classification | Homework 1 due | - |
| 4 | Tree Learning | Classification | - | - |
| 5 | Kernel Methods | Tree Learning | - | - |

| | | | | |
|---|---|---|---|---|
| 6 | *Flexibility Week* | *Flexibility Week* | - | - |
| 7 | Neural Learning | Kernel Methods | Homework 2 due | - |
| 8 | Ensemble Learning | Neural Learning | - | - |
| 9 | Unsupervised Learning | Ensemble Learning | Project due | - |
| 10 | Learning Theory | Unsupervised Learning | - | - |
| 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. (http://web.stanford.edu/~hastie/ElemStatLearn/) * Springer, 2009.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. An Introduction to Statistical Learning with Applications in R (https://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf) . * Springer, 2017.
- Flach, P. Machine learning: The Art and Science of Algorithms that Make Sense of Data. (http://www.cs.bris.ac.uk/~flach/mlbook/) Cambridge University Press, 2012.
- Rogers, S. and Girolami, M. A First Course in Machine Learning (2nd Edition) (https://www.amazon.com.au/Course-Machine-Learning-Pattern-Recognition-ebook/dp/B01N7ZEBK8) . Chapman and Hall/CRC, 2016.
- Barber, D. Bayesian reasoning and machine learning. (http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage) * Cambridge University Press, 2012.
- Bishop, C. Pattern recognition and Machine Learning. (http://www.microsoft.com/en-us/research/people/cmbishop/) Springer, 2006.
- Charniak, E. Introduction to Deep Learning (https://www.footprint.com.au/product-detail.asp?SubSection=Browse+by+Subject&product=9780262039512&ParentPage=product-listing%2Easp%3Fkeywords%3D%26scope%3D%26Category%3DCMP) . MIT Press, 2019.
- Shalev-Shwartz, S. and Ben-David, S. Understanding Machine Learning: From Theory to Algorithms (https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf) . * Cambridge University Press, 2014.
- Goodfellow, I., Bengio, Y. and Courville, A. Deep Learning. (http://www.deeplearningbook.org) * MIT Press, 2016.
- Blum, A., Hopcroft, J. and Kannan, R. Foundations of Data Science (https://www.cs.cornell.edu/jeh/book.pdf) . * 2018.
- Deisenroth, M., Faisal, A., and Ong, C. Mathematics for Machine Learning (https://mml-book.github.io) . * Cambridge University Press, 2020.
- Geron, A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition (https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/) . O'Reilly Media, Inc, 2019.

- Murphy, K. Probabilistic Machine Learning: An Introduction (https://probml.github.io/pml-book/book1.html) . * MIT Press, 2021.

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 myExperience.

Following the previous offering of this course, students requested that the online delivery of the course should be improved. Two specific issues were that the duration of lecture videos should be reduced and that they should be more interactive. These issues will be addressed in this term, by revising some of the lecture material, and enabling interaction via Teams.

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

Resource created 6 months ago (Friday 21 May 2021, 12:35:51 PM), last modified 5 months ago (Friday 11 June 2021, 02:41:28 PM).

Comments

(/COMP9417/21T2/forums/search?forum_choice=resource/61170)

(/COMP9417/21T2/forums/resource/61170)

There are no comments yet.